UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Chimiche

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE MOLECOLARI
INDIRIZZO SCIENZE CHIMICHE
CICLO XXV

STRUCTURE AND INTERACTIONS OF THE STAS DOMAINS AND OF THE MYB/SANT DOMAINS BY MEANS OF NMR TECHNIQUES

Direttore della Scuola : Ch.mo Prof. Antonino Polimeno

Supervisore :Ch.mo Prof. Massimo Bellanda

Dottoranda : Meri Montecchio

TABLE OF CONTENTS

FIRST PART

1. Introduction

3. Materials and methods

4. Results and discussion

SECOND PART

6. Introduction

8. Materials and methods

9. Results and discussion

LIST OF ABBREVIATIONS

ABACUS: Applied BACUS
ASA: Anti Sigma factor Antagonist
CSP: Chemical Shift Perturbation
DTT: Dithiothreitol
FAWN: Fragment Assignment With NOE
FPLC: Fast Protein Liquid Chromatography
HSQC: Heteronuclear Single Quantum Coherence
IMAC: Immobilized Metal ion Affinity Chromatography
IVS: Intervening Sequence or Intrinsic Voltage Sensor
IPTG: Iso Propyl-β-D-1-ThioGalactopyranoside
NOE: Nuclear Overhauser Effect
OHC: Outer Hair Cell
PAT: Partial Anion-Transporter
SDS-PAGE: Sodium Dodecyl Sulphate-PolyAcrylamide Gel Electrophoresis
STAS: Sulphate Transporter Anti Sigma factor antagonist
SulP: Sulphate Permease
SUMO: Small Ubiquitin-like Modifier

SOMMARIO

Il presente lavoro è stato suddiviso in due parti: la prima concernente i domini STAS, mentre la seconda i domini Myb/SANT. Il *fil rouge* che unisce le due parti è costituito dalla metodica e dagli scopi, ovvero lo studio della struttura e delle interazioni di questi domini tramite NMR in soluzione.

La prima famiglia di domini trattati, gli STAS (Sulphate Transporter and Anti Sigma factor antagonist), è stata originariamente classificata sulla base della similarità con gli ASA (Anti Sigma factor Antagonist), domini batterici contenuti in proteine coinvolte in risposte contro gli stress ambientali e nell'induzione della sporulazione. Gli STAS sono presenti in piante, batteri, funghi e animali, in proteine con i ruoli più diversi, come sensori di luce, ossigeno, nucleotidi ciclici, trasportatori vari, ecc... Presentano una struttura con cinque β-sheets e quattro α-eliche, intercalati in maniera irregolare. Le strutture di domini STAS di trasportatori attualmente note (Protein Data Bank, Gennaio 2013) sono cinque: quattro batteriche, una di mammifero, nessuna di vegetale. L'unica struttura nota di mammifero corrisponde allo STAS di prestina di *Rattus norvegicus*, mancante della Intervening Sequence (IVS), una sequenza variabile da proteina a proteina della stessa famiglia: uno degli obiettivi consisteva nell'ottenere questo dominio completo. La succitata prestina fa parte della famiglia degli SLC26A, trasportatori anionici transmembrana con un ruolo nel mantenimento dell'equilibrio elettrolitico a livello degli epiteli. Prestina però è un membro anomalo di questa famiglia, in quanto, nei mammiferi, non funziona da trasportatore bensì da proteina motore nell'amplificazione cocleare del suono. Oltre all'interesse per lo studio del suo dominio completo dello STAS di prestina, l'attenzione è stata focalizzata anche su possibili partner di interazione. Uno di essi, ipotizzato sulla base sul modello di attivita di un trasportatore di ammonio in cui un dominio C-terminale interagisce con un loop citoplasmatico, poteva essere il cosiddetto motivo di Saier, costituito da una tripletta di amino acidi ripetuta abbastanza conservata. Il costrutto del motivo di Saier è stato modellato sui dati strutturali ottenuti sul trasportatore batterico BicA, un modello che si è dimostrato non applicabile per prestina, in quanto il costrutto, una volta clonato ed espresso, ha mostrato gravi problemi di solubilità in fase di purificazione, già a livello di proteina di fusione. Il dominio STAS su cui sono stati concentrati i maggiori sforzi è quello appartenente al trasportatore di solfato SULTR1;2 di *Arabidopsis thaliana*. Questo

proteina presenta una intervening sequence di soli 10 aminoacidi e attualmente non sono note strutture di STAS vegetali. Dopo una prova preliminare con un costrutto del dominio già disponibile in laboratorio, ne sono stati clonati e testati altri cinque che differivano tra loro nella lunghezza della prozione N-terminale della sequenza.. Non è stato tuttavia possibile ottenere un campione adatto per studi NMR ma è stato scelto il costrutto più promettente per ottimizzazione del tampone finale, sul quale nuovi tentativi sono in programma.

La seconda parte del progetto, svolta presso lo Structural Genomics Consortium (SGC) di Toronto, Canada, ha riguardato i domini Myb/SANT. Questi domini di circa 50 amino acidi presentano entrambi una struttura con tre eliche, dove la seconda e la terza elica (quelle più C-terminali) costituiscono un motivo Helix-Turn-Helix (HTH). Elementi addizionali, come ulteriori eliche o β-hairpins, possono essere presenti. Ciò che li differenzia è la distribuzione della carica superficiale: positiva sulla terza elica e negativa sulla prima elica per i Myb, invertita per i SANT. I Myb sono noti principalmente come domini che legano il DNA con la loro terza elica, anche se nuove evidenze dimostrano che la prima elica può interagire con le code basiche degli istoni; i SANT invece legano le code degli istoni con la terza elica. Le strutture di due domini di questa famiglia sono stati risolti e presentati in questo lavoro: la ripetizione R1 del DNA-binding domain della proteina hDmp1 e il dominio SANT2 di NCoR2. hDmp1 è un oncosoppressore che contiene un DNA-binding domain costituito da tre ripetizioni imperfette: è stata risolta la struttura della ripetizione più N-terminale, R1, che ha dimostrato la tipica struttura a tre eliche dei Myb e le stesse proprietà elettrostatiche di superficie. Tuttavia, questa porzione del dominio non si lega al DNA, di cui sono state testate sequenze note e altre nuove. Questo è in accordo con quanto è già noto sui DNA-binding domain di struttura simile, dove l'interazione con il DNA riguarda solo le ripetizioni R2 e R3, mentre R1 sembra avere una funzione accessoria. L'altro dominio risolto è il SANT2 di NCoR2, una proteina che agisce da repressore dei recettori nucleari in assenza di ligando. Questa proteina possiede due domini SANT: la struttura di quello situato nell'estremità più N-terminale è già stata risolta (PDB: 1XC5) e studiata, in particolare come partner di interazione dell'enzima istone deacetilasi 3 (HDAC3). Il dominio SANT2 ha dimostrato però proprietà peculiari e differenti dai SANT: oltre ad avere una lunga elica addizionale all'estremità C-terminale e una zona di conformational averaging tra questa elica e quella precedente, presenta proprietà elettrostatiche di superficie tipiche dei Myb, non dei SANT. Non sono note interazioni con il DNA, ma con le code istoniche dell'istone H4, con cui sono stati fatti degli esperimenti di binding monitorati tramite ^{15}N-HSQC. La regione interessata da questo legame però confina con una mutazione che è stata rilevata nel campione NMR in fase di assegnazione: è quindi in programma la produzione della proteina wild-type e la ripetizione degli esperimenti con le code di H4.

ABSTRACT

The present work is divided in two parts: the first one concerns the STAS domains, while the second one the Myb/SANT domains. The guiding thread linking the two parts is constituted by the method and the aims: these are the study of the structure and of the interactions of these domains by means of NMR techniques.

The first family, the STAS (Sulfate Transporter and Anti-Sigma factor antagonist) was originally discovered by similarity with the ASA (Anti Sigma factor Antagonist), bacterial domains contained in proteins involved in the environmental stress response and in the induction of sporulation. The STAS are present in plants, bacteria, fungi and animals, in proteins with different roles, like light/oxygen/cyclic nucleotides sensors, transporters of various kinds, etc… They present a structure with five β-sheets and four α-helices, interspersed in a irregular way. The STAS structures from transporters currently known (Protein Data Bank, January 2013) are currently five: four bacterial one, one of mammal, none vegetal. The only mammalian structure known corresponds to the STAS of prestin from *Rattus norvegicus*, missing the Intervening Sequence (IVS), a sequence largely variable among the same family of proteins: one of the aims was achieving the full domain. The aforementioned prestin belongs to the SLC26A family, transmembrane anion transporters with a role in maintaining the electrolytic balance in epithelia. However, prestin is an anomalous member of this family, since it does not work as a transporter but as the motor protein responsible of the cochlear sound amplification. Besides the interest in the study of the full STAS domain, the attention was focused also on possible interaction partner. One of them, based on the activity model of an ammonium transporter where its C-terminal domain interacts with a cytoplasmic loop, could have been the so called Saier motif, constituted by a repeated amino acid triplet with a certain degree of conservation. The construct for the Saier motif was modelled on structural data available for the bacterial transporter BicA, a model that demonstrated to be not applicable to prestin, since this construct, cloned and expressed, had severe problems of solubility during the purification, already as a fusion protein. The STAS domain were most of the efforts were focused on was the one belonging to the sulfate transporter SULTR1;2 from *Arabidopsis thaliana*. This domain presents a 10 amino acid intervening sequence and no STAS structures from plants are currently known. After a trial with a construct already available, five new ones were cloned and

tested. Anyway, it has not been possible obtaining a NMR sample but the most promising construct has been chosen for further attempts to optimize the final NMR buffer.

The second part of this thesis, carried out at the Structural Genomics Consortium (SGC) in Toronto, Canada, deals with the Myb/SANT domains. These 50 amino acid domains present a three helix structure, where the second and the third helix (the most C-terminal ones) constitute a Helix-Turn-Helix (HTH) motif. Additional elements, such as further helices or β-hairpins, could be present. The difference of this domains stays in the distribution of the superficial charge: positive on the third helix and negative on first one for the Myb, the opposite for the SANT. The Myb are known chiefly as DNA-binding domains with their third helix, even if new evidence support that the first helix can interact with the basic histone tails; the SANT instead bind the histone tails with their third helix. Two domains of this family were solved and presented in this work: the repeat R1 of the DNA-binding domain of hDmp1 and the SANT2 domain of NCoR2. hDmp1 is a tumour suppressor protein with a DNA-binding domain constituted by three imperfect repeats: the structure of the most N-terminal repeat, R1, was solved, which showed the typical three helix structure and the same electrostatic surface properties of the Myb domains. Anyway, this portion of the domain does not bind DNA, tested with known and new sequences. This result is in agreement with what is known about DNA-binding domains of similar structure, where the interaction with DNA concerns only the R2 and R3 repeat, while R1 seems to have a secondary function. The other domain solved is the SANT2 of the protein NCoR2, a protein acting as nuclear receptor repressor in the absence of ligand. This protein has two SANT domains: the structure of the most N-terminal one was already solved (PDB. 1XC5) and studied, in particular as interaction partner of the enzyme histone deacetylase 3 (HDAC3). Nevertheless, the SANT2 domain has shown particular properties: besides a long additional helix at the C-terminal and a region of conformational averaging between these helix and the previous one, it presents electrostatic surface properties typical of the Myb, not of the SANT domains. No DNA interactions are known, but with the H4 histone tails, used for binding experiments monitored by ^{15}N-HSQC. However, the concerned region lies in proximity of a mutation that was detected in the NMR sample during the assignment: the production of the wild-type protein and the repetition of the binding experiments have been planned.

1. INTRODUCTION

1.1 The STAS domain

Gram-positive bacteria, such as *Bacillus subtilis*, respond to unfavourable environment conditions in different ways, depending on the type of stress and on its intensity. Severe stress leads the bacteria to turn into quiescent and dehydrated forms, the spores. The process leading to the sporulation involves a cascade of transduction signals that starts with the sigma factor σ^F, which leads the bacterial RNA polymerase to the transcription of sporulation-specific genes. When its inactivation is required, σ^F is bound to the anti-σ SpoIIAB, which, in turn, is regulated by the anti-anti-σ or anti-σ antagonist SpoAA. In this way, the anti-sigma factors antagonists activate the bacterial RNA polymerase by inhibiting its anti-sigma factors (1). More frequently, when the intensity of the stressful event is lower, like small changes in environmental pH, temperature, ethanol, blue light, the effect is the activation of another subunit controlling the RNA polymerase, σ^B, which leads to the transcription of the stress genes. This sigma factor is released by RsbU, a phosphatase which is controlled by three proteins: RsbR, RsbS and RsbT, constituting a complex called the stressosome (2) (3). In both cases, some of the key proteins involved contain or exclusively consist in a particular domain called ASA (Anti Sigma factor antagonist), of approximately 150 amino acids. In fact, in the sporulation process, SpoAA is composed by a single ASA domain, like RsbS, while the RsbR is a multidomain protein containing one of these ASA domains (3).

This domain is not just present in the bacterial proteins as the aforementioned anti-sigma factor antagonists of σ^B and σ^F: unexpectedly, a significant sequence similarity between the ASA and the C-terminal domains of a kind of anion transporters present in bacteria, fungi, plants and animals, the SulP family, was found and the new domain was called STAS (Sulfate Transport and Anti Sigma factors antagonist) (4). Later on, the same domain was discovered also in the phototrasduction mechanism in the LOV-STAS domains of the protein YtvA in *Bacillus subtilis*, giving account to the response of σ^B and the stressosome to the blue light (5) (6).

The structural elements of STAS usually are 4 β-strands and 5 α-helices, with a similar folding. However, even if the fold is conserved, the sequence similarity can be low, in particular in the loop between α1 and β3: this segment is also the most evident difference with the bacterial ASA, where it is absent. In other organisms, this “intervening sequence” (IVS) or “variable loop” can reach up to 150 amino acids and it is predicted to be unstructured in solution . In the tract spanning between β3 and α2, there is a conserved loop, where no insertion is admitted, probably suggesting that this a functionally important region (4).

The STAS is generally considered as a protein interaction domain, but it covers also different functions, such as nucleotidase/nucleotide-binding transduction units in sensor proteins, or it is important for the correct membrane targeting and activity of the sulphate transporter Sultr1;2 in Arabidopsis thaliana (7). Although little is known about STAS activity, it seems that phosphorilation can play a role in its modulation: the Ser57 in SpoIIAA is phosphorilated by the protein kinase SpoIIAB and the substitution of the equivalent residue (Thr587) in the STAS domain of the Sultr1;2 abrogates its sulphate transport activity. Interestingly, the phosporilable residues are localized in the conserved loop (8). Another possible mechanism of regulation involves the interaction of the STAS domain with a cytoplasmatic loop belonging to the same protein: this kind of interaction between a cytoplasmic loop and a C-terminal extention has already been demonstrated for the AMT ammonium transporter in *Arabidopsis thaliana* (9), so it is not excluded that it could occur also in the STAS-containing SulP transporters. However, the complete set of functions, interactions and regulation of the STAS domains is yet to be defined. In figure 1 some of the STAS structures solved are reported.

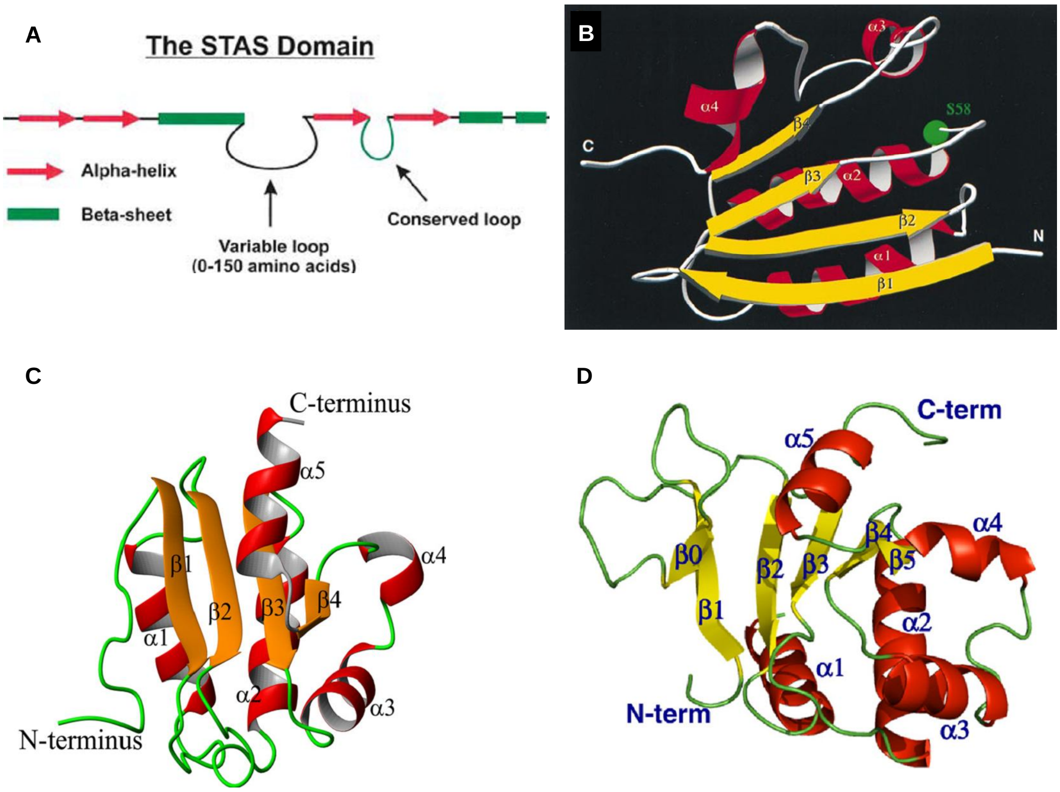


Figure 1: (A) rappresentation of the general sequential organization of the STAS domain, from ref. (6); **(B)** solution structure of spoIIAA (10) (PDB: 1AUZ); **(C)** solution structure of the STAS domain of Rv1739c from *Mycobacterium tuberculosis* (11) (PDB: 2KLN); **(D)** crystal structure of the STAS domain of the prestin (SLC26A5) from *Rattus norvegicus*, without the intervening sequence (12) (PDB: 3LLO).

1.2 The SLC26 transporter family: structure, functions and pathology

The balance of ions in the epithelial tissues is of critical importance in the homeostasis: it is fundamental then that the transport system is efficient and that any imbalance can be buffered properly (13). The disruption of this equilibrium can lead to diseases, such as cystic fibrosis and congenital chloride diarrhoea (14). The maintenance of this electrolytic balance is supported by a system of transporters with different affinity for various ions: one of this family is the SLC26 transporters, a subfamily of the wider SulP (sulfate permease) proteins, encompassing bacteria, plants and animals (13).

In humans, the SLC26 family has 11 member, with SLC26A10 being considered a pseudogene. These transporter can be classified in three groups, based on the different transport activity: sulphate transporters (SLC26A1, SLC26A2), coupled Cl^- /$HCO3^-$ exchangers (SLC26A3, SLC26A4, SLC26A6) and ion channel (SLC26A7, SLC26A9). The transport activity of SLC26A8 and SLC26A11 has not been elucidated and SLC26A5 does not seem to work as an anion transporter in mammals or has a weak activity (13). The SLC26 transporters differ not only for the activity, but also for the tissue distribution and for their involvement in diseases The details about each human SLC26 transporter are reported in table 1.

The SLC26 family members are transmembrane proteins of 700-1000 amino acids, sharing a sequence identity of 21-43% and working as dimers (15).The central part of the proteins consists in the hydrophobic transmembrane α-helices, where the anion binding sites and the pores are localized. The exact number of membrane spanning segments is still debated, but most likely there are between 10 and 14 helices (16). It is also probable that the anion binding sites can be in proximity of helices that breaks and turn in the membrane, exposing backbone amide and carboxyl groups in a hydrophobic environment (17). The most conserved regions of the SLC26 transporters are the transmembrane hydrophobic elices, encompassing also the cytosolic N-terminal “sulfate transport” consensus part. A second portion of conserved residues is toward the C-terminal end of the transmembrane region and it is characterized by the presence of the almost invariant triplet NQE, as observed by Saier (18). At the cytosolic C-terminal end there is a STAS domain (16), in some transporters followed by a PDZ-binding motif (19).

STAS functions are not well understood, but evidence supports the speculation that it can be a protein-protein interaction domain in the SLC26 transporters. In fact, it has been demonstrated that the STAS domain of SLC26A8 can interact with MgcRacGAP (20), the STAS domains of SLC26A3 binds to the R domain of CFTR (21) and the mouse Slc26a6 STAS to the carbonic anydrase II (22).

Table 1: Features of the member of the SLC26A family in human; adapted from (14).

Gene name	**Protein name**	**Known substrates**	**Tissue distribution**	**Disease associations**	**Gene locus**	**GenBank accession ID**
SLC26A1	Sat-1	SO_4^{2-}, oxalate, Cl^-	Kidney, liver	Unknown	4p16.3	AF297659
SLC26A2	DTDST	SO_4^{2-}, Cl^-	Multiple tissues	Chondrodysplasias	5q31-34	U14528
SLC26A3	DRA, CLD	SO_4^{2-}, Cl^-, HCO_3^-, OH^-, oxalate	Intestine, pancreas, sweat glands, prostate	Congenital chloride diarrhea	7q31	L02785
SLC26A4	Pendrin	Cl^-, HCO_3^-, I^-, formate, fructose, mannose	Inner ear, thyroid, kidney, brain	Pendred syndrome, Non-syndromic deafness (DFNB4)	7q31	AF030880
SLC26A5	Prestin	?	Outer hair cells of cochlea	Non-syndromic hearing loss	7q22	AC005064
SLC26A6	CFEX, PAT-1	SO_4^{2-}, Cl^-, HCO_3^-, OH^-, oxalate, formate	Multiple tissues	Unknown	3p21.3	AF279265
SLC26A7	None	SO_4^{2-}, Cl^-, oxalate	Kidney	Unknown	8q23	AF331521
SLC26A8	Tat1	SO_4^{2-}, Cl^-, oxalate	Brain, sperm	Unknown	6p21	AF331522
SLC26A9	None	SO_4^{2-}, Cl^-, oxalate	Lung	Unknown	1q31-32	AF331525
(SLC26A10)	None	Pseudogene	Brain	Unknown	12q13	NM_133489
SLC26A11	None	SO_4^{2-}	Multiple tissues	Unknown	17q25	NM_173626

1.3 A divergent SLC26 protein, prestin, is responsible for the cochlear amplification in mammals

The sense of hearing is mediated by different organs in the mammalian ear. In the organ of Corti, in the cochlea, two kind of cells take part in the hearing: the inner hair cells (IHCs) perceive the sound and transmit the sensory information to the brain, while the outer hair cells (OHCs) are motile cells, with particular stereocilia (hair bundles) whose stimulation by the sound waves triggers the opening and closing of ion channels in the membrane (23) (figure 2). The change in the membrane potential due to the ions passage has a dramatic effect on the shape of OHCs, which are contracted in response to depolarization or elongated because of hyperpolarization: this conformational change occurs at microsecond rates, leading to a sound amplification of thousand of times (24). Another interesting feature of OHC is the “Non Linear Capacitance”, which corresponds to the ability of separating charges at the two opposite surfaces of a dielectric (a non conductor) when they are maintained at different voltages, but induced in a voltage-dependent way (25).

At the basis of this change of cell length there is the protein SLC26A5, known as prestin. It was identified by subtractive cloning in the search of a protein not present in IHCs and that could give account for the electromotility of OHCs: it resulted as a 80 KDa transmembrane protein with unique properties (26). Prestin shares just a weak transport activity with the SLC26A family (27), but it has the special capacity of working as a “motor protein”, coupling the transmembrane voltage to the OHC electromotility, without the need of energy supply such as ATP hydrolysis (28). It is interesting to note that prestin is still an electrogenic antiporter in non mammalian vertebrates such as zebrafish and chicken, where it exchanges oxalate or sulfate for chloride with 1:1 stoichiometry but it does not produce any electromotility (29).

The mechanism underlying the capacity of prestin of sensing the “gating current” generated by the charge movement is still debated between two theories. The first hypothesis considers the intracellular anions as the real voltage sensors of prestin and it is called “the partial anion transporter model” (30). In fact, mutagenesis experiments concerning substitutions of non conserved charged amino acids in the putative membrane domain with other charged or neutral amino acids did not abolish the electrical signature of prestin, although in some mutants the non linear capacitance was altered. On the other hand, the removal of Cl^- from the cytoplasm caused a loss of non linear capacitance, while it was preserved with all the monovalent anions tested in patch-clamp experiments (with a relative preference of: $I^- \approx Br^- > NO_3^- > Cl^- > HCO_3^- > F^-$). However, the only anions supposed to be present at millimolar concentration in the cytoplasm are Cl^- and HCO_3^- and so these are the natural candidate as the extrinsic voltage sensors. The proposed mechanism starts with the anion binding to a site and its subsequent translocation through the membrane by the transmembrane voltage. The direction of the translocation is toward the extracellular surface or the cytoplasm in response to hyperpolarization or depolarization respectively. This movement of the anion across the membrane has an effect on the surface area that triggers a conformational change with morphological consequences at the cellular level: the OHC surface area decreases and the cell is contracted when the anion is close to the cytoplasmic side, while the shift of the anion toward the extracellular surface causes cell elongation due to an increase in the cell surface area.

A different theory about prestin's mechanism ("the intrinsic voltage sensor model") states that the voltage sensor is not constituted by the intracellular anions, but from prestin itself (31). In particular, it was suggested that there is a charged portion of prestin that traslocates upon changes of membrane polarization, inducing a conformational change in the protein. The purported segment of prestin responsible for the anion sensing was mapped by mutagenesis in a putative transmembrane helix, in a portion of the protein which shares a high similarity with SLC26A6, the closest family member that does not show any voltage sensitivity. According to this theory, the anions act as allosteric modulators on prestin's structure. The figure 3 illustrates the two different theories.

To propagate the conformational change of prestin to the OHC, the protein must be present in the cell lateral membrane at a high density, up to 10 millions per cell (32) and may interact with proteins involved in maintaining the cell structure: as a matter of fact, one of its binding partner is the Microtubule-associated Protein MAP1S. In particular, the interaction concerns the tract between the heavy and the light chain of MAP1S and part of the STAS domain of prestin (the first three β-strands, the first α-helix and the IVS) (33).

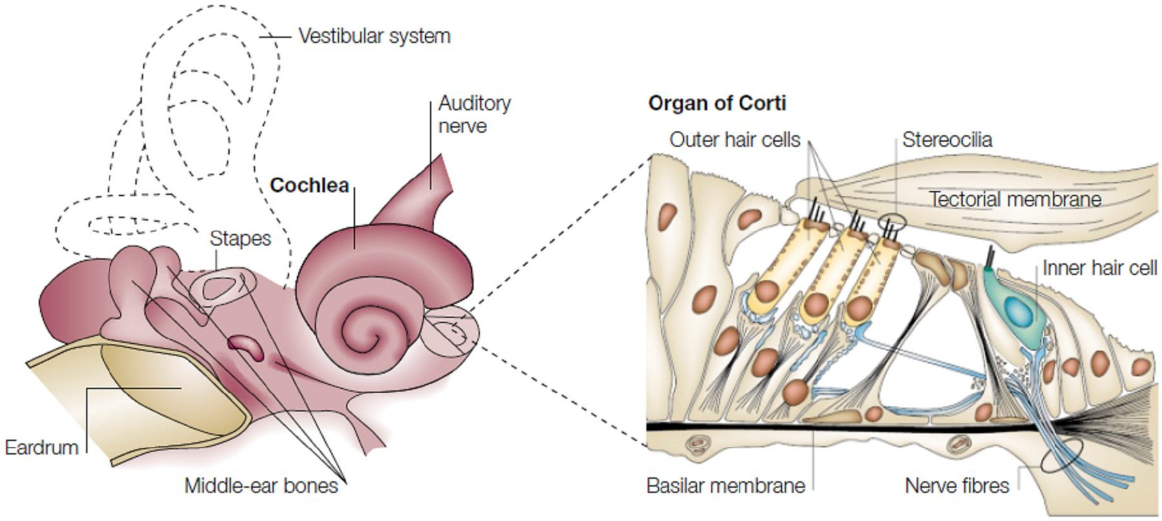


Figure 2: Anatomical representation of the cochlea and of its different cellular components (34).

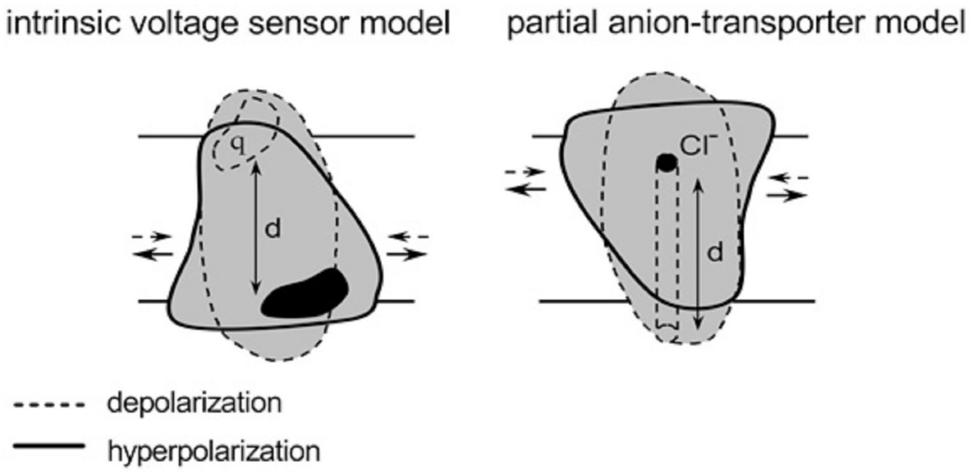


Figure 3: Illustration of the two different theories about prestin electromotricity (31). The depolarization of the membrane corresponds to contracted state of prestin, while the hyperpolarization to an expansion of the protein. The membrane cycle between hyperpolarization and depolarization is caused by the traslocation of a prestin charged portion *q* in the intrinsic voltage sensor model (IVS) (left) or anion-binding site (for Cl^- or bicarbonate) in the partial anion-transporter model (PAT) (right).

1.4 The sulfate transporter SULTR1;2 and its STAS domain in *Arabidopsis thaliana*

Sulfur in plants is necessary for the production of compounds like the amino acids methionine and cysteine, glutathione, vitamins and cofactors such as thiamine, coenzyme A and biotin (35). The first step is the acquisition of sulphur from the soil as sulfate (SO_4^{2-}), followed by the transport through other cell membranes. In the cells, the sulphate is then stored in vacuoles or metabolized in chloroplasts and plastids. The shift of sulphur from one plant district to an other is controlled by sulphate transporters (36) (figure 4).

The sulfate uptake is carried out by members of the transporter SLC26 family. In *Arabidopsis thaliana*, there are 14 sulfate transporters, 12 of wich containing a STAS domain, grouped in 5 categories based on the sequence similarity (37). The family members involved in the first step of the sulfur metabolism are two high affinity transporters, SULTR1;1 and SULTR1;2, expressed in roots hair and epidermal and cortical cells (38). This two transporters present a certain degree of functional and spatial distribution overlap but SULTR1;2 is prevalent over SULTR1;1, providing around 80% of the sulphate uptake. SULTR1;1 has a lower K_m value for the uptake of trace amounts of sulphate and it seems to gain more functional importance in case of sulphate deprivation, which triggers an increase in its expression, while the expression of SULTR1;2, which is expressed constitutively, is just slightly boosted in these conditions (39) (figure 5). They are closely related, with an amino acid sequence identity of 69%.

From the structural point of view, SULTR1;2 is predicted to have 12 transmembrane helices and the STAS domain is located in its cytosolic C-terminal end. STAS seems to have some fundamental roles for SULTR1;2 activity: it is important for the localization of the transporter to the plasma membrane and for the regulation of its transport activity (7) (8). In particular, the swapping of the STAS domain between different transporters affects the kinetic of the transport, suggesting that STAS performs transporter-specific tasks (7). Based on comparison with other transporters, it is speculated that STAS can do inter- and intra-molecular interactions, with proteins, small molecules and cytosolic portions of the transporter itself, regulating the transporter activity. For the STAS of SULTR1;2, a direct interaction with a enzyme involved in sulfur metabolism, the cysteine synthase (O-acetylserine (thiol)lyase or OASTL) was demonstrated with the two hybrid assay (40). The binding has a negative effect on SULTR1;2 activity but positive on OASTL in *in vitro* experiments; OASTL can bind also SULTR1;1 STAS in the two hybrid system experiment but their binding has not any influence on OASTL activity (40).

The STAS domain present also a particular feature in the sequence : a couple of adjacent cysteines (Cys645 and Cys 646). The Cys 645 is conserved in many of the *Arabidopsis* sulphate transporters and replaced by a hydrophobic amino acid in the other tansporters, while the second cysteine is present only in SULTR1;2, substituted by polar amino acids in other sulfate transporters (8). Site directed mutagenesis experiments on these two residues demonstrated that they are not essential for the sulfate uptake but they are necessary for the optimal activity of SULTR1;2 (8). Another interesting residue is also the threonine-587, which corresponds to the serine-58 of SpoIIAA, that undergoes phosphorilation: it is not clear yet if Thr-587 is phosphorilated too, but its deletion or substitution abrogates completely the sulphate uptake by SULTR1;2 (8).

A better understanding of the mechanism of sulfur uptake by plants can be useful in finding solutions for agriculture in sulfur-deficient territories (41) (42)

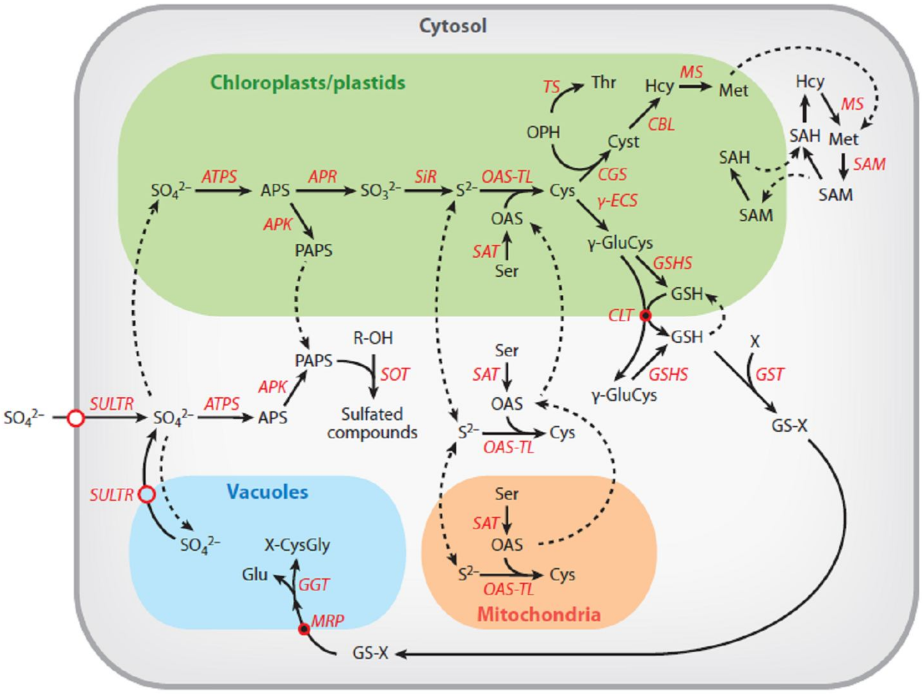


Figure 4: Sulfur cycle in *Arabidopsis thaliana* (36).

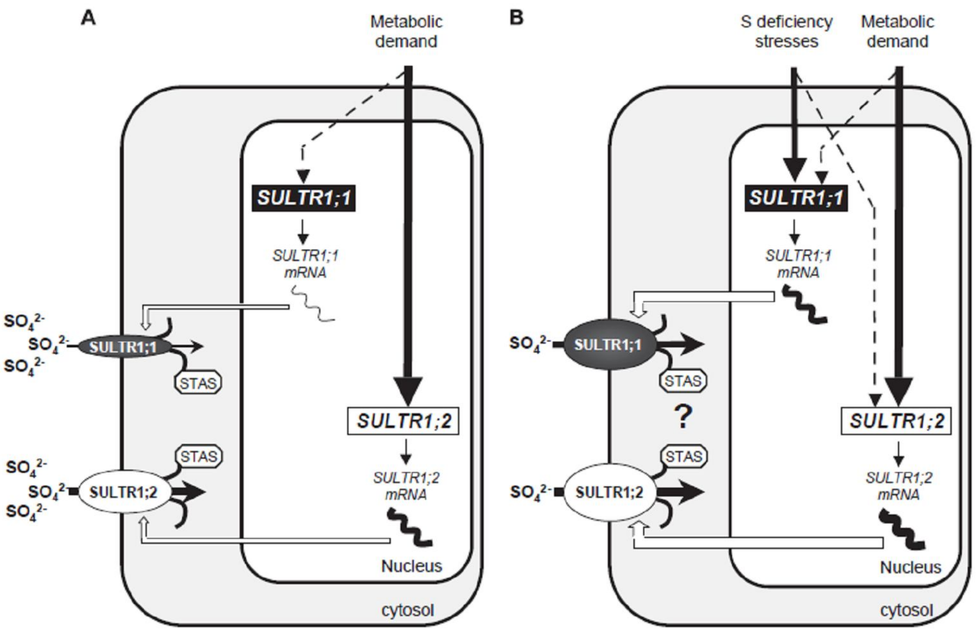


Figure 5: (A) Activity of SULTR1;1 and SULTR1;2 in normal conditions and **(B)** under sulfur deprivation (43).

2. AIMS

During the first part of this PhD, the project concerned the study of selected STAS domains with the objectives described in the following.

Achieving the complete structure of the prestin STAS from *Rattus norvegicus*, with IVS

Referring to a previous PhD project, where the structure of an engineered STAS domain of prestin from Rattus norvegicus was solved by crystallography and studied also by NMR (12), we wanted to go further trying to achieve the full domain structure. The crystal structure misses the intervening sequence, originating a general model of the mammalian STAS. The consequence of this deletion is a wider angle between α1 and α2, compared with the STAS of YchM and Rv1739c.

The interest in the full structure was not only related in obtaining the STAS of prestin and not of an engineered prototypical version, but also in achieving a structure with the intervening sequence, useful for possible interaction studies, specifically for prestin. The role of the IVS is still unknown and it is supposed to be unstructured. Currently, pathologic mutations in that region have been detected just in pendrin (6).

Obtaining a portion of rat prestin, the Saier domain, as a possible intramolecular interaction partner of STAS

The investigation was opened also for possible inter- and intra-molecular interaction partners. In particular, a tract of the SLC26A family is conserved and in some predictions it is proposed as cytosolic or partially cytosolic (34) (44). This protion is called “the Saier motif” and it is characterized by the presence of the triplet -NQE- (18). In some transporters, an interaction between a C-terminal domain and a cytosolic loop has been demonstrated, as a regulatory mechanism (9). We investigated then if the Saier motif could act as cytosolic loop interacting with the STAS of the same protein (rat prestin). Since no structure of SLC26A transporter is available, we based the design of the construct on the topological studies reported in ref. (45), about the bacterial SulP transporter BicA, where the Saier motif is cytosolic (figure 6).

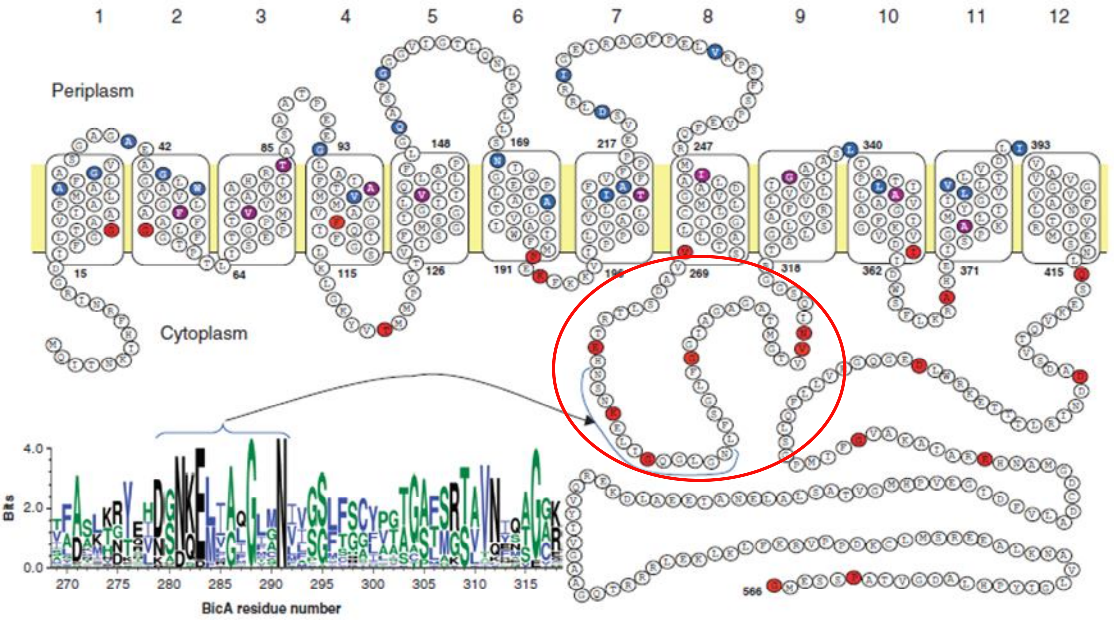


Figure 6: Topology of the bicarbonate transporter BicA from Synechoccus PCC7002 (45). The Saier motif is circled with a red line.

Obtaining a suitable NMR sample of the STAS domain of SULTR1;2 from *Arabidopsis thaliana* for structure determination

A STAS domain from the vegetal reign was also considered: the one from the sulfate transporter SULTR1;2 form *Arabidopsis thaliana*. In fact, the majority of the STAS structures solved are bacterial for the greatest majority, with one from mammals and none from plants. Moreover, this STAS has the IVS, which is not present in bacteria and was deleted in the mammal one, but it is relatively short (10 amino acids, based on the comparison with spoIIAA from *Bacillus sphaericus*). In the case that the small IVS had not a too deleterious effect on NMR spectra (line broadening for instance), there could be not the possibility of solving a full STAS with the IVS.

3. MATERIAL AND METHODS

3.1 Expression and purification of the STAS domain of prestin from *Rattus norvegicus* with the intervening sequence

100 μl of a frozen aliquote of *Escherichia coli* BL21(DE3) were cultivated in 3 ml of LB medium for 5 h and then transferred in 100 ml of minimal medium M9 for an overnight culture, at 37°C, 180 rpm. The overnight culture was diluted to 1 l in M9 and then divided in three flasks, to test different induction OD_{600nm}: 0.6, 0.8 and 1. The induction was carried out with 1 mM IPTG for 4 h. The bacteria were harvested by centrifugation at 4°C, 10000 rpm, 20 min. The pellet was solubilised in 8 ml of the binding buffer (see above) with the protease inhibitors cocktail and lysed by sonication (9 s pulse, 9 s pause, for 10 min). The cell debris were removed by centrifugation at at 4°C, 14000rpm, 20 min and the final volume reached 16 ml with the binding buffer. The sample was loaded in a HisTrap column (GE Healthcare) with a peristaltic pump, and then eluted with an AKTA-purifier FPLC system at 150 mM imidazole (20 mM NaH2PO4/Na2HPO4, 500 mM NaCl, 10 mM β-mercaptoethanol, 150 mM imidazole, pH 8). The buffer was exchanged with one with the same composition but without imidazole. The cleavage was carried out overnight, with the Ulp protease (about 10 units/mg of estimated protein amount), at 4°C, without any stirring. The cleaved protein was purified from the SUMO-histag fragment and the Ulp protease with a second affinity chromatography. The protein was concentrated and the buffer was changed to one suitable for NMR spectroscopy: 32 mM KH^2PO^4/K^2HPO^4, 50 mM NaCl, 10 mM DTT, 1 mM EDTA, 10% D2O.

The same expression and purification was performed with the bacterial strain *Rosetta* of *Escherichia coli*, with induction at OD 0.8.

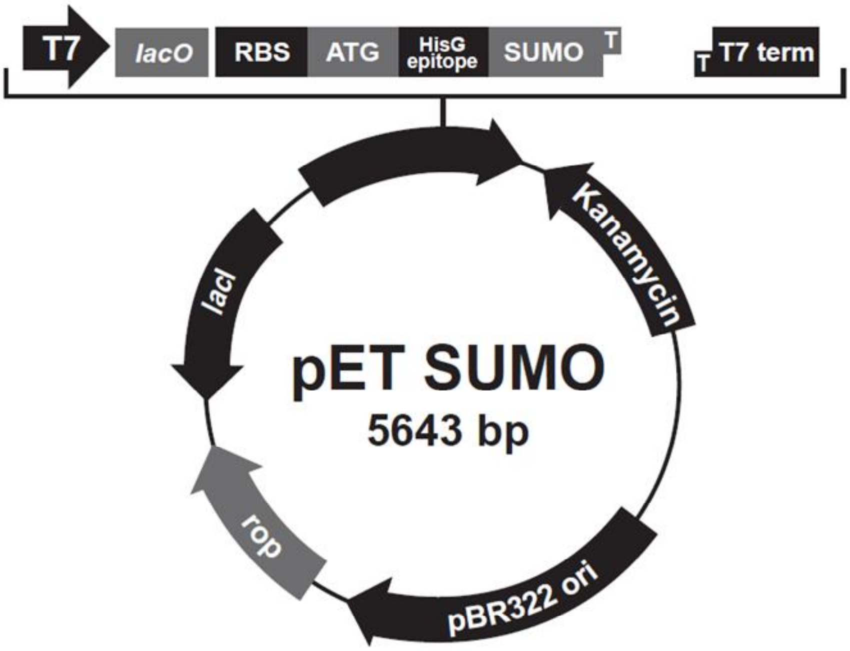
3.2 Cloning of the Saier motif of prestin from *Rattus norvegicus*

The construct for the expression of the Saier motif of prestin from *Rattus norvegicus* (residues 359-409) was constructed in the Champion^{TM} pET SUMO Protein Expression System by Invitrogen Life Technologies® (figure 7), with the SUMO protein at the N-terminal and the kanamicin resistance. The primer set for the Saier motif was: forward 5'-AAAACCTTGGCAAATAAGCATGG-3', reverse 5'-CTATTACTTCCCTCCAGTTCCTTC-3'., which was amplified from the rat prestin cDNA with 30 PCR cycles organized as: 45 s at 95°C, 40 s at 58 °C and 25 s at 72°C, preceded by 5 min at 95°C to activate the polymerase only in the first step and followed by 20 s at 72°C to add the polyadenilation tails and then 4°C in the end. The pET SUMO plasmid was provided already linearized with blunt ends by the supplier and the PCR products were ligated with the T4 DNA ligase.

Bacteria *Escherichia coli One shot* by Invitrogen Life Technologies® were transformed with the ligation reaction by heat shock. The positive clones were selected on a plate with Luria Broth agar with 50 μg/ml of kanamicin. The plasmids of six different positive clones were purified by miniprep (NucleoSpin®) and the correct direction of the insertion was checked by agarose gel, after digestion with the restriction enzymes Xba and SacI. The only plasmid that gave the two gel bands of the expected length was sequenced and found without mutations. The purified plasmid was used to transform the bacteria *Escherichia coli Mac One* by Invitrogen Life Technologies® by heat shock. These bacteria are optimized for plasmid propagation. The bacteria were cultivated overnight at 37°C, 170 rpm in a thermostated orbital shaker and stored in aliquots of 1.2 ml with the 30% of glycerol and 50 μg/ml of kanamicin in LB at -80°C, preceded by freezing with liquid nitrogen.

The bacteria *Escherichia coli BL21(DE3)*, suitable for protein expression, were transformed with the plasmid of the only positive clone by heat shock, cultivated for 1 h at 37°C, 200 rpm and then stored at 4°C for immediate cultivation for protein expression or -20°C for long term storage.

Figure 7: Plasmid map of the Champion^{TM} pET SUMO Protein Expression System by Invitrogen Life Technologies®.



3.3 Expression and purification of the Saier motif of prestin from *Rattus norvegicus*

A colony from the plate of *Escherichia coli BL21(DE3)* were cultivate overnight in 20 ml of LB. 10 ml of culture were transferred to 500 ml of LB medium and incubated at 30°C, 200 rpm, for 4 h. When the bacteria growth reached an optical density at 600 nm of 0.6, the protein expression was induced with 1 mM IPTG and the culture was transferred at 30°C, to decrease the formation of the inclusion bodies. After 6 h, the bacteria were harvested by centrifugation at 4°C, 10 min, 7000 rpm and the pellet was stored at -20°C.

The lysis was performed in 20 ml of a affinity chromatography buffer with 10 mM imidazole, called “binding buffer” (20 mM NaH_2PO_4/Na_2HPO_4, 500 mM NaCl, 10 mM β-mercaptoethanol, 10 mM imidazole, pH7.4) with a cocktail of protease inhibitors (mini complete protease inhibitors by Roche®). The bacteria were lysed by sonication and the cell debris were spun down by centrifugation, at 4°C, 14000 rpm, 20 min or by three cycles of French Press. The final volume was doubled with the binding buffer and it was loaded slowly with a peristaltic pump in a 1 ml affinity column HisTrap (GE Healthcare), functionalized with Nickel-NTA resin. The elution was carried out by connecting the column with the AKTA purifier – FPLC system (GE Healthcare) and monitored at 280 nm. The protein eluted at 150 mM imidazole (20 mM NaH2PO4/Na2HPO4, 500 mM NaCl, 10 mM β-mercaptoethanol, 150 mM imidazole, pH7.4), but it precipitated just after the elution and it was not possible to proceed with the purification protocol. The expression and purification were attempted twice: in both the cases, the fusion protein precipitated just after the first IMAC chromatography.

3.4 Expression and purification of the construct 517-653 of the STAS domain of SULTR1;2 from *Arabidopsis thaliana*

This construct, cloned previously, was expressed at 37°C for 4 h, 170 rpm in Minimal Medium M9 with $^{15}NH_4Cl$ in a thermostated shaker. The purification was carried out as described for the STAS of prestin, with the AKTA purifier system.

3.5 Cloning of five new constructs of the STAS domain of SULTR1;2 from *Arabidopsis thaliana*

For the STAS domain of SULTR1;2 from *Arabidopsis thaliana*, 5 constructs were designed, as reported in table 2 and 3:

Protein	Number of aa	Molecular weight (Da)	pI
522-653+SUMO	252	28364.1	5.54
522-653	133	14966.1	5.11
525-653+SUMO	249	27975.7	5.54
525-653	130	14577.7	5.11
527-653+SUMO	247	27775.5	5.63
527-653	128	14377.5	5.27
531-563+SUMO	243	27347.0	5.63
531-653	124	13949.0	5.27
525-653+TrxA	264	28948.0	5.31
525-653+GAM*	132	14750.0	5.11

Table 2: List of the constructs prepared for the STAS domain of SULTR1;2 of *Arabidopsis thaliana*. The features of each construct are reported as fusion protein, followed by the cleaved protein. * The construct made in pETM20 is fused with the tioredoxin A (TrxA) and after the cleavage with the TEV protease the aminoacids glycine, alanine and methionine (GAM) are left at the N-terminal end of the protein. Sequences based on the Uniprot file Q9MAX3 for SULTR1;2.

Protein	Forward and reverse primers 5’-3’	Tm
522-653+SUMO	AGTCAGTATCCTGAAGCCACTA TGCTCCAGTAACGCCGGCG	64°C
525-653+SUMO	AGCGAAGCCACTATGGTTCC TGCTCCAGTAACGCCGGCG	64°C
527-653+SUMO	AGTACTATGGTTCCAGGGGTTC TGCTCCAGTAACGCCGGCG	66°C
531-563+SUMO	AGCGGAGTTCTTACTATTCGTG TGCTCCAGTAACGCCGGCG	64°C
525-653+TrxA	ACACCATGGAAGCCACTATGGTTCC ACACTCGAGTCAGACCTCGTTGGAGA	64°C

Table 3: List of the primers prepared for the STAS domain of SULTR1;2 of *Arabidopsis thaliana*.

For the four constructs prepared with the SUMO expression system, the procedure was the same as described for the cloning of the Saier motif.

The STAS construct 525-653 was prepared also in the pETM20 vector, a non patented plasmid provided by the EMBL, Heidelberg, Germany (figure 8).

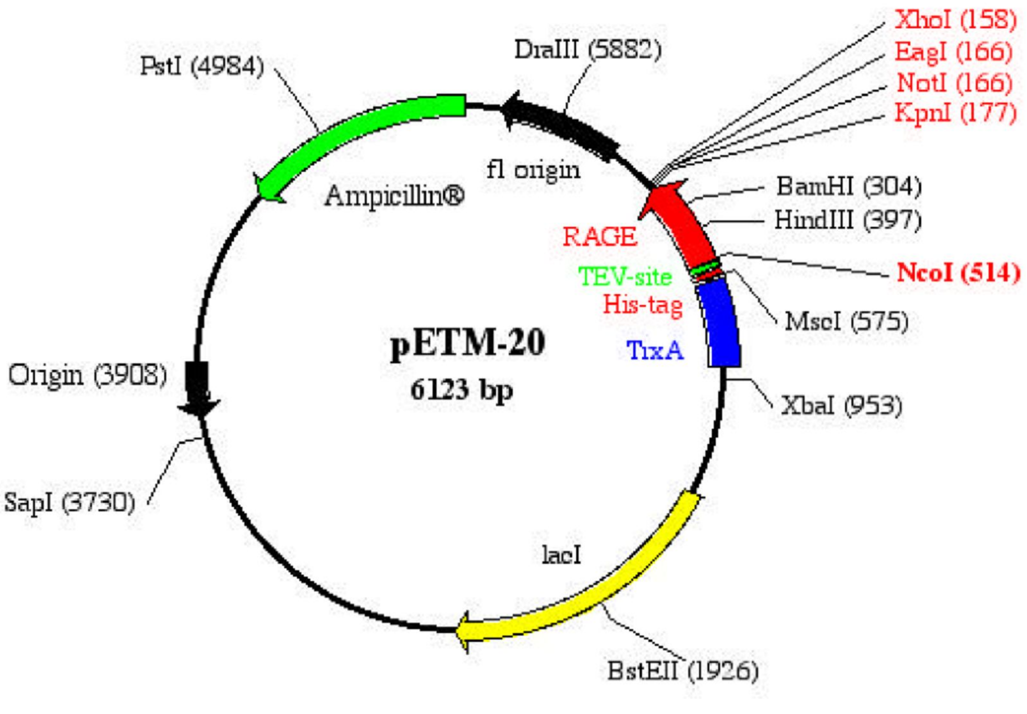


Figure 8: Plasmid map of the vector pETM20.

The plasmid and the PCR products were digested with the restriction enzymes NCoI and XhoI, for 2 h 30 min at 37°C, with 20 min at 65°C in the end to inactivate the enzymes. The digested PCR product and plasmid were separated in a gel and the correct gel bands were cut and purified with the PureLink® Quick Gel Extraction Kit by Invitrogen^{TM}. The plasmid and the PCR product were ligated with the T4 ligase, over night at 10°C. The bacteria *Escherichia coli Mac One* were transformed with the ligation reaction by heat shock and plated on LB agar; four colonies were cultivate in 3 ml of liquid LB, overnight. The plasmids were purified with a miniprep kit and digested with the restriction enzymes BamHI and XbaI to verify the correct insertion, which generated two fragments of 727 and 5433 bp. The correct clones were sequenced and a correct one was used to transform the bacteria *Escherichia coli Mac One* by heat shock. The bacteria were cultivated overnight at 37°C, 170 rpm in a thermostated orbital shaker and stored in aliquots of 1.2 ml with 30% of glycerol and 50 μg/ml of kanamicin in LB at -80°C, preceded by freezing with liquid nitrogen.

For protein expression, the bacteria *Escherichia coli BL21(DE3)* were transformed with the plasmid of the only positive clone by heat shock, cultivated for 1 h at 37°C, 200 rpm and then stored at 4°C for immediate use or -20°C for long term storage.

3.6 Expression and purification of the five new constructs of the STAS domain of SULTR1;2 from *Arabidopsis thaliana*

Preliminary expression trials were performed for each construct, to test two different expression temperatures. 100 μl of a frozen aliquote of Escherichia coli BL21(DE3) transformed with the different plasmids were cultivate in 3 ml of LB overnight and then divided and transferred in 100 ml of minimal medim M9 at 37°C, 180 rpm, for the growth step. Two flasks were prepared for each construct. At OD 0.6, the temperature was kept at 37°C for one flask and lowered to 20°C for the other flask, and then 1 mM IPTG was added for the induction. A sample of each culture was collect every hour, for 4 h. The outcome of the expression trial was checked by SDS-PAGE (with 16% gels).

For the constructs 522-653, 525-653 and 527-653, fused with the SUMO protein, and the construct 525-653 in the pETM20 plasmid, the protein expression and purification was carried out as described for the STAS domain of *Rattus* *norvegicus*, with the AKTA FPLC (GE Healthcare®). The protein expression was induced at 0.8 OD, at 25°C, overnight. For the cleavage with the TEV protease of the construct 525-635, fused with the Thioredoxin A, the recommended imidazole concentration is 150 mM. The NMR buffer was: 32 mM KH^2PO^4/K^2HPO^4, 50 mM NaCl, 10 mM DTT, 1 mM EDTA, 10% D2O.

The construct 525-653 in pETM20 was further used for the optimization of the purification buffer. Two flasks of 1 L of bacterial colture each were prepared. The protein expression and lysis were performed as described previously, with the sole exception of the induction carried out at 20°C instead of 25°C. The purification was carried out with NTA-nickel beads (Qiagen®) in plastic columns and elution for gravity, instead of the AKTA FPLC. The product of one liter of bacterial colture was purified with the same buffers used for the AKTA FPLC, while for the other liter the NaCl in the buffers was completely substituted by 50 mM arginine and 50 mM glutammic acid (46). This sample was used to test different NMR buffers.

1. 50 mM phosphate, 400 mM NaCl, 10 mM DTT, EDTA 1 mM, 0.05% NaN_3, pH 6.5, 10% D_2O
2. 50 mM phosphate, 50 mM L-Arg, 50 mM L-Glu, 10 mM DTT, EDTA 1 mM, 0.05% NaN_3, pH 6.5, 10% D_2O
3. 50 mM HEPES/NaOH, 50 mM L-Arg, 50 mM L-Glu, 10 mM DTT, EDTA 1 mM, 0.05% NaN_3, pH 7, 10% D_2O
4. 50 mM acetate, 400 mM NaCl, 10 mM DTT, EDTA 1 mM, 0.05% NaN_3, pH 5, 10% D_2O
5. 50 mM phosphate, 50 mM L-Arg, 50 mM L-Glu, 2% glucose, 10 mM DTT, EDTA 1 mM, 0.05% NaN_3, pH 6.5, 10% D_2O

4. RESULTS AND DISCUSSION

4.1 Results of the expression and purification of the STAS domain from prestin of *Rattus norvegicus*

The STAS domain from the rat prestin 505-727 (histag-SUMO-STAS: Mw: 38002,9; 342 aa; STAS: 24678,0; 223 aa) was first expressed in the *Escherichia coli* strain BL21(DE3). The achieved yield was unsatisfactory (figure 9), while for the same construct missing the intervening sequence the yield was good (12). The intervening sequence then was supposed to hamper somehow the protein expression: excluding that the presence of the sequence itself is badly tolerated or degradation-prone by the bacteria, we found that the intervening sequence is rich of rare codons (figure 10). This is a problem that generally occurs when an eucaryotic protein is expressed in a prokaryote, where the codons usage is different.

To overcome this problem, we exploited the bacterial strain *Rosetta*, which is optimized for 7 rare codons. Anyway, this was not sufficient to increase the yield to achieve a NMR sample (figure 11).

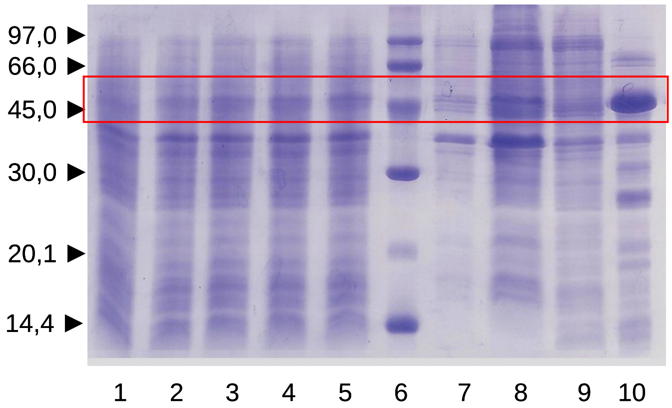


Figure 9: SDS-PAGE, gel15%. *E. coli* BL21(DE3). Loading order: (1) not induced expression; (2) 1h expression; (3) 2h expression; (4) 3h expression; (5) 4h expression; (6) Mw marker (KDa); (7) pellet; (8) supernatant; (9); flowthrough of the IMAC loading; (10): protein eluted at 150mM imidazole from IMAC

```
SPSYTVLGQLPDTDVYIDIDAYEEVKEIPGIKIFQINAPIYYANSDLYSSALKRKTGVNPAI
IMGARRKAMRKYAKEVGNANIANATVVKVDAEVDGENATKPEEEDDEVKFPPIVIKTTFPEE
LQRFLPQGENIHTVILDFTQVNFMDSVGVKTLAGIVKEYGDVGIYVYLAGCSAQVVNDLTSN
RFFENPALKELLFHSIHDAVLGSQVREA
```

```
agt  ccg  agc  tac  aca  gtc  ctg  gga  cag  ctc  cct  gac  acc  gat  gta
tac  att  gac  att  gat  gcc  tat  gag  gag  gtg  aaa  gaa  att  cct  gga
ATA  aaa  ATA  ttc  caa  ATA  aat  gcc  cca  att  tac  tat  gca  aat  agc
gac  ttg  tac  agc  agc  gca  ctg  aaa  AGA  aag  act  ggc  gtg  aac  cca
gca  atc  atc  atg  gga  gca  AGA  AGA  aag  gcc  atg  AGG  aag  tac  gcg
aag  gag  gtt  gga  aat  gcc  aac  ATA  gct  aac  gcg  act  gtt  gtc  aaa
gtg  gat  gca  gaa  gta  gat  gga  gaa  aat  gct  acg  aag  cct  gaa  gaa
gag  gac  gac  gaa  gtc  aaa  ttt  CCC  cca  ATA  gtc  atc  aaa  aca  aca
ttt  cct  gaa  gag  ctg  cag  AGG  ttt  ttg  CCC  cag  ggg  gaa  aat  atc
cac  act  gtc  att  CTA  gac  ttc  acg  cag  gtc  aat  ttt  atg  gat  tct
gtt  gga  gtg  aaa  act  ctg  gct  ggg  att  gtg  aaa  gag  tac  ggc  gat
gtt  gga  att  tat  gtg  tac  tta  gca  gga  tgc  agt  gca  caa  gtt  gtg
aat  gac  ctc  acc  agc  aac  cgt  ttt  ttt  gaa  aat  cct  gcc  tta  aag
gag  ctt  ctg  ttc  cac  agt  atc  cat  gac  gcg  gtc  ttg  ggc  agc  caa
gtc cgt gag gcg atg
```

Red = rare Arg codons **AGG, AGA, CGA**
Green = rare Leu codon **CTA**
Blue = rare Ile codon **ATA**
Orange = rare Pro codon **CCC**

Figure 10: Rare codons highlighted with different colors in the construct for the STAS of *Rattus norvegicus* with the intervening sequence (underlined). Calculated with: http://nihserver.mbi.ucla.edu/RACC/.

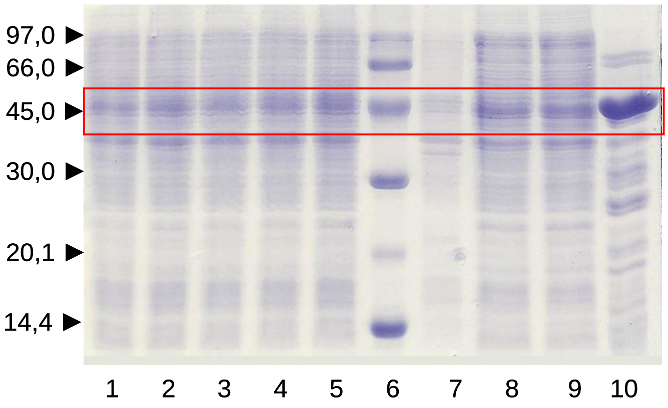


Figure 11: SDS-PAGE, gel15%. *E. coli* BL21 Rosetta.Loading order: (1) not induced expression; (2) 1h expression; (3) 2h expression; (4) 3h expression; (5) 4h expression; (6) Mw marker (KDa); (7) pellet; (8) supernatant; (9); flowthrough of the IMAC loading; (10): protein eluted at 150mM imidazole from IMAC.

4.2 Result of the expression and purification of the Saier motif

The expression and purification of the Saier motif (histag-SUMO-Saier: Mw:18556,8; 168 aa; Saier: 5372,0; 51 aa) was attempted twice. In the first attempt, the cellular pellet was lysed by sonication (figure 12). What remained in solution was loaded anyway in the IMAC column, but the protein precipitated during the loading. It is interesting to note that the fusion protein presented a second band in the SDS-PAGE, suggesting that it underwent a degradation process. Moreover, the fused SUMO protein should have increased the solubility but this did not happen, showing a dramatic aggregation during the first chromatographic step of the purification.

In the second attempt, the bacteria were squeezed with three cycles of French Press, which gave a more efficient lysis and most of the protein was in the supernatant (figure 13).

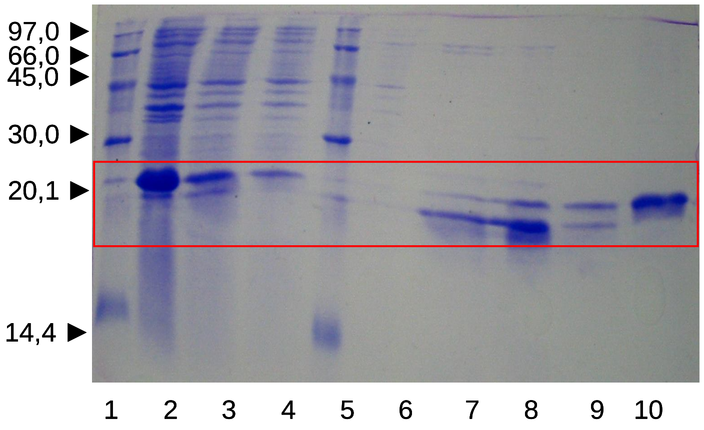


Figure 12: SDS-PAGE of the fusion protein Saier-SUMO-histag, first attempt, gel 17%. Loading order: (1) Mw marker (KDa); (2) pellet; (3) supernatant; (4) flowthrough; (5) Mw Marker (KDa); (6) elution at 10 mM imidazole from IMAC; (7) (8) protein eluted at 150mM imidazole from IMAC, two different fractions; (9) (10): 300mM imidazole from IMAC, two different fractions.

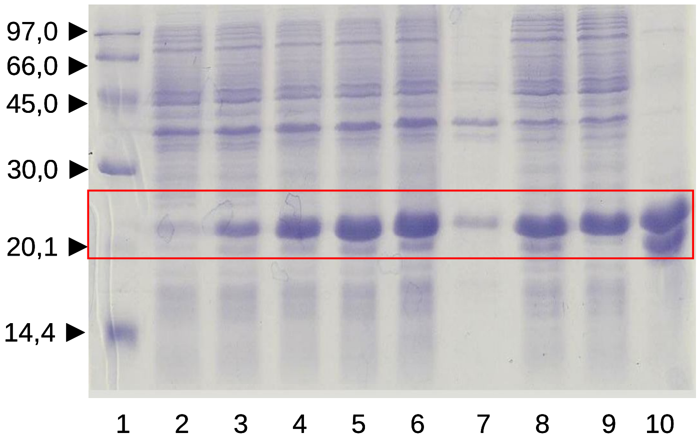


Figure 13: SDS-PAGE the fusion protein Saier-SUMO-histag, second attempt, gel 17%. Loading order: (1) Mw marker (KDa); (2) non-induced expression; (3) 1h expression; (4) 2h expression; (5) 3h expression; (6) 4h expression; (7) pellet; (8) supernatant; (9) flowthrough (10): protein eluted at 150 mM imidazole from IMAC

Anyway, the Saier motif was over expressed but during the purification it had serious aggregation problems. Most likely the topological model that we used to design the construct Saier motif is not applicable to the prestin of *Rattus norvegicus* and it could be a re-entrant loop or a transmembrane segment in prestin.

4.3 Expression and purification of the construct 517-653 of the STAS domain of SULTR1;2 from *Arabidopsis thaliana*

A construct of the STAS of SULTR1.2, previously cloned, encompassing the segment 517-653 (histag-SUMO-STAS: Mw: 30497,7; 271 aa; STAS-SULTR1.2: 17172,7; 152 aa), was expressed for 4 hours, at 37°C in Minimal Medium M9 supplemented with $^{15}NH_4Cl$. The expression was not high and a lot of over-exspressed protein remained in the cellular pellet after the lysis (lane 6, fig.14). We continued with the purification of the soluble fraction (lane 7, fig.14).

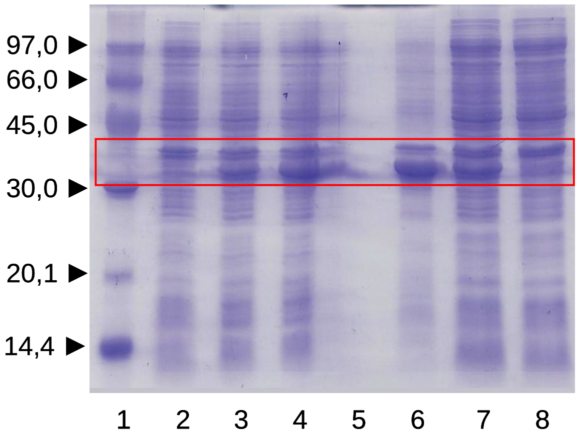


Figure 14 SDS-PAGE, gel15%. Loading order: (1) Mw marker (KDa); (2) not induced expression; (3) 2h expression; (4) 4h expression; (5) (empty); (6) pellet; (7) supernatant; (8) flowthrough of the IMAC loading.

After the removal of the histag-SUMO, a process of aggregation started, as showed by the asymmetric shape and the long tail of the first two peaks (0 mM imidazole), which correspond to the cleaved protein (figure. 15). The SDS-PAGE confirmed that the first two peak are composed by the same protein (fig. 16). We tried to achieve a NMR sample but the protein appeared strongly aggregated in the HSQC experiment (figure 17) and precipitated completely after few minutes in the magnet The precipitated protein was clearly visible in the NMR tube.

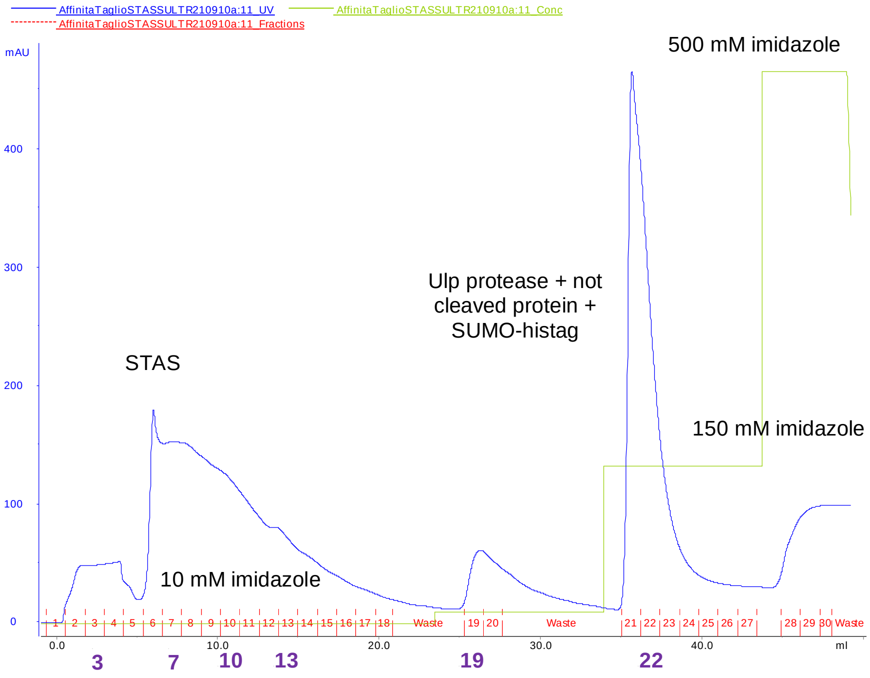


Figure 15: Chromatogram of the second IMAC purification STAS-SULTR1.2 (after cleavage).

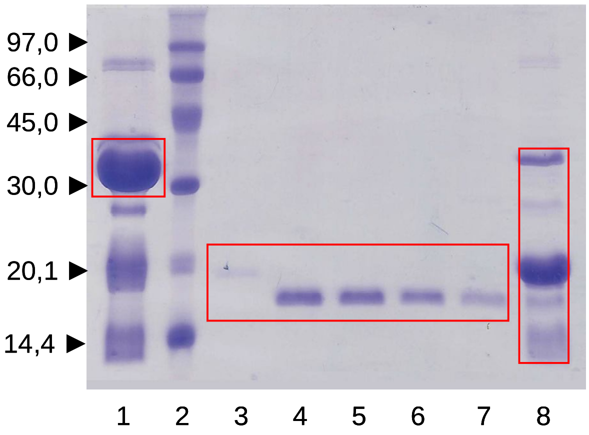


Fig.16: SDS-PAGE, gel 15%. Loading order: (1) protein eluted at 150mM imidazole from the first IMAC; (2) Mw marker (KDa); (3) fraction 10mM imidazole from the second IMAC (fraction 3 of fig. 14); (4) (5) (6) fractions 10mM imidazole from the second IMAC (fractions 7, 10, 13 respectively, of fig.14); (7) fraction 20mM imidazole from the second IMAC (fraction 19 of fig.14); (8) fraction 150mM imidazole from the second IMAC (cut protein+SUMO-histag+SUMO protease, fraction 22 of fig. 14).

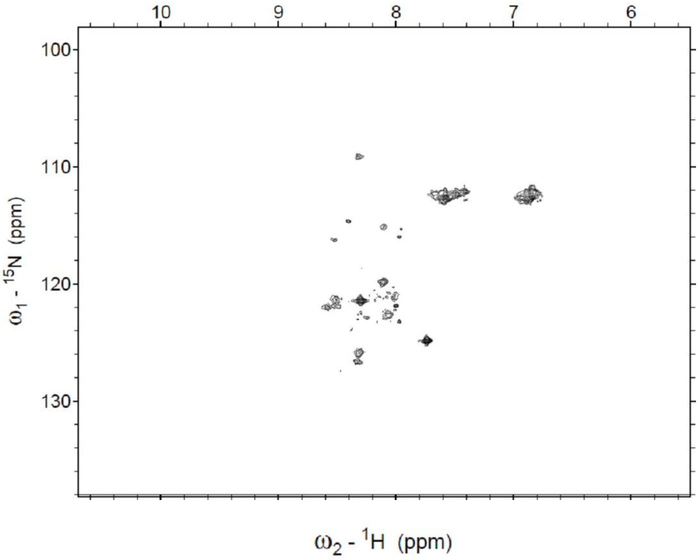


Figure 17: HSQC of the STAS-SULTR1.2 517-653, just after few minutes of the experiment. The collapsed signals are due to the aggregation process of the protein.

<u>4.4 Expression trials of the constructs 525-, 527-, 531-653 of the STAS domain of SULTR1;2 from *Arabidopsis thaliana*</u>

The expression of the new constructs in the SUMO or pETM20 plasmid of the STAS domain of SULTR1;2 was tested for 4 h, at 20° and 37°C, in 100 ml of Minimal Medium M9. At 37°C, bacteria grow very well but the temperature could be critical for protein stability, while at 20° the risk of protein precipitation in the inclusion bodies is decreased but the bacterial growth is slowed down: the expression at this temperature is consequently run for longer periods, usually overnight.

The expression showed a dependence on the incubation temperature for what concern the SUMO plasmid, since the protein was hardly expressed at 20°C (figure 18 A, 20 A, 21 A), but copiously at 37°C (figure 18 B, 20 B, 21 B), while for the pETM20 a good expression level was achieved also at 20°C (figure 19). The following SDS-PAGE gels represent some of the expression trials.

The construct 522-653 in the SUMO plasmid was included in these trials but was tested later.

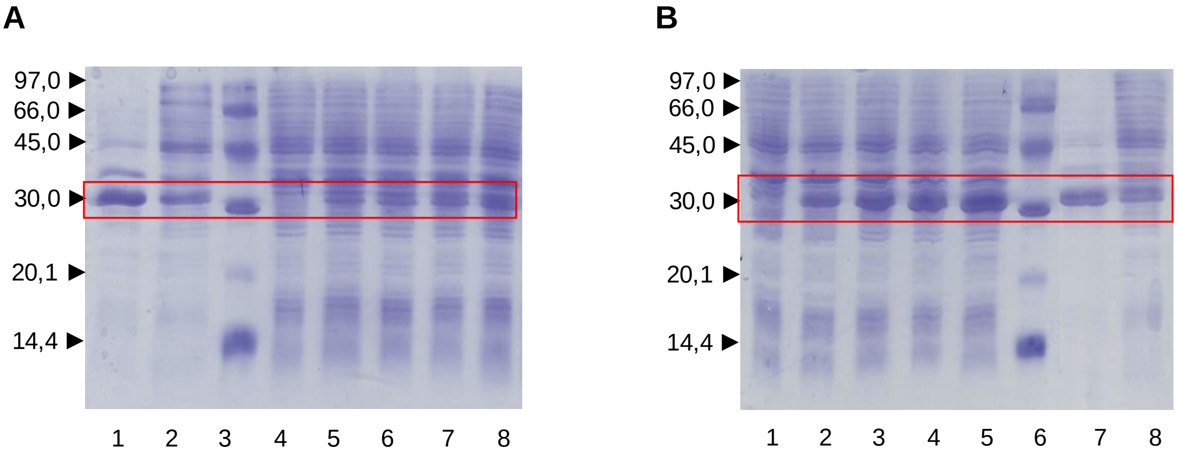


Figure 18: (A) SDS-PAGE of the expression of SUMO-STAS 525 at 20°C, gel 15%. Loading order: (1) pellet; (2) supernatant; (3) Mw marker (KDa); (4) 0h expression; (5) 1h expression; (6) 2h expression; (7) 3h expression; (8) 4h expression. **(B)** SDS-PAGE of the expression of SUMO-STAS 525 at 37°C, gel 15%. Loading order: (1) 0h expression; (2) 1h expression; (3) 2h expression; (4) 3h expression; (5) 4h expression. (6) Mw marker (KDa); (7) pellet; (8) supernatant.

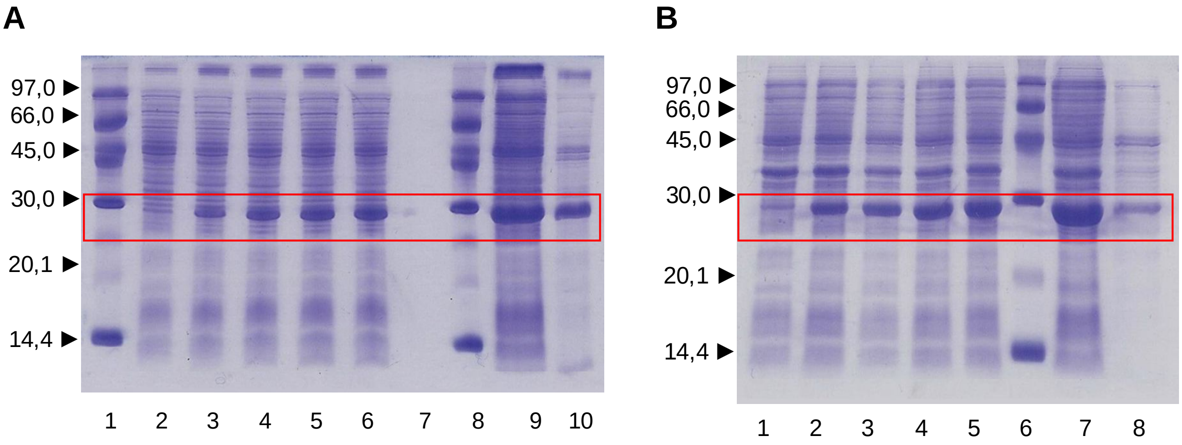


Figure 19: (A) SDS-PAGE of the expression of pETM20-STAS 525 at 20°C, gel 15%. Loading order: (1) Mw marker (KDa); (2) 0h expression; (3) 1h expression; (4) 2h expression; (5) 3h expression; (6) 4h expression; (7) (empty); (8) Mw marker (KDa); (9) pellet; (10) supernatant **(B)** SDS-PAGE of the expression of pETM20-STAS 525 at 37°C, gel 15%. Loading order: (1) 0h expression; (2) 1h expression; (3) 2h expression; (4) 3h expression; (5) 4h expression. (6) Mw marker (KDa); (7) pellet; (8) supernatant.

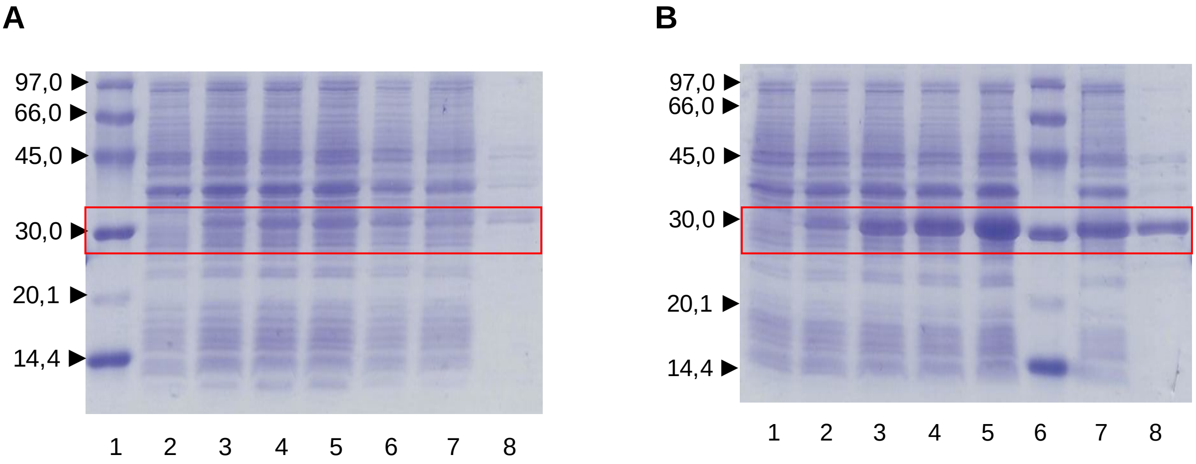


Figure 20: (A) SDS-PAGE of the expression of SUMO-STAS 527-653 at 20°C, gel 15%. Loading order: (1) Mw marker (KDa); (2) 0h expression; (3) 1h expression; (4) 2h expression; (5) 3h expression; (6) 4h expression; (7) (empty); (8) Mw marker (KDa); (9) pellet; (10) supernatant **(B)** SDS-PAGE of the expression of SUMO-STAS 527 at 37°C, gel 15%. Loading order: (1) 0h expression; (2) 1h expression; (3) 2h expression; (4) 3h expression; (5) 4h expression. (6) Mw marker (KDa); (7) pellet; (8) supernatant.

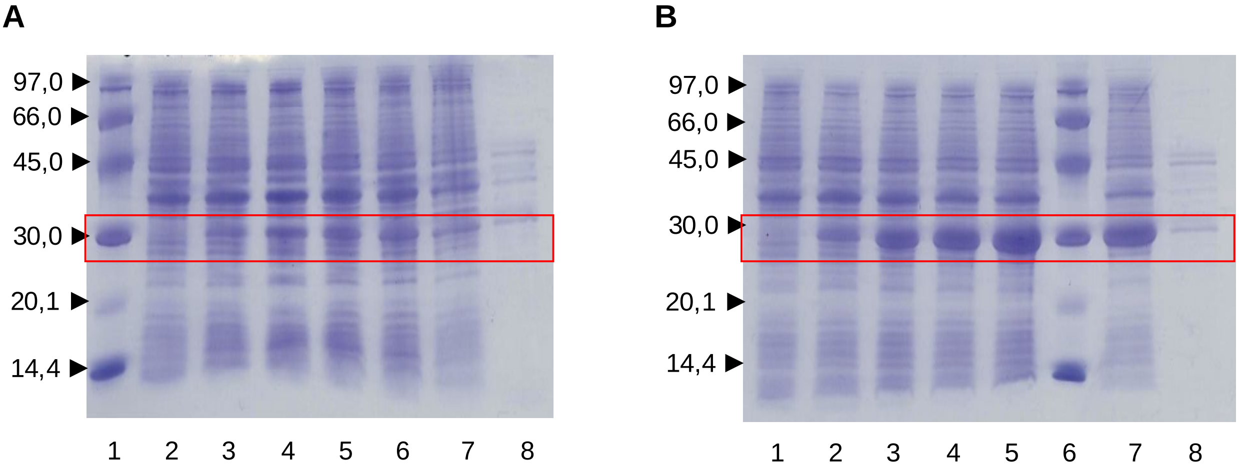


Figure 21: (A) SDS-PAGE of the expression of SUMO-STAS 531 at 20°C, gel 15%. Loading order: (1) Mw marker (KDa); (2) 0h expression; (3) 1h expression; (4) 2h expression; (5) 3h expression; (6) 4h expression; (7) pellet; (8) supernatant. **(B)** SDS-PAGE of the expression of SUMO-STAS 531 at 37°C, gel 15%. Loading order: (1) 0h expression; (2) 1h expression; (3) 2h expression; (4) 3h expression; (5) 4h expression. (6) Mw marker (KDa); (7) pellet; (8) supernatant.

4.5 Expression and purification of the construct SUMO-STAS 527-653

The construct SUMO-STAS 527-653 was the first one we tried to express and purify in larger amount, with a overnight expression at 25°C and purification with the FPLC (figure 22 and 23). The cleavage step was critical: we performed it at 4°C, under agitation and this was fatal to the protein that clearly underwent precipitation. The purification was carried on with the fraction of protein that was still in solution, which anyway precipitated soon afterwards (figure 24). It is interesting to note that in the SDS-PAGE lanes where the cleaved STAS should be, approximately at the level of 14 KDa, there was just a faint band, but there was a big one at 20KDa.

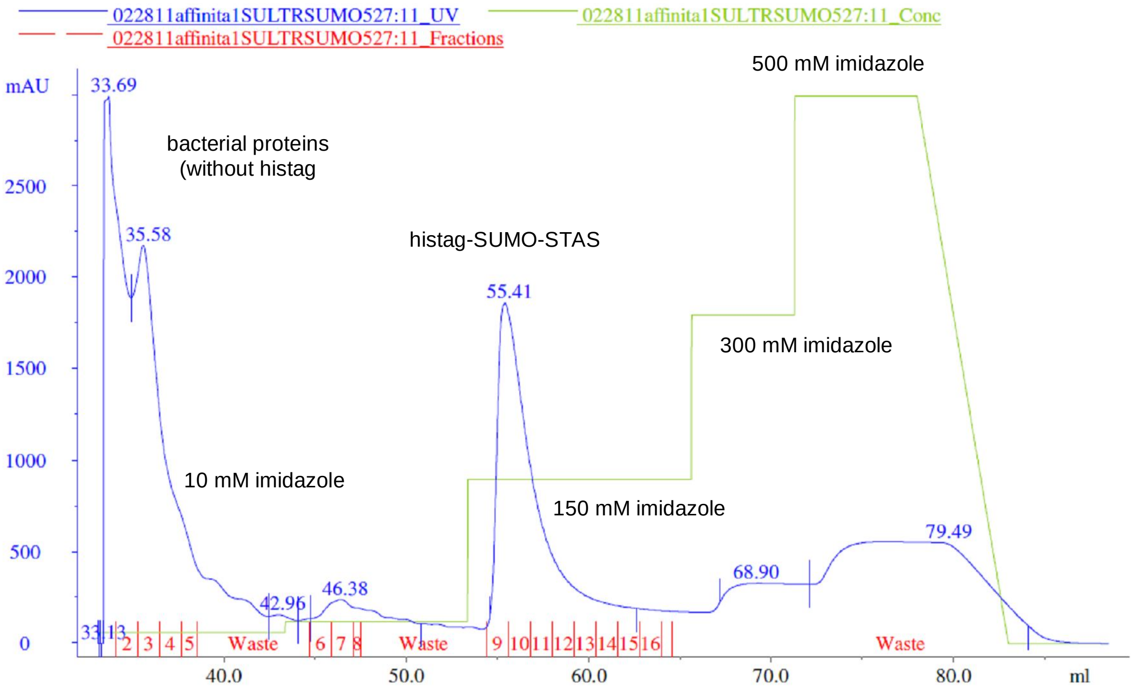


Figure 22: First affinity chromatography of SUMO-STAS 527-653.

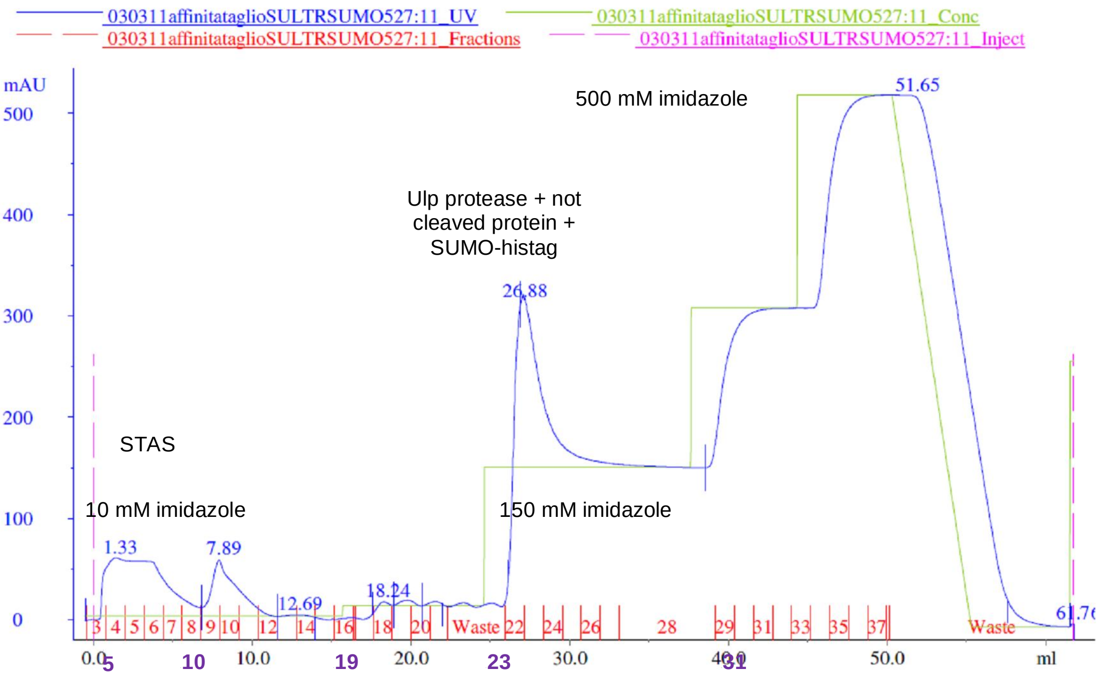


Figure 23: Second affinity chromatography of SUMO-STAS 527-653.

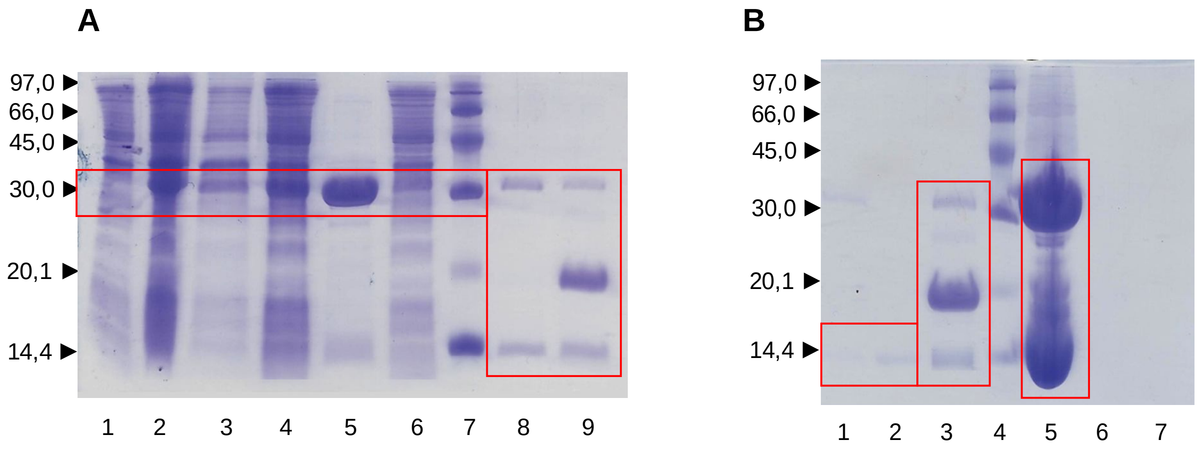


Figure 24: (A) SDS-PAGE of the expression and purification SUMO-STAS 527-653 at 25°C overnight, gel 15%. Loading order: (1) 0h expression; (2) 15h expression; (3) pellet; (4) supernatant; (5) protein of interest with histag, pre-cleavage, 150 mM imidazole, first affinity purification, ; (6) flowthrough of the IMAC loading; (7) Mw marker (KDa); (8) pellet of the protein of interest, after cleavage with stirring; (9) supernatant of the cleavage.**(B)** SDS-PAGE of the expression and purification SUMO-STAS 527-653 at 20°C overnight, gel 15%. Loading order: (1) second affinity purification, fraction 5, 10 mM imidazole; (2) second affinity purification, fraction 10, 10 mM imidazole; (3) SUMO protease/uncut protein, second affinity purification, fraction 23, 150 mM imidazole; (4) Mw marker (KDa); (5) pellet of the cleavage(overloaded); (6) fraction 19, second affinity purification; (7) fraction 31, second affinity purification.

The protein expression was satisfactory but not its stability in solution after the removal of SUMO, probably exacerbated by the stirring conditions: for the following trials a cleavage with very mild agitation was preferred.

4.6 Expression and purification of the construct SUMO-STAS 522-653

The longest of the constructs, SUMO-STAS 522-653, was expressed and purified (figure 25, 26 and 27) in the same conditions as the construct starting from 527, with the exception of the expression temperature set at 20°C and of the cleavage, which was carried out at 4°C, overnight, with very mild stirring. As it happened for the other SUMO construct, the 527-653, in the SDS-PAGE, in the lanes where the STAS protein should be, there was just a light band at about 14 KDa (its Mw) and a big one at 20 KDa.

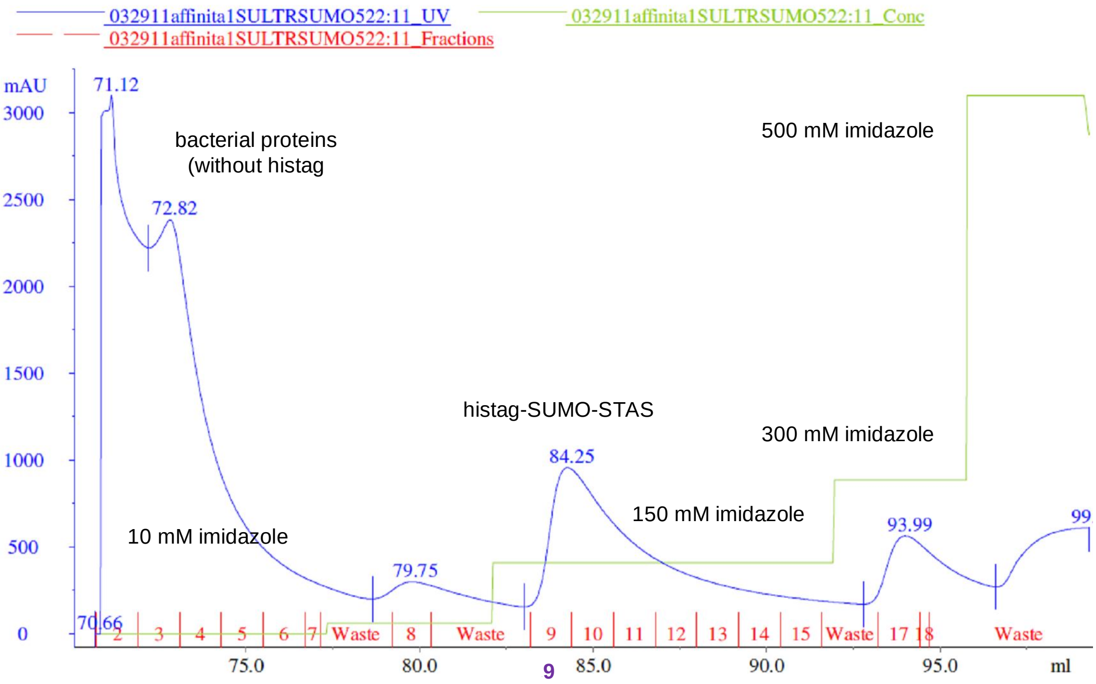


Figure 25: First affinity chromatography of SUMO-STAS 522-653.

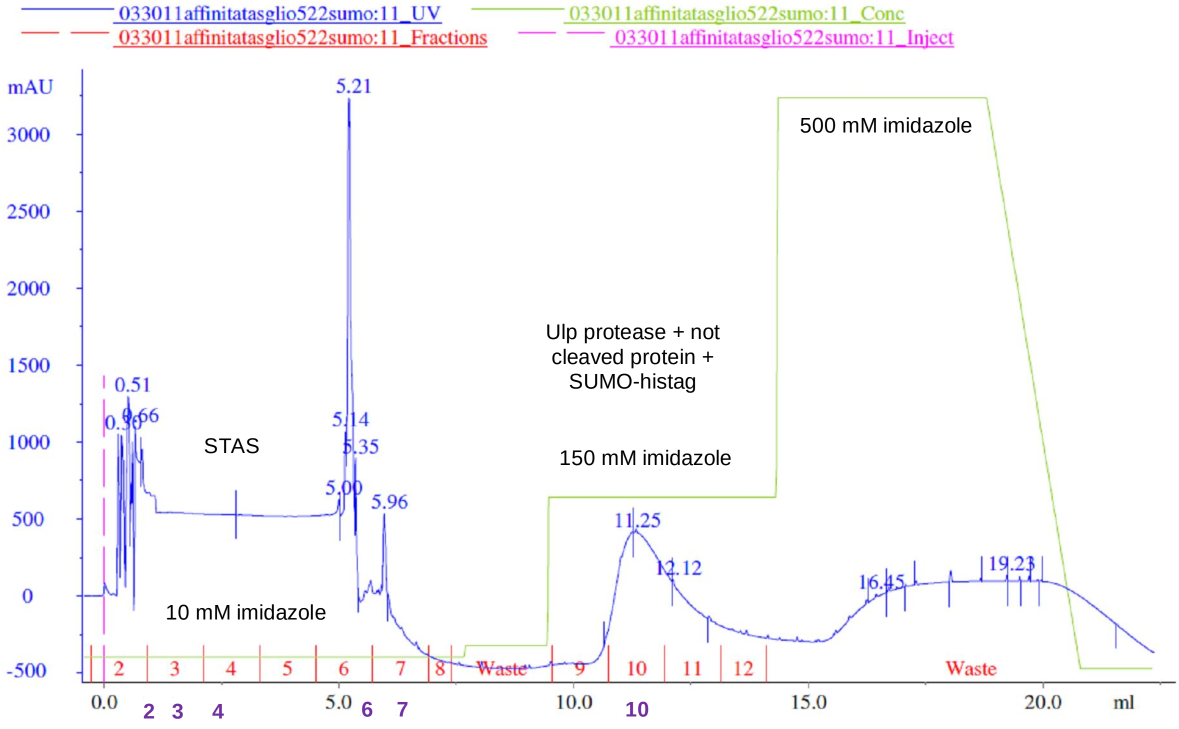


Figure 26: Second affinity chromatography of SUMO-STAS 522-653. The spikes on the first part of the chromatogram are due to a technical problem (air bubbles in the FPLC).

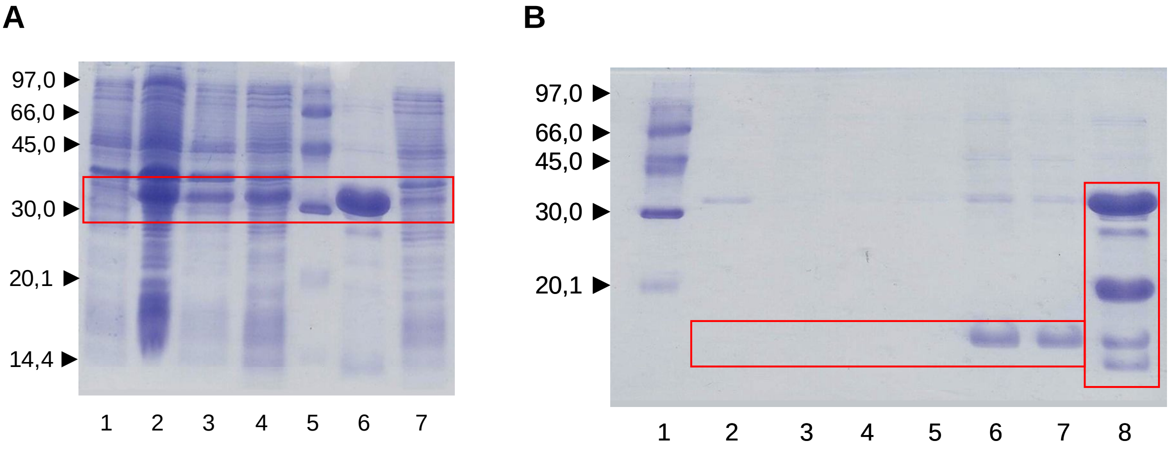


Figure 27: (A) SDS-PAGE of the expression and purification SUMO-STAS 522-653 at 20°C overnight, gel 15%. Loading order: (1) 0h expression; (2) 15h expression; (3) pellet; (4) supernatant; (5) Mw marker (KDa); (6) protein of interest with histag, pre-cleavage, first affinity purification, fraction 9, 150 mM imidazole; (7) flowthorough. **(B)** SDS-PAGE of the expression and purification SUMO-STAS 522-653 at 20°C overnight, gel 15%. Loading order: (1) Mw marker (KDa); (2) second affinity purification, fraction 2, 10 mM imidazole; (3) second affinity purification, fraction 3, 10 mM imidazole; (4) second affinity purification, fraction 4, 10 mM; (6) second affinity purification, fraction 6, 10 mM; (7) second affinity purification, fraction 7, 10 mM; (8) SUMO protease/uncut protein, second affinity purification, fraction10, 150 mM.

4.7 Expression and purification of the construct pETM20-STAS 525-653

For its first trial, the construct pETM20-STAS 525-653 was expressed at 25°C overnight as usual, and purified with the FPLC (figure 28, 29 and 30). The cleavage was carried out under no agitation.

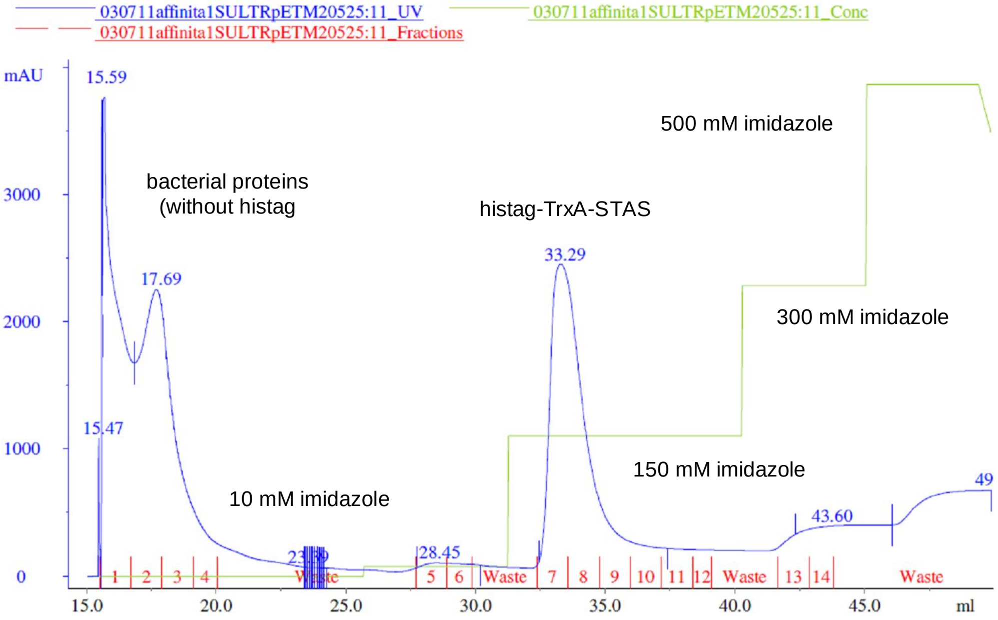


Figure 28 : First affinity chromatography of pETM20-STAS 525-653.

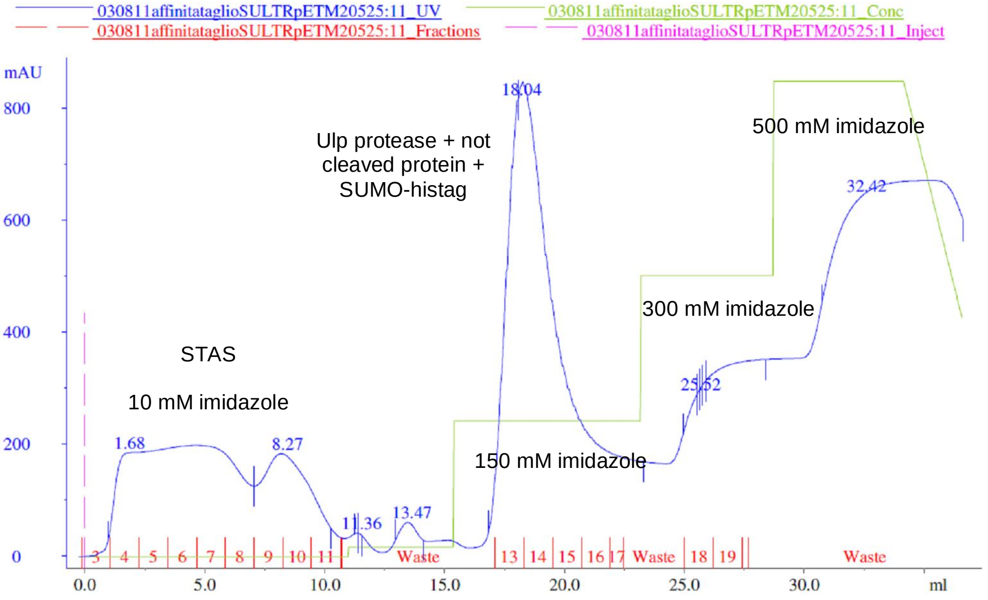


Figure 29: Second affinity chromatography of pETM20-STAS 525-653.

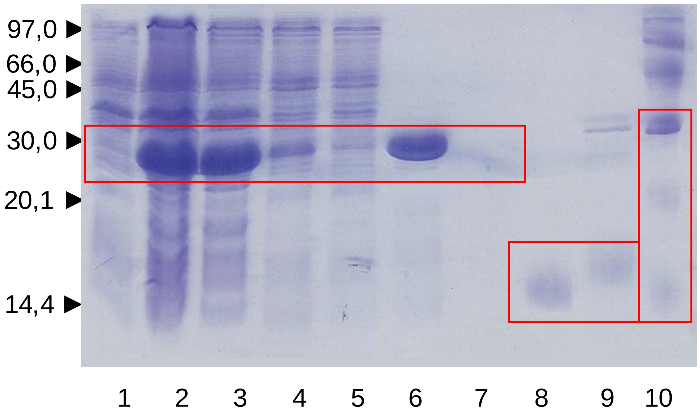


Figure 30 : SDS-PAGE of the expression and purification of pETM20-STAS 525-653 at 20°C overnight, gel 15%. Loading order: (1) 0h expression; (2) 15h expression; (3) pellet; (4) supernatant; (5) flowthrough; (6) protein of interest with histag, pre-cleavage, 150 mM imidazole, first affinity purification, fraction 78; (7) protein of interest, after cleavage, second affinity purification, fraction 5, 10 mM imidazole; (8) protein of interest, after cleavage, second affinity purification, fraction 10, 10 mM imidazole;.(9) TEV protease/ uncut protein, second affinity, fraction 14, 150 mM imidazole; (10) Mw marker (KDa).

A second trial on this construct was attempted to test if the stability could be positively influence by a different buffer. Two liters of colture were prepared and the protein expression was carried out at 20°C. Both the coltures were purified with just nickel NTA beads in plastic columns under gravity , instead of FPLC. One liter of colture was purified using the traditional buffers with NaCl (figure 31, A), but the other one was purified by using the same buffers with the NaCl substituted by 50 mM Arg and 50 mM Glu (figure 31, B) (46). The sample in the buffer with Arg and Glu has a better outcome than the one in NaCl. The cleavage with the TEV protease was also performed for a total time of 3 day, at 4° and with no stirring, since a overnight reaction was not sufficient to cut all the protein.

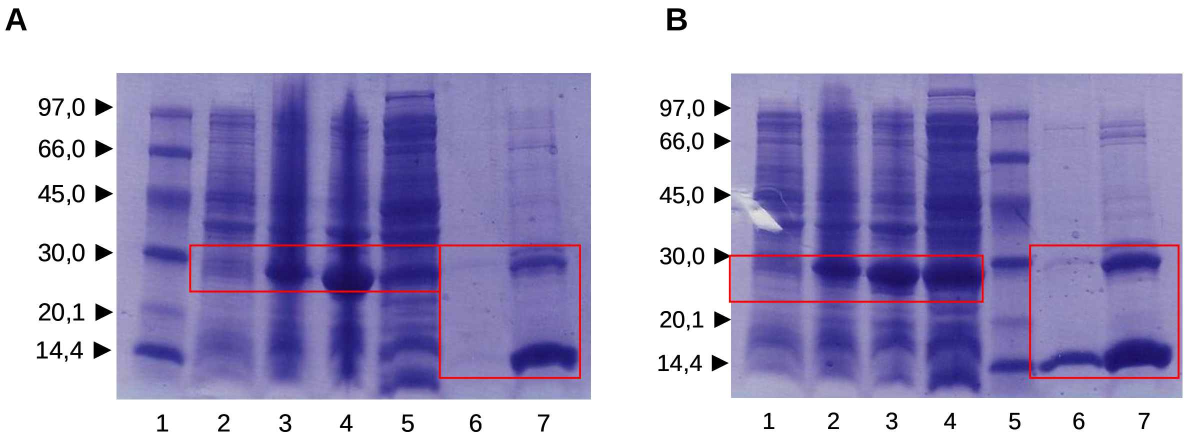


Figure 31: SDS-PAGE of the expression and purification of the construct 525-653 in the vector pETM20, gel 15%. **(A)** Purification performed with traditional buffers with NaCl. (1) Mw marker (KDa); (2) 0h expression; (3) 15h expression; (4) pellet; (5) supernatant; (6) protein of interest, after cleavage and after the second affinity purification, 0 mM imidazole (7) sample at the end of the cleavage, before purification, 150 mM imidazole. **(B)** Purification with buffers where NaCl was substituted with 50 mM Arg and 50 mM Glu (1) 0h expression; (2) 15h expression; (3) pellet; (4) supernatant; (5) Mw marker (KDa); (6) protein of interest, after cleavage and after the second affinity purification, 0 mM imidazole (7) sample at the end of the cleavage, before purification, 150 mM imidazole.

Since the sample was split in five other samples with different buffers, (listed in Materials and Methods) the final concentration in each tube was approximately 80 uM. The samples were first compared as 1D spectra: the buffer n. 5 (50 mM phosphate, 50 mM L-Arg, 50 mM L-Glu, 2% glucose, 10 mM DTT, EDTA 1 mM, 0.05% NaN_3, pH 6.5, 10% D_20) gave the strongest signal and best signal resolution. All the samples tested in the different buffers were all collected in one, concentrated and tested in the buffer n. 5 (50 mM phosphate, 50 mM L-Arg, 50 mM L-Glu, 2% glucose, 10 mM DTT, EDTA 1 mM, 0.05% NaN_3, pH 6.5, 10% D_2O) but at a lower pH (6.0). One HSQC of this sample was acquired (figure 32). Excluding the signals likely due to side chains, the HSQC peaks were approximately 85, out of 126 expected (132 amino acids, with 6 prolines). However, at a higher concentration and lower pH, the sample gave signs of aggregation and precipitated after 24 hours.

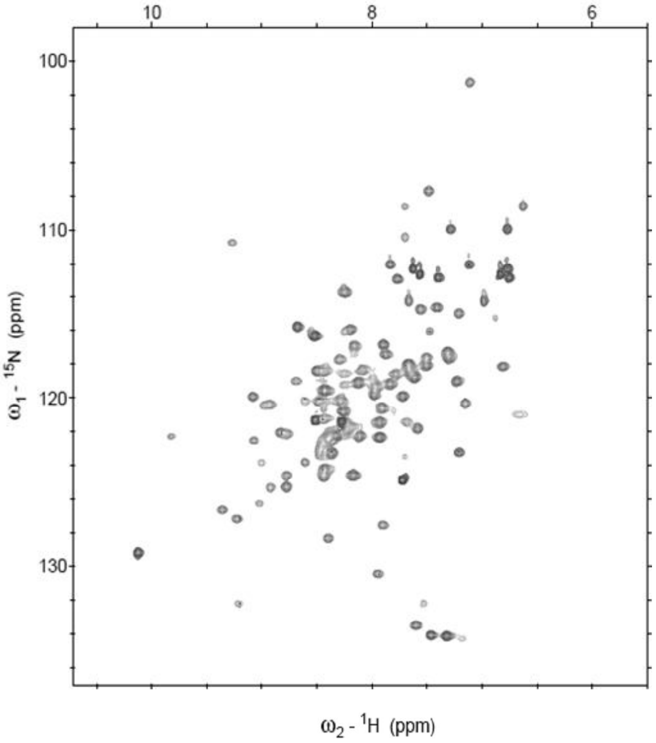


Figure 32: HSQC of the construct pETM20 525-653 of the STAS of SULTR1;2 in the buffer: 50 mM phosphate, 50 mM L-Arg, 50 mM L-Glu, 2% glucose, 10 mM DTT, EDTA 1 mM, 0.05% NaN_3, pH 6.0, 10% D_2O. Acquired at 600 MHz.

5. CONCLUSIONS AND FUTURE PERSPECTIVES

The complete STAS of prestin from *Rattus norvegicus*, with the IVS

On 2010 the first structure of a mammalian STAS was solved. This was an engineered version of the domain, deleted in the intervening sequence, which is the most evident difference with the bacterial ASA domains. It is predicted to be unstructured and its functional meaning is still obscure, but, since it is highly variable among the different proteins, it could be involved in transporter-specific interactions. Our aim was to obtain the full STAS of prestin, to complete the structure and to use it to investigate the particular interactions of prestin. The first expression and purification was attempted in the bacterial strain BL21(DE3) and the outcome was a very poor expression of the protein. Since the engineered STAS was expressed in abundant amount, the low yield of the full STAS could be evidently explained by the presence of the IVS. Unless its presence was destabilizing for the protein by itself, the characteristic on the IVS that could hamper its expression was hypothesized to be the presence of a large number of rare codons. Luckily, the strain Rosetta of *Escherichia coli* has been optimizied to overcome this problem, but in our case it did not change the result of a second expression, giving again a poor yield.

The Saier motif

The structure of prestin and of the SLC26A6 transporters in general is still unknown and debated. Consequently, any speculation on their organization in transmembrane segments is based on comparison with information about other transporters and experimental data. do not allow Solving the structure of transmembrane proteins is a very challenging task with the current technologies;anyway, extramembrane domains and independent fragments could be be isolated and solved separately. This is what we tried to do with the Saier motif, a portion of prestin characterized by the presence of the conserved triplet -NQE-. In the topological study reported in (45), the Saier motif is suggested to be cytosolic, and since some cytosolic portions are known to do interamolecular interaction with C-terminal domains (9), we wanted to evaluate if the same interaction can occur in prestin. The design of the construct was based then on the topology of BicA and it was expressed and partially purified: the protein showed immediately solubility problems, since it aggregated even as a fusion protein designed to increase the solubility. The presence of a second band in the SDS-PAGE highlighted also possible degradation problems. The unsuccessful production of the Saier motif suggested that the topological model we used could probably not be applied to prestin and that the domain could belong to a transmembrane segment or being a loop re-entering in the membrane. The study on the Saier domain was consequently postponed to when reliable data about its topology will be available.

The STAS of SULTR1;2 from *Arabidopsis thaliana*

Most of our efforts was concentrated on getting a sample of the STAS domain of the sulfate transporter SULTR1;2 from *Arabidopsis thaliana*, as a STAS complete of IVS and as the first plant STAS. An old construct was tested first, and it precipitated. Five new constructs were therefore designed in two different plasmids and tested at different expression temperatures. Some of them was also expressed and purified in larger amount. We learnt some lessons from all the expression and purification attempts and we tried to address the detected problems:

- No meaningful differences were detected between the constructs, except that for an expression of four hours at 20°C the pETM20 construct was better than the SUMO. This difference was smoothed over by the overnight expression.
- The best expression temperature was 20°C, overnight: about the 40-50% of the protein remained in the pellet, but at 37°C the percentage was much higher. The soluble fraction seemed anyway sufficient.
- The cleavage of the fusion protein under stirring was deleterious, probably accelerating the aggregation process: we tried to slow it down by eliminating the stirring and lengthening the incubation time. The aggregation process was still present but reduced.
- The soluble fraction seemed enough to reach a NMR sample: we preferred to avoid refolding and to focus on optimizing the buffer conditions.
- Since the constructs had the same behavior pretty much but the pETM20 was more cost effective, we decided to continue with this construct.

- During the buffer screening, we notice that the substitution of NaCl with 50 mM Arg and 50 mM Glu, an expedient often used to increase protein solubility at relatively high concentration, helped to achieve a better yield. This allowed the achievement of a sample at a concentration suitable for NMR experiments and the acquisition of a ^{15}N-HSQC spectrum which showed a well dispersion of the peaks. This is indicative of the presence of a folded protein. The number of observed signals was anyway significantly lower than the expected one and this could be caused to line broadening due to conformational exchange in some portion of the protein.

We planned to screen other buffers, also with detergents, to evaluate if the concentration can be raised without compromising the stability and to achieve a higher percentage of peaks in the HSQC experiment.

2. INTRODUCTION

6.1 The Myb/SANT domains

The Myb/SANT domains are present in a variety of proteins with different functions, from chromatin remodelling enzymes to DNA-binding proteins (47). Their most striking characteristic is that, although their functions are heterogeneous, their structure is rather conserved: they are constituted by three helices, where the second and third helix are displaced in a Helix-Turn-Helix (HTH) motif, of approximately 50 amino acids (48). Three tryptophans (or other bulky hydrophobic amino acids), regularly repeated every 18-19 amino acids and defined as “the tryptophan cluster”, contribute to create the hydrophobic core (49) (50) (51) (figure 33, A and B)

So, since the structure is the same, what makes these domains functionally different? A simple answer is in the surface electrostatic properties: in fact, the binding surface of these domains is on the third helix and it is positively charged for the Myb (52) and negatively for the SANT (47) (53) (figure 33, C). These clearly enables them to bind different partners characterized by opposite charge, like DNA or histone-tails for the Myb and the SANT respectively: indeed, as an example, the name Myb comes from the three tumour suppressor DNA-binding transcription factors Myb (A-, B- and c-Myb), while the acronym SANT is derived from four histone-interacting enzymes involved in chromatin-remodelling complexes (switching-defective protein 3 (Swi3), adaptor 2 (Ada2), nuclear receptor co-repressor (NCoR), transcription factor IIIB (TF IIIB)) (47). However, the answer to what makes these two domains different is more complex, because there is evidence that also the first helix can be a binding surface (54).

A further modulation of the function can be due to accessory structural elements, such as additional helices, beta-hairpins or a slight different spatial arrangement of the helices. For example, the C-terminal β-hairpin of the Myb3 of *Thricomonas vaginalis*, contributes to the stabilization of the interaction with DNA (55), or the R2 repeat of the DNA-binding domain of c-Myb undergoes a phenomenon of thermal (56) and conformational instability of its third helix (due to an unusual cavity in the hydrophobic core) that is attenuated upon the binding with the DNA (57) (58) (59).

Not all the Myb bind DNA, such as the single Myb present in the human protein Rap1 (60), which binds to hTRF2, a telomeric repeat binding protein (60). Interestingly, its *Saccharomyces cerevisiae* homologue has not lost the DNA-binding capability. Other typical Myb with a very weak or non-existent DNA interaction are the first repeat of many Myb-like DNA-binding domain, which are organized as multiple imperfect Myb repeats (from two to four) (61). In those domains, the DNA-binding activity resides in the repeats R2 and R3 (62), while R1 is dispensable, as it occurs in many of the Myb proteins in *Arabidopsis thaliana*, where R1 is not present for most of them.

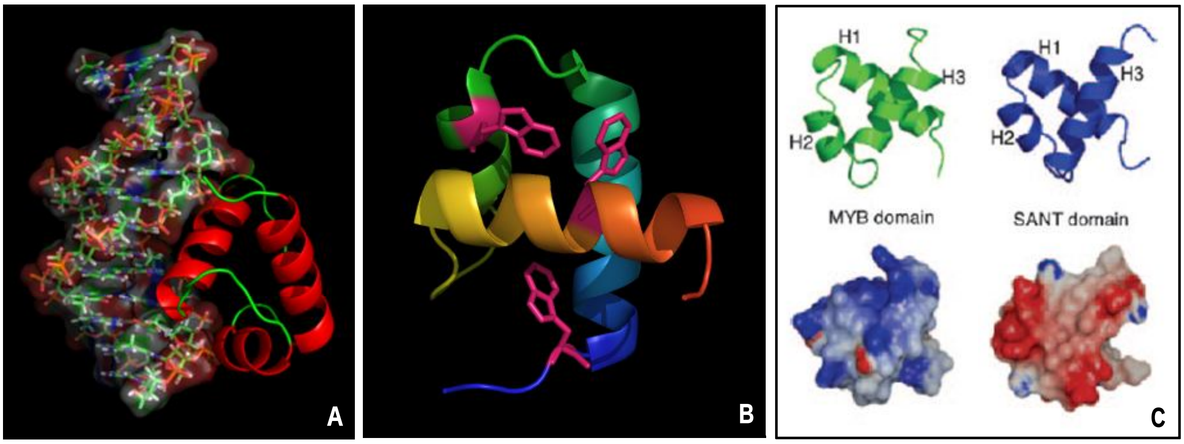


Figure 33 : (A) the Myb-like DNA-binding domain from hTRF1 (PDB: 1IV6); **(B)** a R1 domain from the mouse protein c-Myb, with the tryptophan cluster in magenta (PDB:1GUU); **(C)** comparison of the third helix of a Myb and a SANT domain, from the point of view of the electrostatic surface (adapted from (63)).

6.2 The cyclin D-interacting Myb-like protein 1(Dmp1)

The cyclin D-interacting Myb-like protein was discovered first in mouse (64) and then in human (hDmp1) (65) as phoshorylation substrate by the cyclin D-dependent kinases CDK4 and CDK6.

hDmp1 exists in three isoforms: the α is the 760 amino acid full-length isoform of the protein, with a central DNA binding domain flanked by cyclin D binding sites and acidic trans-activating domains. The other two isoforms, β and γ, are truncated before the DNA-binding domain (66) (figure 35). Dmp1 does not belong to any gene family, but its DNA binding domain is classified as a domain with three Myb-like repeats (67).

The isoform α of Dmp1 is a tumour suppressor transcription factor that exert its anti-oncogenic activity controlling the transcription of Arf (68), a pro-apoptotic protein. In fact, Dmp1 induces in primary cells an Arf-p53-dependent cell cycle arrest (68) (figure 35). hDmp1 recognizes the consensus sequence CCCG(G/T)ATGT (64), which can be found not only in the promoter of Arf (68), but also in the promoter of CD13/Aminopeptidase N: in this case, the D-type cyclins can hamper the binding (69). In fact, cyclins D have the ability to control hDMP1 forming heterodimers and they can block its interaction also with CDK4 and DNA (70). Other negative regulators of the activity of hDmp1 are: physiological mitogenic stimuli, overexpressed E2F (71), genotoxic stimuli by NF-B (72), while the oncogenic Ras through the Ras-MEK-ERK signaling (73) and HER2/neu (74) have an indirect positive effect (figure 34). The Dmp1 activity can be controlled also by the formation of heterodimers of the isoform α, the only one with the DNA-binding domain and the tumour suppressor activity, with the isoforms β and γ (66). A wrong control of Dmp1 activity can lead to pathologies: for example, the isoform β is involved in some forms of leukemia, due to splicing aberrations (66), but more frequently and understandably is the α isoform mostly involved in cancer. In fact, the deletion of the hDmp1 locus is associated with the development of carcinomas (75), such as non-small cell lung carcinomas (76) (77) and hematopoietic cancers (75). In *Dmp1-null* (*Dmp1*$^{-/-}$, with both the alleles inactivated) mice , with the gene disruption at the Myb-like DNA-binding domain of *Dmp1*, it has been observed a development of pulmonary adenomas/ adenocarcinomas, vascular tumors, hepatocellular adenomas/ adenocarcinomas, B-cell lymphomas and other tumours from the second year of life, with the tumorigenesis process accelerated by carcinogens (78). Other less severe effects included for example urologic dysfunctions (78). Also the *Dmp1*$^{+/-}$ mice showed a certain propensity to the development of spontaneous cancers, suggesting that Dmp1 is haploinsufficient for its anti-oncogenic activity (79). Recently, Dmp1 was discovered as a interaction partner of another tumour suppressor, p53, with a positive function in its stability, localization and activity, in an Arf-independent way (80).

mDmp1 exerts an upregulation in genes involved in cell proliferation (adrenomedullin), apoptosis (Bcl-3), transcription (JunB, Egr1), DNA methylation (methyl-CpG binding domain protein 1), angiogenesis and metastasis (thrombospondin-1), and a negative effect on Gas1 and Ect2, which are involved in cell cycle and proliferation, as demonstrated in lung tissue of null mice with GeneChip Microarrays (81). For some of these genes (amphiregulin, thrombospondin-1, JunB and Egr1) a binding of mDMP1 to their promoter has been demonstrated (81).

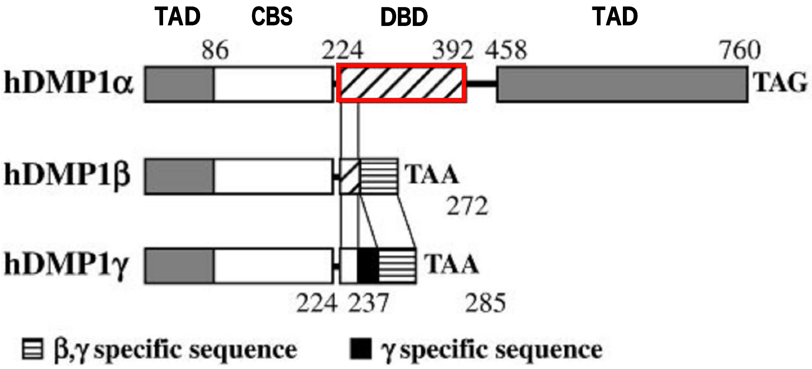


Figure 34: The three isoforms of hDMP1 (TAD: transactivation domain; CBS: cyclin D binding domain; DBD: DNA-binding domain) (66).

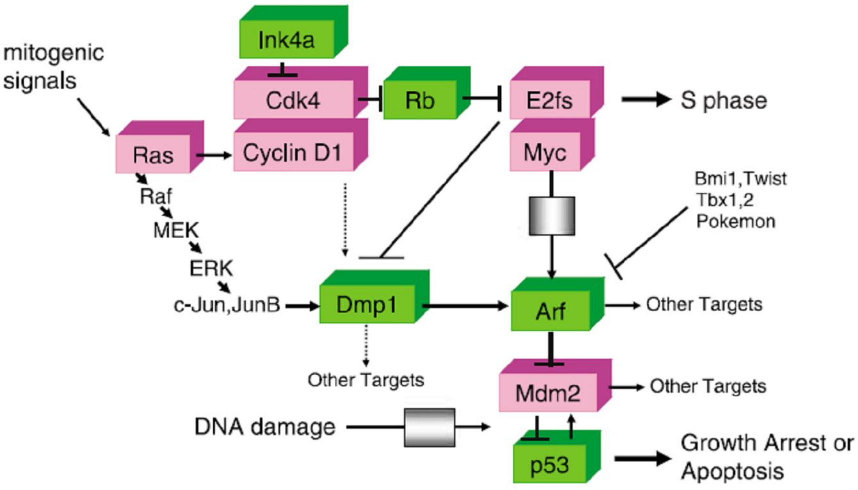


Figure 35: Dmp1 is involed in pathways linking many other tumour suppressors, like Retinoblastome, Arf, p53, Ink4a (82).

6.3 The Nuclear Receptor CoRepressor 2 (NCoR2)

NCoR2 (Nuclear Receptor Corepressor 2) is a repressor of nuclear receptor activity. In the absence of the ligand, the repressor recruits other proteins forming a complex that keeps the nuclear receptor in an idle state. In the presence of the ligand (a hormone), the repressor complex is removed and the activator complex is formed: now the nuclear receptor can dimerize and bind to the “hormone responsive element” DNA sequence, starting the transcription (83) (84).

NCoR2, together with its homologue NCoR1 (also known as SMRT and NCoR respectively), are among the most studied repressors (85) (86). They are codified by different loci but they share 45% of sequence identity and similar structural organization: three or four transcriptional repression domain (RDs) and two or three nuclear receptor interaction domain (NRs) depending on the isoforms, respectively at the N-terminal and C-terminal end (84) (figure 36). They interact with several unligated nuclear receptors: the best known are thyroid hormone receptor T3R (85) (86), the retinoid hormone receptor RAR (86), the peroxisome proliferator activated receptor PPAR (87), the vitamin D receptor VDR (88); other receptors are reviewed in the reference (84). The interaction with the steroid hormone receptors is an exception, because it occurs or it is facilitated by the presence of the hormone antagonist (89). It is interesting to note that in the disease where the corepressors are involved, the pathology is mainly explained by an aberrant corepressor-nuclear receptor interaction due to mutation in the receptor, with outcomes ranging from low severity diseases to leukemia (90) (91) (92).

In the binding with the nuclear receptors, the corepressors are assisted by other protein components with different roles, like, for example, organizing the chromatin condensation (93). For NCoR2, the best known of these partners are: G protein pathway suppressor 2 (GPS2) (94), the Transducin-like protein 1 (TBL-1) (95), the Transducin-like 1 related protein (TBLR-1) (96) and mSin3 (97). The interaction occurs on the surface of the RDs on the corepressor, which is supposed to be mainly unstructured, in order to accommodate specific but transient interactions (99). The two SANT domains of NCoR2 seem to be among the few structured regions of the corepressor (84). Even if they belong to the same family, the functions of these two domains are distinct: the N-terminal SANT domain is known also as “Deacetylase Activation Domain” (DAD) because it binds and activate the histone deacetylase 3

(HDAC) (100), while the other SANT domain is known to be an interaction partner of the histone tails of H4, earning the definition of "Histone Interacting Domain" (HID) (101). The H4 histone tails are bound only when they are deacetylated, suggesting that this SANT domain stabilizes the product rather than the substrate of HDAC3 (102).

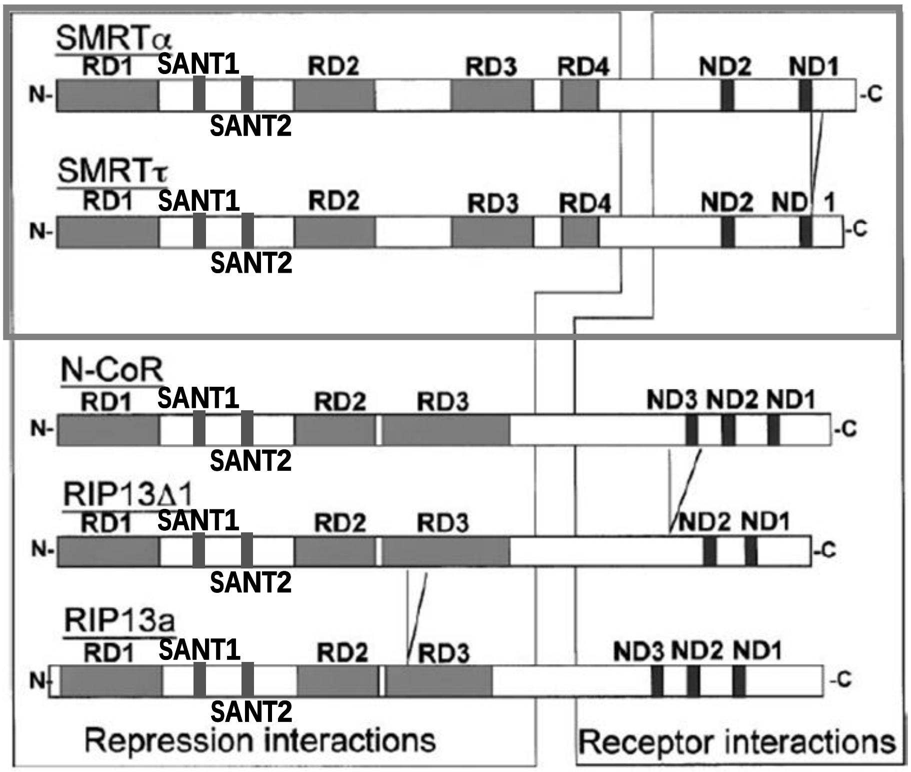


Figure 36: The two isoforms of NCoR2 (RD: repression domain; ND: nuclear receptor interaction domain) (84).

6.4 Fast NMR methods for structural proteomics according to the SGC Toronto procedures

The mainstream method to solve protein structures by NMR was originally developed by Bax et al. (103) (104) (105) (106). This approach can be divided in two main steps: the first one consists in determining the sequential assignment with $^{13}C/^{15}N$-experiments based on scalar couplings, while the second one considers the information extracted from the NOESY experiments to fold the structure. Some drawbacks are associated with this methodology: the experiments sometimes do not provide all the scalar connections needed because of sensibility limitations; the manual assignment is subjective, prone to inaccuracies and time consuming.; increasing the resolution of the spectra means lengthening the experimental time, with repercussions on protein stability for delicate samples. To address these issues, additional methods have been introduced, like TALOS, a computational approach to predict dihedral angle restraints (107); the introduction of the residual dipolar couplings (108); experiments with increased dimensionality but with extended acquisition time (109) (110) or with less paths of magnetization transfers, for an improved sensitivity (105).

Recently, the introduction of software with a comprehensive approach toward the interpretation of data (for instance, with PINE-SPARKY(111)) and the development of algorithms for automated structure calculation based on constraints extracted from the NOESY experiments (such as CYANA (112)) demonstrated that a new inclusive conception of solving protein structure by NMR was coming up. These new methods have some drawbacks too, like the fact that any inaccuracies will have consequences in the final structure.

The approach developed at the Structural Genomic Consortium (SGC) in Toronto, Canada, is in the frame of this new conception of assignment strategy and structure determination and, at the same time, it tries to address some problematic issues with new strategies (113).

The highlights of this method can be listed in four points:

1. A selection of experiments for assignment and structure calculation constituting the minimal dataset. The experiments have been chosen on the basis of sensibility, with fewer magnetization transfers and relaxation losses compared with the traditional approach. The total acquisition time

is reduced but the dataset is self-consistent, providing all the information needed for assignment and structure calculation.

2. A further reduction of the experimental time is achieved by acquiring the experiments in a scheme known as *Non-Uniform Sampling* (NUS) (114). NUS is applied with two main purposes: in three-dimensional triple resonance experiments, where sensibility is not an issue, it decreases the acquisition time by sampling less points following a precise scheme, without affecting the resolution; in experiments such as 3D- HCCH-TOCSY and ^{13}C- and ^{15}N- NOESY, where the spectra can be very crowded, the digital resolution is increased without lengthening the experimental time. The sparse dataset is restored to a full matrix with the *Multi Dimensional Decomposition* (MDD) algorithm and it is ready to be converted in file formats for NMR spectra analysis (as for example SPARKY (111),).
3. The HNCA and the NOESY experiments provide the information needed for the sequential assignment. In this way, the NOEs are not just seen as constraints for structure calculation but also as a source of information to assembly the sequence. This use of the NOESY in an extensive way as an assignment tool is a novelty.
4. All the steps are facilitated by two *Graphical User Interfaces* (GUI). The first GUI is the MDD-GUI, which leads the user to the reconstruction of the 3D *sparse* dataset acquired in NUS modality. The other one is the Fragment Monte Carlo-GUI (FMC-GUI) which helps the user in the peak lists management, resonance assignment, conversion of input file formats for other software and a number of other process, from the beginning to the preparation of the files for the final deposition of the structures in databases.

FMC-GUI is not just a management tool: it is integrated with ABACUS (115) (113), a procedure of critical importance for the resonance assignment and structure calculation in a semi-automated fashion.

6.5 The minimal ABACUS dataset: an accurate selection of heteronuclear 2D- and 3D-NMR experiments

For the determination of the structure of short peptides, the NMR experiments based only on the proton frequency and two dimensions (homonuclear experiments) can be sufficient to gather all the information required for sequential assignment, determination of the secondary structure elements and 3D structure. But if the sample presents a higher level of complexity, such as in a protein, the number of protons that generates peaks is sensibly larger and the resulting spectra are too crowded. This problem can be faced in two ways: using nuclei different from ^{1}H (heteronulei) and increasing the number of dimensions in the experiments.

The choice of the nuclei is related with their presence in the sample under investigation and on the abundance of the isotopes with a spin number ½ : for example, phosphorus is abundant in nucleic acid and the natural abundance of its useful isotope ^{31}P is 100%. For proteins, the useful heteronuclei are ^{15}N and ^{13}C, but their natural abundance is just 0.365% and 1.108% respectively: it is anyway possible to sensibly enrich a protein sample with the right isotopes just supplementing them in the minimal medium where the bacteria expressing the protein grow.

Once obtained a protein sample containing ^{1}H, ^{15}N and ^{13}C, these nuclei are exploited for experiments where they can be involved in magnetization transfers. Obviously, depending on the experiment, the scales are referred also to the chemical shift of the heteronuclei, which are wider than the one of the proton: 100-140 ppm for ^{15}N and 0-200 ppm for ^{13}C, depending on the experiment. Consequently, the possibility of spreading the peaks along larger scales reduces the problem of the signal crowding.

The fundamental 2D-heteronuclear NMR experiment for protein is the *Heteronuclear Single Quantum Coherence* (HSQC) (116). As all 2D experiments, it is divided in four phases: preparation, evolution, mixing, detection (figure 37). In this spectrum, each scalarly coupled ^{15}N and ^{1}H produce a peak: most of them come from the backbone, few others, in specific regions of the spectrum, from the side chains containing the N-H (figure 38). The magnetization is transferred from ^{1}H to ^{15}N and then back to ^{1}H with the INEPT sequence, to increase the overall sensitivity of the experiment, according to the scheme on figure 39 .

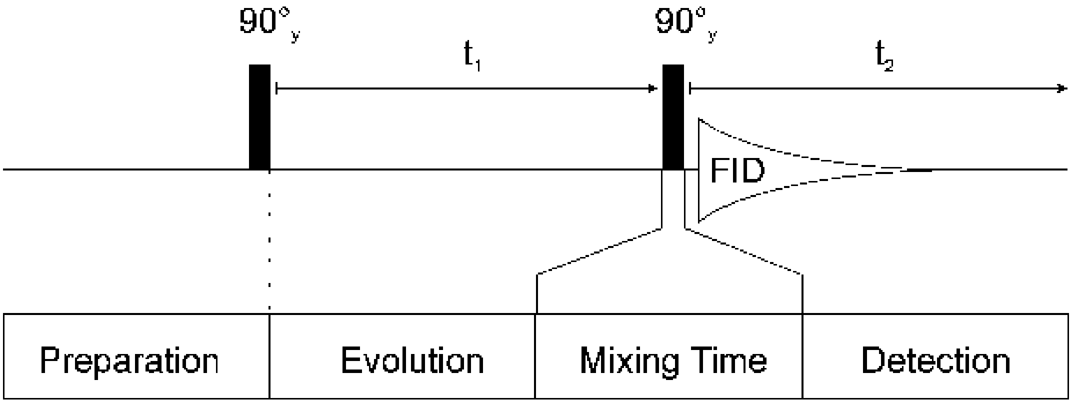
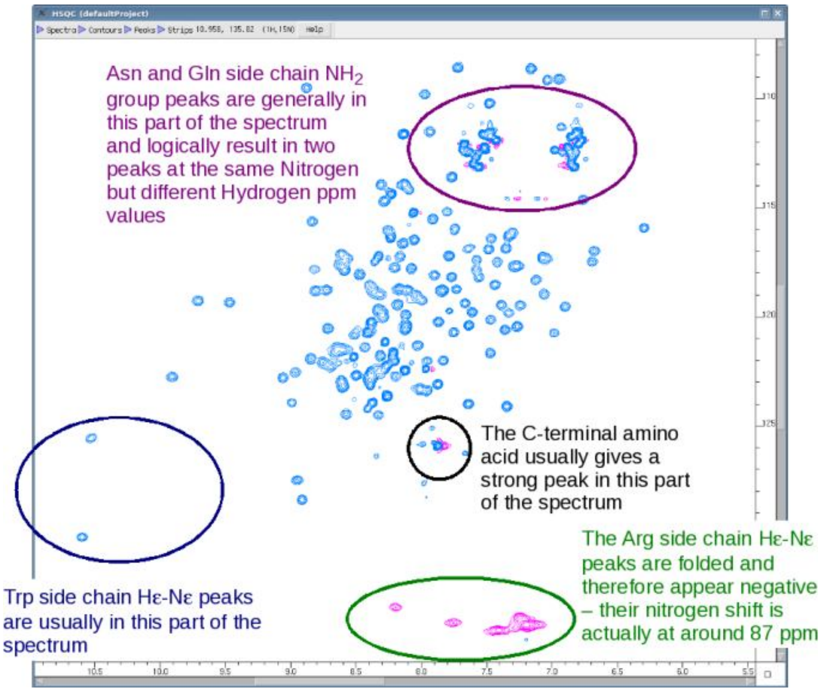


Figure 37: The four sections of a 2D-NMR experiment.
(from http://www.cryst.bbk.ac.uk/PPS2/projects/schirra/html/2dnmr.htm).



Asn and Gln side chain NH_2 group peaks are generally in this part of the spectrum and logically result in two peaks at the same Nitrogen but different Hydrogen ppm values

The C-terminal amino acid usually gives a strong peak in this part of the spectrum

Trp side chain Hε-Nε peaks are usually in this part of the spectrum

The Arg side chain Hε-Nε peaks are folded and therefore appear negative – their nitrogen shift is actually at around 87 ppm

Figure 38: A typical HSQC spectrum of a protein (http://www.protein-nmr.org.uk/solution-nmr/).

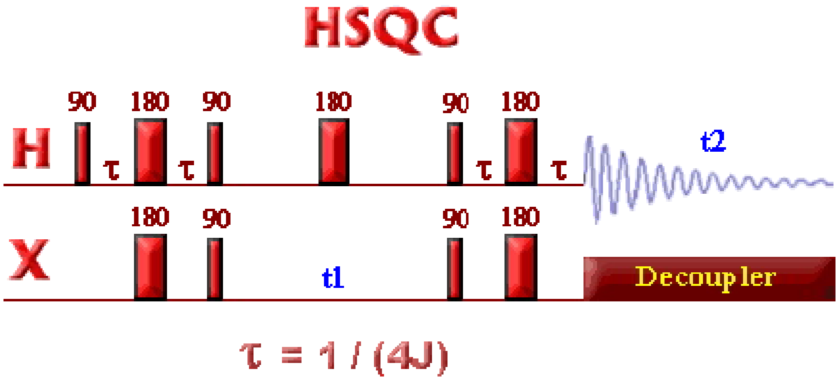


Figure 39: Pulse sequence of an HSQC experiment
(from http://www.chem.queensu.ca/facilities/nmr/nmr/webcourse/hmqc.htm).

Nevertheless, 2D experiments are still not sufficient for the assignment and the determination of the protein structure, because of crowding and necessity of finding more complex links. A third dimension in then added to a 2D experiment. The 3D experiments in addition to reduce signal overlaps can provide different information depending on the magnetization transfer and on the coupling (scalar -for chemical bonds- or dipolar -for space proximity-) exploited (figure 40 and 41). The main drawback associated with these experiments is the long acquisition time that can challenge the stability of delicate samples, but methods for decreasing it have been developed (such as the *Non-Uniform Sampling*, NUS).

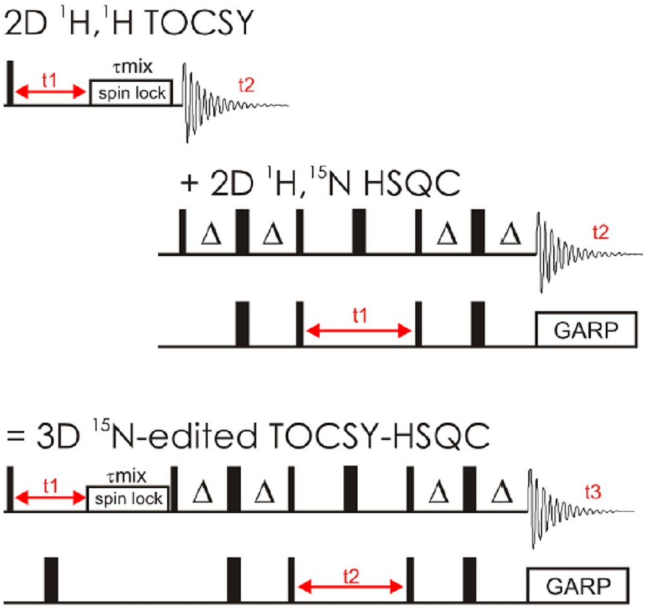


Figure 40: Scheme describing the creation of a 3D NMR experiment (from http://www.bioc.aecom.yu.edu/labs/girvlab/nmr/course/COURSE_2012/3DNMR_lecturenotes.pdf).

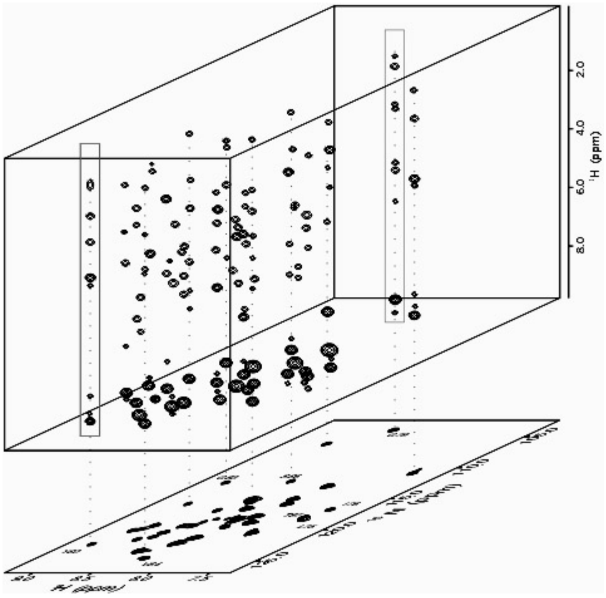


Figure 41: Aspect of a 3D NMR spectrum. The interpretation of the spectrum is based on 2D projections (from http://www.bioc.rice.edu/bios576/nmr/nmr.html).

The structural information to fold a protein can be found in the 3D NOESY experiments (that retain also sequential information) (figure 42). In these experiments, the magnetization is transferred not through the scalar coupling (covalent bond) but through the dipolar coupling exploiting the NOE effect, generating a signal when the excited atoms are at 5 Å or less of distance. This information becomes distance restraints to fold the structure.

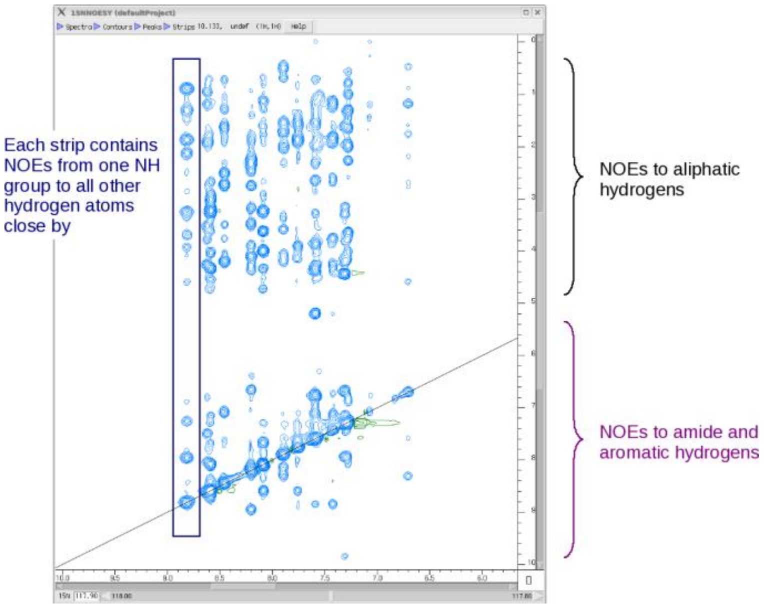


Figure 42: 2D projection of a portion of ^{15}N-HSQC-NOESY (from http://www.protein-nmr.org.uk/solution-nmr/spectrum-descriptions/15n-noesy-hsqc/).

The choice of 2D- and 3D-NMR experiments for proteins is rather wide. For the ABACUS dataset, the choice has been made based on the number of magnetization transfers and relaxation losses: so, the CBCA(CO)NH and HNCA have been preferred toward the less sensitive HNCACB (104), or the HBHA(CO)NH to HC(C)CONH. Anyway, new experiments can be added. The dataset can be divided in two parts: the first one includes the ^{15}N-rooted experiments and it is called FAWN (Fragment Assignment With NOE), while the second one the ^{13}C-edited experiments that, together with FAWN, constitutes the full ABACUS dataset (figure 43).

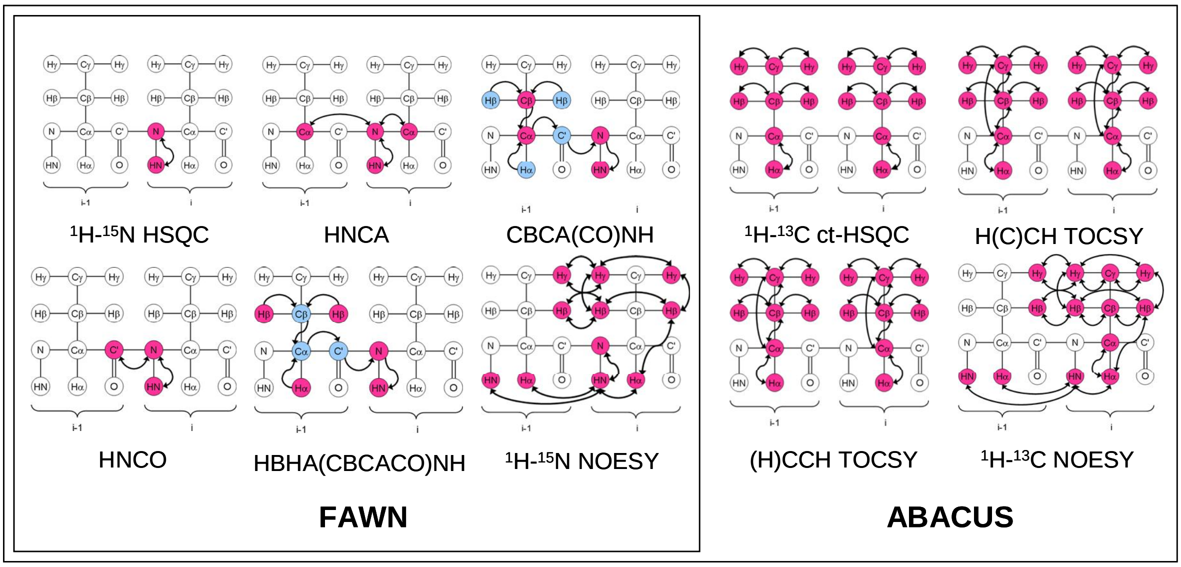


Figure 43: Definition of the ABACUS dataset and of the magnetization transfers involved in the different experiments. (Pictures from http://www.protein-nmr.org.uk/solution-nmr/).

6.6 Semi-automated resonance assignment and structure calculation according to the ABACUS procedure

The ABACUS procedure (115) is based on BACUS (117), an automated Bayesian analysis procedure that achieves protein structures starting from unassigned NOESY restraints. It groups many programs with different functions: predicting the amino acid type, assembling the sequence, managing the peaklists, preparing the format file for other software, etc…(115) (118). All the data needed are the primary sequence and the peaklists generated from the manual peak picking of the spectra from the minimal dataset: a list of unassigned ^{1}H, ^{13}C and ^{15}N resonances divided in Peptide Bond fragments (PB-fragments, figure 44) and ^{15}N- and ^{13}C-edited 3D NOESY peak lists. ABACUS do not need the user to label the side chains, since it is an operation it performs by itself. It can also generate expected peak lists for the NOESY spectra, to help in the analysis, and files in the suitable formats for other programs, such as CYANA (112) for structure calculation and CNS (119) for refinement.

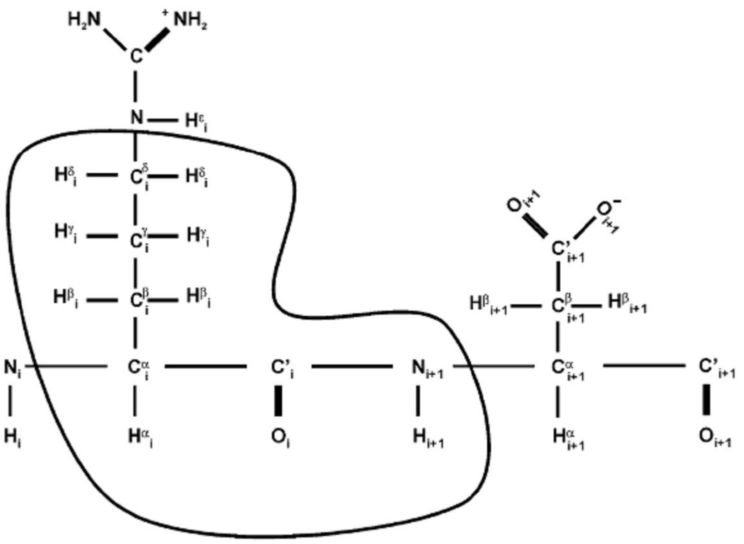


Figure 44: Description of PB-fragment (113).

The starting point consists in determining the amino acid type probability of each fragment with the program TYPESYST. This software takes advantage of the statistics collected in the BMRB, since the experimental ^{1}H, ^{13}C and ^{15}N resonances are strongly dependent on the amino acid type. The following step concerns the identification of the NOEs by matching their chemical shifts with the spin system peaks, with BACUS. The NOE links are divided in different categories (COSY peaks, TOCSY peaks, intra-residue connectivities, etc…) and the likelihoods of these links related with each category are determined with a database of ultra-high resolution x-ray structures, with a resolution better than 1.0 Å and with a pairwise sequence similarity lower than the 25% (118).

After these two steps, the PB fragments have been created but not sequentially connected yet: the two programs; FINDSEQ and FMC, links the fragments (118). FINDSEQ creates "clouds of fragments" relying on the Bayesian inference. FMC (Fragment Monte Carlo) assembles the fragments trying to reach a minimum in a pseudo-energy function, using a thermodynamic analogy. It substitutes the program LINKMAP, present in the previous version of ABACUS (115), and it addresses some of its issues: the possibility of broadening the positions available to a number bigger than the sequence, in order to accommodate the His-Tag, contaminants and the not assigned fragments, and a reduced risk of stucking the simulation in a local minimum. In the Monte Carlo stochastic sampling, a pseudo-energy profile is described for each assignment: the starting point is a "highly energetic" assignment, which is modified at every iteration by swapping the fragments in the sequence. Only the swaps that are below a certain pseudo-energy level, lower at every iteration, are accepted and the simulation can be considered accomplished when the pseudo-energy profile is not modified anymore. This method is a variation of the Monte Carlo method called Multi-Canonical (MUCA).

A new feature has been included in the latest version of ABACUS (118), to obtain a rapid sequence assignment for titrations and relaxation experiments: it is called FAWN (Fragment Assignment With NOE) (113). It is an algorithm analogous to ABACUS but do not rely on the ^{13}C-rooted experiments, basing its calculations only on the peptide fragments excluding the information from the side chains, from the γ atoms to the end. Instead, it makes use of the HNCA peak list to sequentially assign the fragments, in addition to the NOE information from the ^{15}N–NOESY.

6.7 Non-Uniform Sampling and Three-Way Decomposition for high-resolution 3D NMR spectra

Sampling a high number of points in a FID results in the improvement of the resolution, but the price to pay is an increase in the experimental time of the spectrum. In particular, in three-dimensional spectra, the time dedicated to the indirectly detected dimensions affect sensibly the total acquisition time. The idea that the acquisition time can be decreased and the final NMR spectra can be reconstructed to a full dataset just based on a selected part of data relies on high redundancy of these data: peaks in 3D spectra are about few thousands and can be described by a few dozens of numbers each, among the millions of points that are collected during a traditional experiment. To override the problem of long experimental time while maintaining the resolution, many acquisition schemes and reconstruction method have been proposed: an interesting one was elaborated by Orekhov *et al.* (114) and included in the method developed at the SGC Toronto. The acquisition method is defined as *non-uniform sampling* (NUS) and the reconstruction scheme is called *three-way decomposition* (TWD), then implemented in the SGC procedure as *multidimensional decomposition* (MDD).

This time-saving scheme considers a 3D experiment as a grid of t1- and t2-values, while the t3-values are acquired just for selected pairs of t1and t2: this data set is referred as *sparse*. The final outcome is reduction of the 70% of the experimental time, but the spectral width is preserved and the resolution as well. Anyway, the sparsely recorded spectrum is full of grid points without experimental data and the FFT algorithm for Fourier transform cannot be applied: the missing points are therefore predicted in the time domain and the grid is filled. The resulting data set can now be processed as a traditional one, with the standard Fourier transform algorithm (figure 45).

The choice of the combinations of t1- and t2-values is not random in the NUS acquisition: it has been demonstrated that sampling at the beginning of the spectra yields a higher sensitivity (120). The fundamental requirement of TWD is that no FID should be missing for any t1- and t2-value. Signals in data set at least with three dimensions can be described by direct products of 1D vectors and the resulting decomposition is unique: this is how TWD handles the data. This concept can be mathematically described in this way:

$$\min G_{ijk} \sum_{ijk} \left| S_{ijk} - \sum_{m=1}^{M} a^{m} \cdot F1_{i}^{m} \cdot F2_{j}^{m} \cdot F3_{k}^{m} \right|^{2} + \lambda \sum_{m=1}^{M} \left(a^{m}\right)^{2}$$

Where S is the 3D input data set (the experimental time-domain spectrum), which can be approximated by 3D components described by the 1D tensor products F1, F2, and F3 along the three dimensions. S and F1, F2 and F3 are made of discrete points and are a 3D matrix and 1D vectors respectively (63). The aim is optimizing the fitting by varying the parameters to minimize the function. The first sum over *m* enumerates the M components, while the second one flattens large differences in the size of the components Tikhonov's factor λ (64) and it is normalized by the T; a^{m} is the amplitude of a component; $F1_{i}^{m}$, $F2_{j}^{m}$ and $F3_{k}^{m}$ are numbers representing the normalized shapes for a component; the indices *i*, *j* and *k* are grid points along the three dimensions (65).

The innovation introduce by Orekhov et al. is the presence of the matrix G, where the elements can assume the value $G_{ijk} = 1$ for a recorded and $G_{ijk} = 0$ for a skipped data point S_{ijk} in the NMR spectrum. In this way, not recorded spectra will not contribute to the penalty function. It is important to point out that S, the input, is sparse, but $F1^{m}$, $F2^{m}$ and $F3^{m}$, the output, are vectors with complete data: this is possible only if S does not lack an entire plane. By multiplying the shapes $F1^{m}$, $F2^{m}$, $F3^{m}$ and the amplitudes a^{m} the full spectrum S^{*} can be achieved, an optimal approximation of the experimental spectrum S.

The degree of *sparsing* R is defined by the ratio between the number of (t1, t2) combinations used and those in the full spectrum. R is a critical factor, together with the number of components M, which can be easily estimated from the number of HN-H groups in the ^{15}N-^{1}H HSQC spectrum: usually, a sparsing percentage of 30% is enough for the reconstruction and it is the percentage used in the ABACUS data set. A comparison between NUS/MDD and traditional spectra is reported below (figure 46).

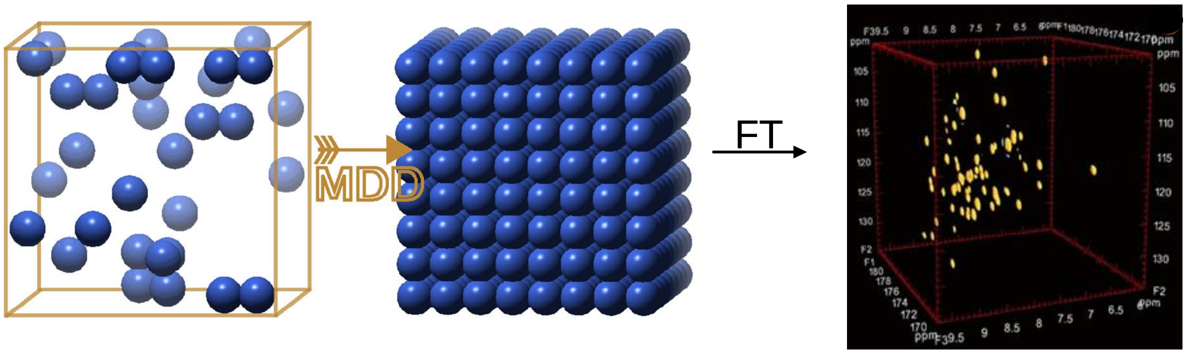


Figure 45: Steps of the NUS protocol. The 3D experiment is acquired as a *sparse* dataset, which is filled with the missing points with the MDD. The fully reconstructed dataset can now be handled as a normal one, ready for operations like the Fourier Transform (FT), to generate the final 3D spectrum. (Pictures reproduced from http://www.extend-nmr.eu/mdd.htm and http://www.rsc.org/Education/EiC/issues/2008september/ThePowerOfNMRInTwoAndThreeDimensions.asp).

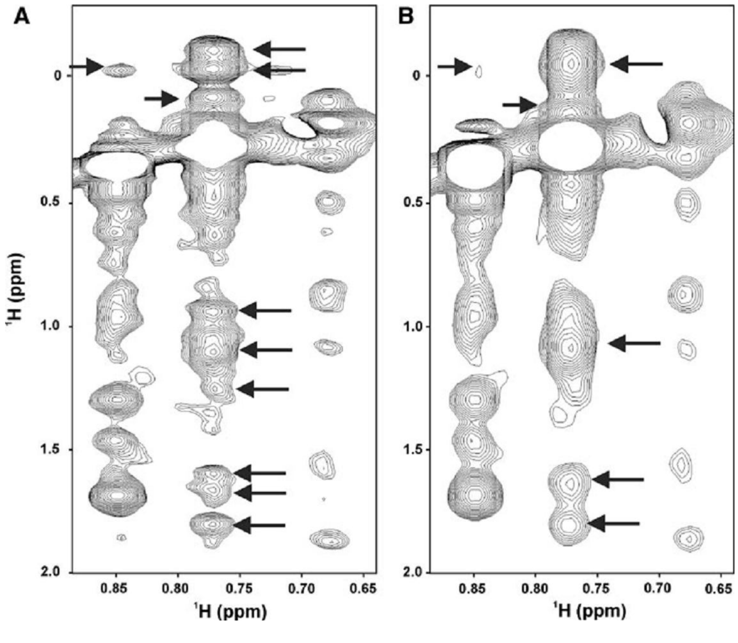


Figure 46: Comparison of ^{13}C-edited NOESY spectra obtained with NUS/MDD and fully sampled of the protein Atu0922 from *Agrobacterium tumefaciens*, 121 amino acids, highlighting the difference in resolution. The two spectra have been acquired in the same amount of time. **(A)** Spectrum acquired with NUS and processed with MDD, 300 points in the ^{1}H indirect dimension; **(B)** Spectrum achieved by reprocessing the spectrum in (A) but with half the number of indirect complex points with the Fourier transform. The arrows indicate peaks that are affected by the resolution. From (113).

7. AIMS

Our aim was determining the structure and characterizing the backbone dynamics of two members of the Myb/SANT family. For one of them we perform additional binding studies.

The structure of the R1 repeat of the DNA-binding domain of hDMP1

^{15}N relaxation measurements of R1

The first domain belongs to the protein hDMP1, a tumour suppressor with transcriptional activity. Its DNA binding domain can be divided in three Myb-like repeats (R1, R2 and R3), where R2 and R3 are the units supposed to interact with the nucleic acid. R1 is considered to be accessory, but actually its functions and interactions are still unknown. We wanted to solve its structure applying the fast NMR methods developed at the SGC Toronto and to analyze its backbone dynamics.

The structure of the SANT2 domain of NCoR2

Preliminary interaction studies of SANT2 with the H4 histone tails

^{15}N relaxation measurements of SANT2

NCoR2 presents two SANT domains, defined as DAD (“Deacetylase Activation Domain, SANT1) and HID (“Histone Interacting Domain”). The structure of DAD has already been solved (63) and we aimed to solve the structure of the HID with the SGC Toronto methods and to analyzed its backbone dynamics. Moreover, we planned to start preliminary investigations of the binding of SANT2 with the H4 histone tails based on the results of Yu *et al.* (101) and Hartman *et al*. (102), to map the interaction surface by titrations monitored with ^{1}H-^{15}N-HSQC.

8. MATERIALS AND METHODS

8.1 NMR structure determination of the R1 Myb-like repeat of hDmp1

The NMR spectra for structure determination were collected at 298K on Bruker Avance 600 and 800 MHz, with cryoprobes. The NMR sample contained 1mM R1 of hDMP1 domain and 10 mM sodium phosphate buffer at pH6.5, 400 mM NaCl, 0.01 mM $ZnSO_4$, 10 mM DTT, 1mM benzamidine, 0.01% (v/v) sodium azide, 5% (v/v) D2O.

The NH-rooted spectra were acquired at 600 MHz (two dimensional ^1H-^{15}N HSQC, three dimensional CBCA(CO)NH, HNCO, HNCA; HBHA(CO)NH, ^1H-^{15}N NOESY), while the CH-rooted ones at 800 MHz (the two dimensional ^1H-^{13}C constant time HSQC, three dimensional, ^1H-^{13}C aliphatic NOESY, ^1H-^{13}C aromatic NOESY, H(C)CH-TOCSY, (H)CCH-TOCSY), with the three dimensional spectra acquired with non-uniform sampling (NUS) protocol (121) (114). The two dimensional spectra were processed with NMRPipe (107), while the non-uniformly sampled three-dimensional spectra were processed with the multidimensional decomposition algorithm of the MDDGUI software (122). Sparky was used for the manual peak picking (111); the ABACUS method for the assignment of the ^1H, ^{15}N and ^{13}C resonances (113) and TALOS to calculate the restraints for the backbone torsion angle from the chemical shifts (107). Automated NOE assignment and structure calculation were performed with CYANA (112). The 20 lowest energy structures were refined by restrained molecular dynamic simulation in explicit solvent with the software CNS (119). The final structures were inspected with MOLMOL (123) and Pymol (124) and evaluated with the PSVS server 1.4 (125).

8.2 Structure deposition of the R1 Myb-like repeat of hDmp1

The structure coordinates and restraints for the R1 repeat of the Myb-like domain of hDMP1 have been deposited in the Protein Data Bank under the code 2LLK. The chemical shifts have been deposited in the BioMagResBank (University of Wisconsin) with the accession no. 18051.

8.3 Relaxation measurement of the R1 repeat of the DNA-binding domain of hDmp1

The NMR experiments acquired for analysis of backbone dynamics of the domain were conventional experiments based on 2D ^{1}H-^{15}N-HSQC adapted to measuren T_1 and T_2 relaxation times and ^{15}N{^{1}H}-NOE (126), acquired at 600 and 800 MHz, on Bruker Avance instruments. The relaxation delay times for T_1 were set as follows: 5, 65, 145, 246, 366, 527, 757, 1148, 1500 ms; CPMG pulse trains of the duration of 34, 51, 68, 85, 102, 119, 136 ms were used for T2. Two experiments for the measurement of the heteronuclear NOE were collected in an interleaved manner at 800 MHz and one at 600 MHz, alternating one experiment with proton presaturation with one without it. The recycle delay was 3 s for T1 and T2 experiments and 3s of presaturation was used for the heteronuclear NOE experiments.

8.4 NMR structure determination of the SANT2 domain of NCoR2

The NMR sample contained 1mM SANT2 domain and 25 mM sodium phosphate buffer at pH 6.5, 200 mM NaCl, 10 mM DTT, 1mM benzamidine, 0.01% (v/v) sodium azide, 5% (v/v) D2O. The spectra acquisition and processing, the assignment and the structure calculation were performed in the same way as reported for the R1 Myb-like repeat of hDmp1.

8.5 Structure deposition of the SANT2 domain of NCoR2

The structure coordinates and restraints for the SANT2 domain have been deposited in the Protein Data Bank under the code 2LTP. The chemical shifts have been deposited in the BioMagResBank (University of Wisconsin) with the accession no. 18492.

8.6 Relaxation measurement of the SANT2 domain of NCoR2

The NMR experiments acquired for analysis of backbone dynamics of the domain were conventional experiments based on 2D ^{1}H-^{15}N-HSQC adapted to measure T_1 and T_2 relaxation times and ^{15}N{^{1}H}-NOE (126), acquired at 500 and 600 MHz, on Bruker Avance instruments. The relaxation delay times for T_1 were set as follows: 5, 65, 145, 246, 366, 527, 757, 1148, 1500 ms; CPMG pulse trains of the duration of 16.33, 32.65, 48.97, 81.61, 114.25, 130.57, 163.3 ms at 600 MHz and 16.8, 33.6, 50.4, 67.2, 84, 100.8, 117.6, 134.4, 151.2, 168 ms at 500 MHz were used for T_2. Two experiments for the measurement of the heteronuclear NOE were collected in an interleaved manner for each NMR frequency, alternating one experiment with proton presaturation with one without it. The recycle delay was 3 s for T1 and T2 experiments and 3s of presaturation was used for the heteronuclear NOE experiments.

8.7 Expression and purification of the SANT2 domain of NCoR2 for the titration experiments

The ^{15}N-labelled SANT2 domain for the titration was expressed with the auto-induction method, developed by Studier (127). This method exploits the induction of the Lac operon by lactose, leading to the protein expression, after a latency time where the bacteria grow metabolizing all the glucose in the medium. In fact, *Escherichia coli* is a diauxic bacterium, which means that it can metabolized two kinds of sugars (glucose and lactose), but, in presence of both, it prefers glucose because it is ready for its metabolism. Moreover, glucose represses the lactose metabolism. In conditions of glucose deprivation , the bacterium can metabolize lactose, but it needs to activate the *lac* operon, which constitutes the ensemble of proteins required for lactose metabolism. The stimulation of the lac operon leads also to the derepression of the viral T7 polymerase, which starts the transcription of the protein of interest in the plasmid.

This was the composition of the expression medium used for auto-induction protocol: 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM $^{15}NH_4Cl$, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 0.2X metals (10 µM Fe + other 9 metals), 0.5 % glycerol (54 mM), 0.05% glucose (2.8 mM), 0.2 % α-lactose (5.6 mM), 0.25% aspartate (18.8 mM), 100 µg/ml kanamicine. 100 µl of bacteria from a frozen aliquote were cultivated in 3 ml of LB medium, for 24 h, then transferred to 50 ml of auto-induction medium for other 24 h, at 37°C, 170 rpm, in a thermostated orbital shaker. The 50 ml of the bacterial colture were added to 450 ml of the same auto-induction medium and cultivate for 24 h at 37°C, 170 rpm. The bacteria were harvested by centrifugation at 7000 rpm, 4°C, for 20 min. The bacterial pellet, stored at -80°C, was resuspended 25 ml of lysis buffer (15 mM imidazole, 50 mM Tris, 500 mM NaCl, 10 nM $ZnSO_4$) and the cells were lysed by sonication (3 s pulse, 3 s pause, for 5 min 30 s). The cell debris were removed by centrifugation (at 4°C, 12000 rpm, 25 min). A suspension of 3 ml of NTA-Nickel beads (50% beads, 50% buffer) was added to the supernatant in a falcon tube and rocked for 1 h at 4°C. The tube was centrifuged for 5 min at 4°C, at 1800 rpm, to spin down the beads. The supernatant was removed, then 12 ml of lysis buffer were added and the tube was rocked again. This cycle was repeated twice with the lysis buffer with 15 mM imidazole and twice with the one with 30 mM imidazole. After the last cycle, without centrifugation, the suspension was loaded on a plastic column for gravity elution. The falcon tube was cleaned with 5 ml of the 30 mM imidazole lysis buffer, which were loaded in the column too. When the liquid reached the limit of the beads front, the elution was carried out with 5 ml of the 500 mM lysis buffer. 1 mM benzamidine and 10 mM DTT were added. The following day, the buffer was changed with 25 mM Na phosphate, 200 mM NaCl, 0.01 mM $ZnSO_4$, 10 mM DTT, 1 mM benzamidine, 0.01% NaN_3, 5% D_2O.

8.8 Preliminary titrations of the SANT2 domain of NCoR2 with a H4 histone tail peptide

The titration of the ^{15}N-labelled SANT2 domain with the unlabeled H4 was followed with ^{1}H-^{15}N HSQC, collected with a Bruker Avance 500 MHz, with a cryogenic microprobe. The 24 amino acid H4 histone tail peptide sequence was: SGRGKGGKGLGKGGAKRHRKVLRD.

The amplitude of the chemical shift perturbation was quantified applying the following formula:

$$\Delta\delta(NH) = \sqrt{\left[\Delta\delta\left({}^{1}H\right)\right]^{2} + \left[\Delta\delta\left({}^{15}N\right)\cdot 0.154\right]^{2}}$$

The titrations experiments were performed twice, using 40 µl/microtube of 60 µM ^{15}N-labelled SANT2 domain in the same NMR buffer used to solve the structure. The first time the molar ratio of H4 histone tails/SANT2 tested were: 0, 2, 5, 10, 20, 30, 40, 50; the second time: 0, 5, 14, 20, 35.

9. RESULTS AND DISCUSSION

9.1 Solution structure of the R1 repeat of the DNA-binding domain of hDmp1

The 51 amino acid R1 unit of hDmp1 presented the typical Myb structure. It has three α-helices (H1: 228-243; H2: 245-254; H3: 256-267), where the second and the last helix form a Helix-Turn-Helix (HTH) motif. In the so called “tryptophan cluster”, only a tryptophan is present (W246), the other two are substituted by Y227 and C264. The statistics for the NMR structure are reported in table 4, its HSQC in figure 47 and its structural ensemble in figure 48.

The surface charge distribution is typical of a Myb domain, as illustrated in figure 49: consequently, the SGC Toronto tested the DNA-binding ability of R1, R2-R3 and R1-R2-R3 toward different DNA sequences (data not reported). As regards R1, it did not bind to any of the sequences tested, while the fragment R2-R3 and the full domain R1-R2-R3 did bind to most of them. This is in line with the expectations about R1 and a possible explanation about its missing DNA-binding capacity can be proposed just looking at the amino acid composition of its third helix. Applying a reductionist approach, the binding with DNA can be divided in hydrophobic/ hydrophilic interactions with the DNA backbone and basis and electrostatic interactions with the charged phosphate groups. In the analysis performed by Hanaoka *et al*. (128) about the interaction of the third helix of the Myb domain of hTRF1 and hTRF2 with the DNA backbone (sugars and phosphate groups) and basis, 16 and 14 amino acids were involved respectively. The comparison with R1 just in terms of amino acid identity shows that in both the complexes 8 amino acids, but different, in the two cases, are conserved (the amino acids written in red in figure 50). Analyzing the problem in terms of charged amino acids, R1 has a total positive charge of +2 in the last helix, while it is +3 for hRAP1, another Myb domain that does not bind DNA but proteins, and +4 for hTRF1 and hTRF2, which have almost a helix turn more than R1 with two positively charged amino acids (table 5). In conclusion, R1 has half of the positive charge and half of the amino acids required for the interaction with the backbone compared with known DNA-binding Myb domain, like in hTRF1 and hTRF2, explaining its incapacity of interacting with DNA.

Table 4: Statistics of the R1 (220-274) NMR structure[a].

Conformationally-Restricting Distance Constraints	
Total	733
Intraresidue [i = j]	155
Sequential [|i – j| =1]	212
Medium Range [1< |i – j| <5]	203
Long Range [|i – j| >5]	163
NOE constraints per restrained residue[b]	13.3
Dihedral angle constraints (φ and ψ)	68
Total number of restricting constraints[b]	835
Total number of restricting constraints per restrained residue[b]	15.2
Restricting long-range constraints per restrained residue[b]	3.0
Number of structures used	20
Residual constraints violations[a,c]	
NOE distance violations/structure	
0.1-0.2 Å	4.15
0.2-0.5 Å	0.35
>0.5 Å	0
Dihedral Angle Violations/Structure	
1-10°	2.5
>10°	0
RMSD Values (Å)	
All residues (backbone atoms)	3.9
Selected structure[d] (backbone atoms)	0.4
All residues (heavy atoms)	4.3
Selected structure[d] (heavy atoms)	1.0
Structure Quality Factors	
PROCHECK (129) G-factors[d] Z score (φ and ψ/all dihedral angles)	1.89/1.24
PROCHECK (129) G-factors[d] mean score (φ and ψ/all dihedral angles)	0.40/0.21
MOLPROBITY (130) clash score (mean/Z-score)	10.57/-0.29
R/P/DP scores[e]	94.6%/94.5%/0.65
Ramachandran Plot Summary from Procheck[f] [%]	
Most favored regions	93.7
Additionally allowed regions	6.3
Generously allowed regions	0
Disallowed regions	0

a: Analysed for residue 220-274.

b: There are 55 residues with conformationally restricting constraints

c: Calculated for all constraints for the given residues, using sum over r^-6

d: Selected residues: with sum of φ and ψ order parameters > 1.8. Selected residue ranges: 227-242; 245-266.

e: Recall: percentage of peaks that are consistent with the structure, they are retrieved by the algorithm and are thus part of the query structure; Precision: percentage of expected peaks from the structure that are already included in the NOESY peak list; DP score: is a normalized F-measure statistic, DP > 0.7 correlates to structures having accuracies of < ~ 2 Å rmsd.

f: Residue selected based on: Dihedral angle parameter, with S(φ)+S(ψ)>=1.8. Selected residue ranges: 227-242; 245-266.

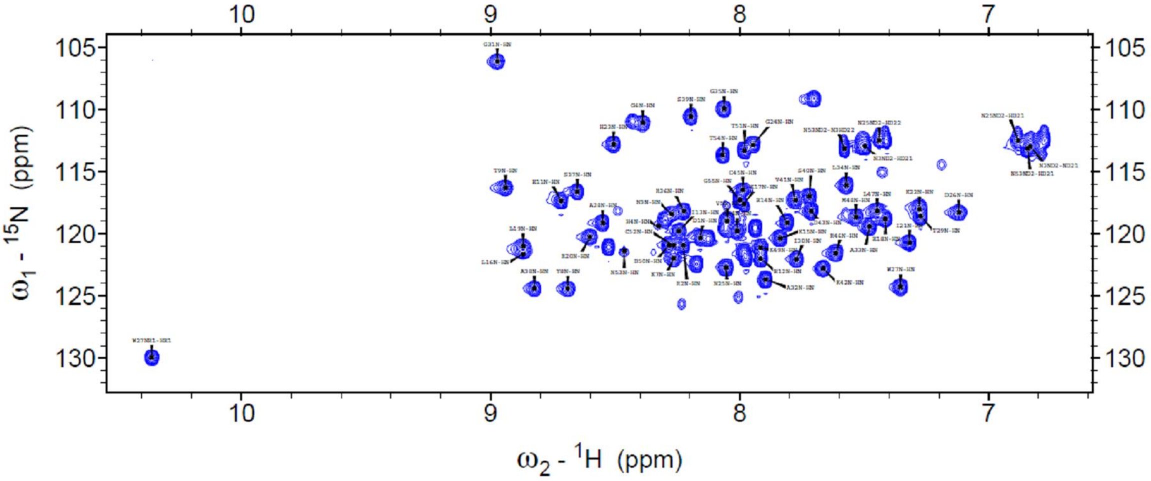


Figure 47: HSQC with the assignment of the R1 repeat of the DNA-binding domain of hDmp1. The unassigned peaks correspond to peaks from the histag.

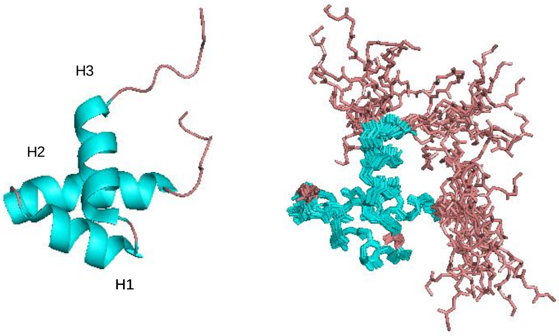


Figure 48: Structural ensemble of the R1 Myb-like domain of hDmp1.

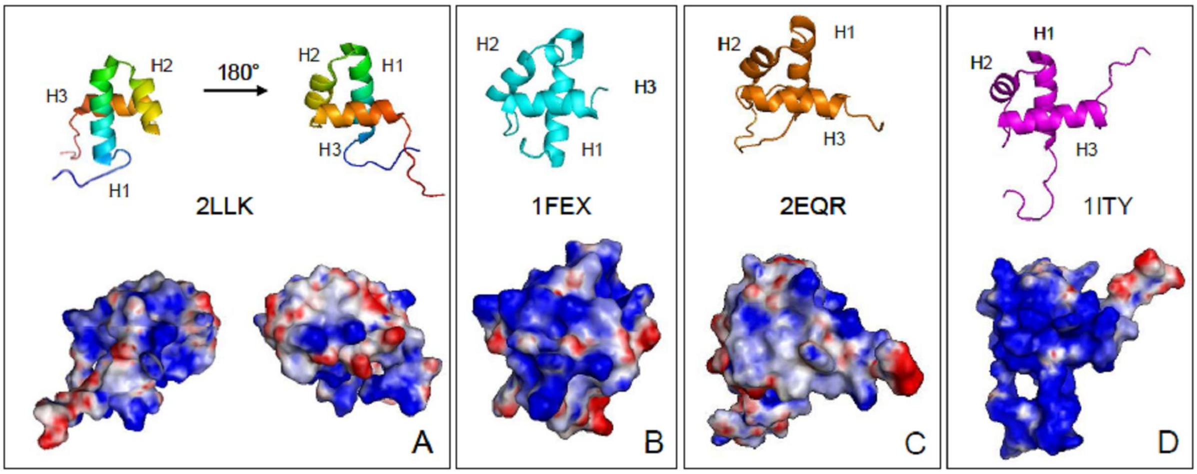


Figure 49: (A) R1 Myb of hDmp1; **(B)** Myb of hRAP1, not DNA-binding; **(C)** SANT of NCoR1; histone tail binding; **(D)** Myb of hTRF1, DNA-binding.

Table 5: Comparison of the amino acid composition of the third helix of some DNA-binding and not binding Myb domains with the Myb-like repeats of hDMP1.

Myb from														Total charge
hTRF1	Ser 417	Val 418	Met 419	Leu 420	**Lys 421**	*Asp 422*	**Arg 423**	Trp 424	**Arg 425**	Thr 426	Met 427	**Lys 428**	**Lys 429**	+4
hTRF2	Ala 484	Val 485	Met 486	Ile 487	**Lys 488**	*Asp 489*	**Arg 490**	Trp 491	**Arg 492**	Thr 493	Met 494	**Lys 495**	**Arg 496**	+4
hRAP1	Trp 46	Gln 47	Ser 48	Leu 49	**Lys 50**	*Asp 51*	**Arg 52**	Tyr 53	Leu 54	**Lys 55**	**His 56**	Leu 57		+3
hDMP1 R1	Ala 257	Ser 258	Ser 259	Val 260	**Lys 261**	*Asp 262*	**Arg 263**	Cys 264	**Arg 265**	Leu 266	Met 267			+2

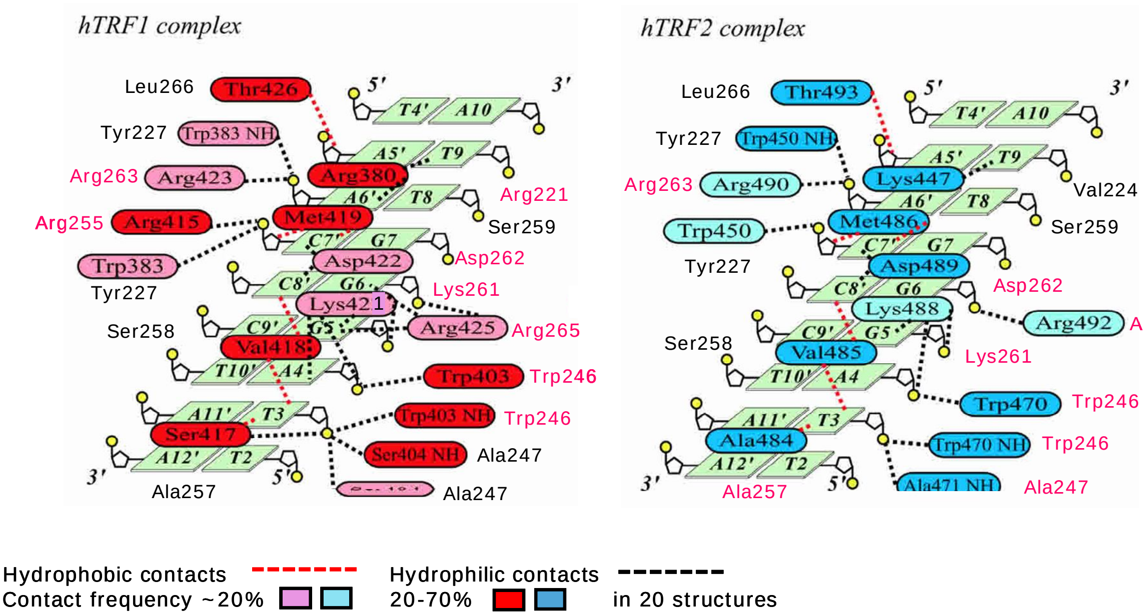


Figure 50: Representation of the interaction between hTRF1 or hTRF2 and DNA, with sugars, phosphate groups and bases. The not circled amino acids belong to R1 of hDmp1 and especially the ones in magenta are the same as hTRF1 or hTRF2. Adapted from (128).

9.2 ^{15}N relaxation measurement of the R1 repeat of the DNA-binding domain of hDmp1

The profile of the relaxation parameters measured for the Myb domain shows that the protein is well folded and it does not present any flexible region, besides the N- and the C-terminal ends. Five amino acids have been excluded from the analysis because of overlaps or bad signal: D220 and H223 in the N-terminal end, D269 and C271 in the C-terminal end, all placed in the flexible tails, except for K236. E230 was not assigned.Even if the 18 amino acid histag was not removed, being highly flexible, this has only a small influence on overall tumbling: a τ_c of 4.4 ns, compatible with a monomer was measured(monomeric state confirmed by gel filtration chromatography).

Figures 51, 52, 53 and 54 represent respectively the T1, T2, T1/T2 and the heteronuclear NOE.

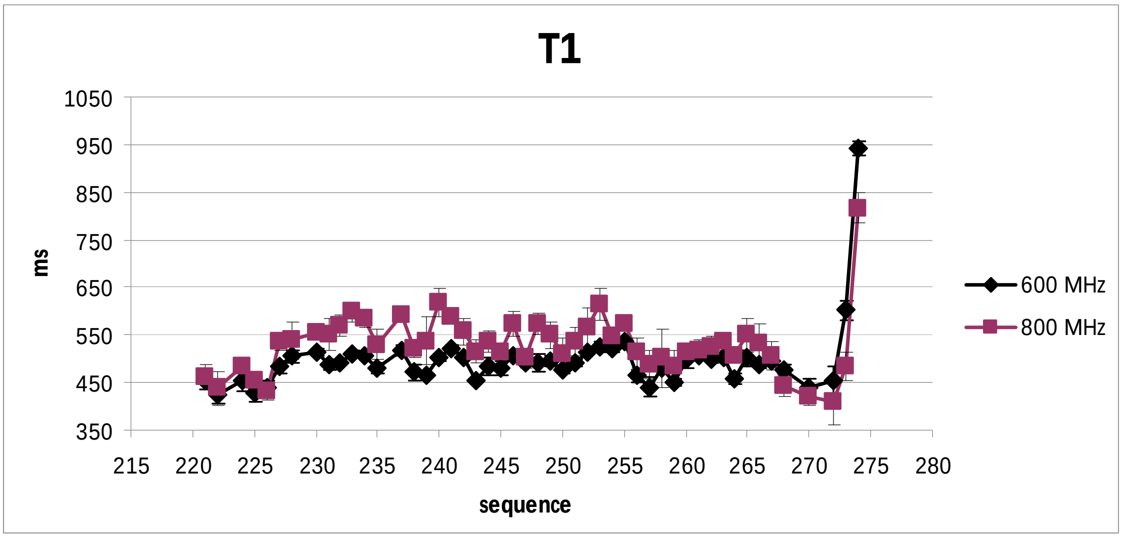


Figure 51: T1 of the R1 repeat of the DNA-binding domain of hDmp1.

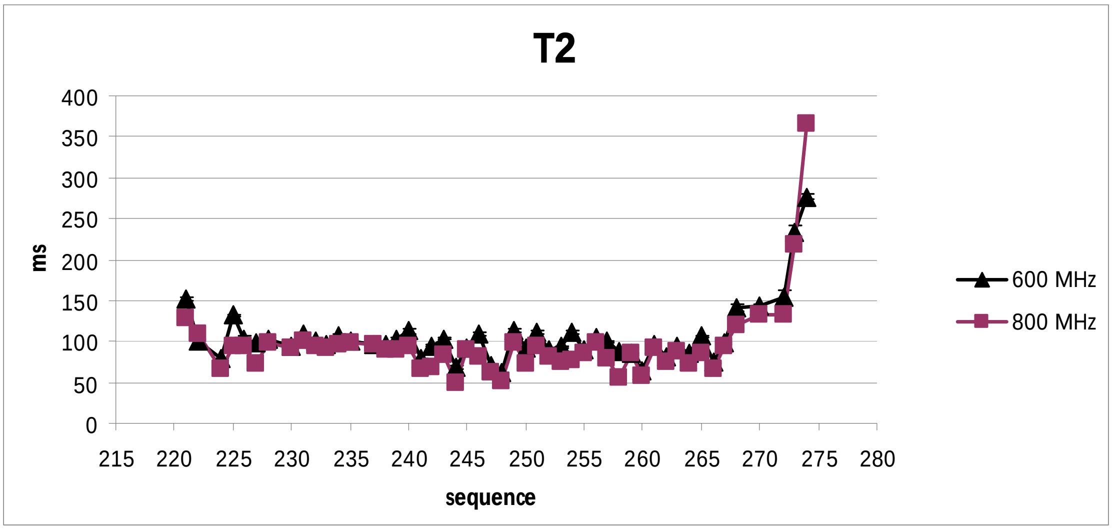


Figure 52: T2 of the R1 repeat of the DNA-binding domain of hDmp1.

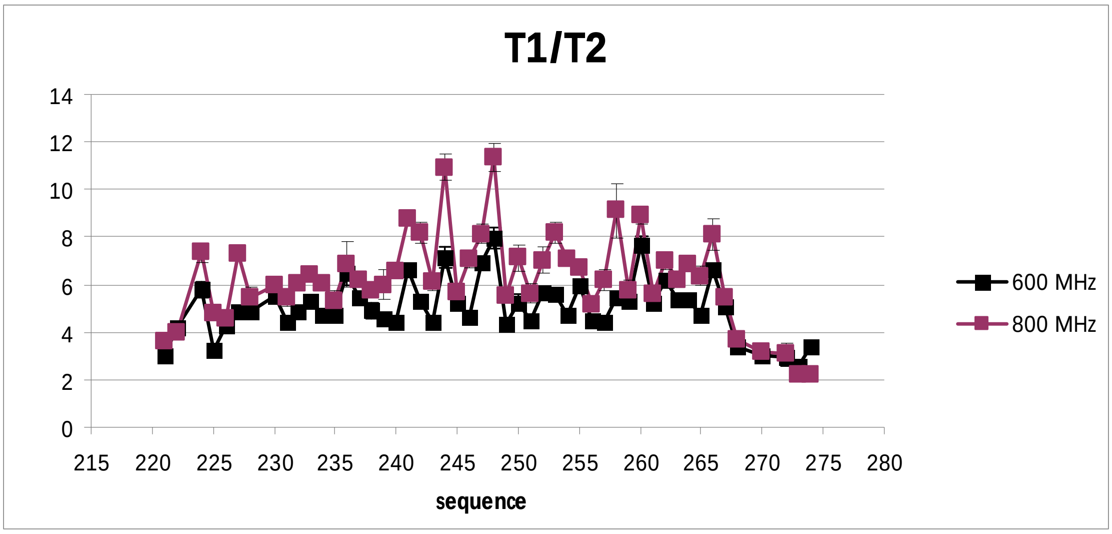


Figure 53: T1/T2 of the R1 repeat of the DNA-binding domain of hDmp1.

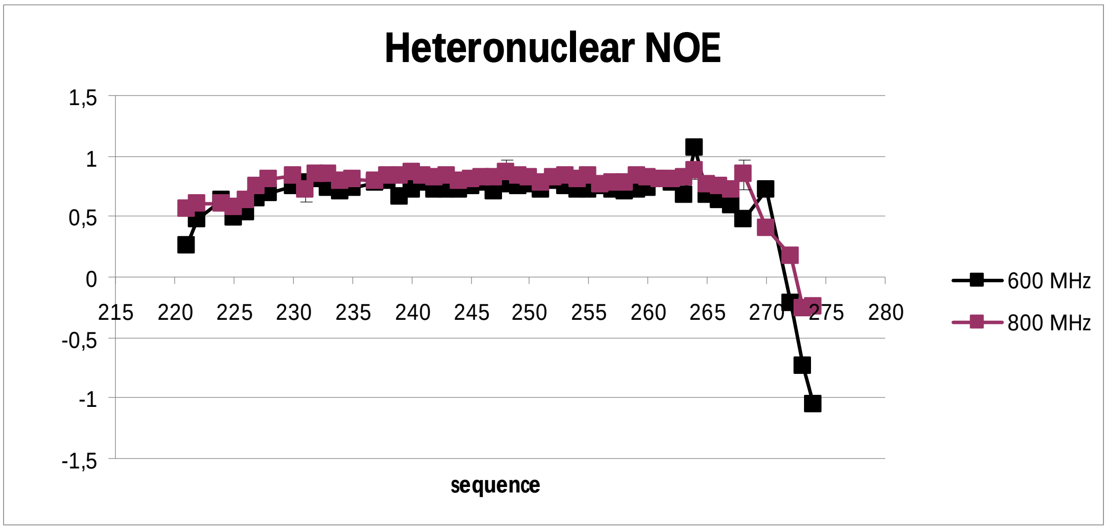


Figure 54: Heteronuclear NOE of the R1 repeat of the DNA-binding domain of hDmp1.

9.3 Solution structure of the SANT2 domain of NCoR2

The structure is made by four helices, wrapped around an hydrophobic core. The statistics for the structures are reported in table 6. The extension of the α-helices is: H1 615-631; H2 634-644; H3 647-655; H4 663-682. The C-terminal helix is longer and it sticks out from the central core. The loop connecting the third and the last helix shows a certain degree of disorder (figure 55).

During the NMR assignment (figure 56), the presence of a mutation was noticed, at position 621, where a glycine substituted a glutammic acid (E261G). Anyway, it is likely that such mutation did not affect the backbone structure, since the final outcome was in line with what expected, based on secondary structure predictions with JPred (131) made with and without the mutation, to check the expected length of the helices. The final structures were also compared to other SANT three dimensional structures found with the Dali server (132): the N-terminal helix, the one affected by the mutation, and the overall structure were apparently not perturbed by the mutation. Anyway, this mutation affects obviously the surface properties and its repercussion are discussed in the following section about the titrations with the H4 histone tails.

In the hydrophobic core the “tryptophan cluster” (48) (49) (61), typical of the Myb domains, is only partially conserved: comparing the SANT2 with the single Myb domain of hTRF1 (PDB code: 1ITY), two tryptophans are conserved (W615 and W634 in SANT2, corresponding to W383 and W403 for hTRF1) and just one for the SANT1 (W432, corresponding to W383 in hTRF1). The tryptophans 403 and 424 in hTRF1 are substituted by phenylalanine 451 and tyrosine 470 respectively in the SANT1, while in SANT2 tyrosine 653 corresponds to tryptophan 424 of hTRF1.

The comparison with its cognate domain, the SANT1 (DAD) of NCoR2 (63), shows evident differences. The first is the position of the helices: they have both four helices, but just the three most C-terminal of the DAD and the three most N-terminal of the HID are overlapping with a backbone RMSD of 2.5 Å, excluding the H0 of SANT1 and H4 of SANT2 (figure 57).

A second characteristic that are not shared is a wide hydrophobic groove on the surface of the DAD, which is not shared with the HID and generally with the Myb and SANT domains (figure 58). This cleft is positioned between the C-terminal half of H3 and the loop between H1 and H2: the angle between H1 and H3 is approximately

118° for the DAD (PDB: 1XC5), 103° for the HID (PDB: 2LTP), 93° and 87° for a Myb (PDB: 1ITY) and a SANT (PDB: 2EQR) domain respectively, measured with Pymol (124) (figure 59). It has already been demonstrated that this groove belongs partially to the interaction/activation surface for the enzyme Histone Deacetylase 3 (HDAC3) (63), an interaction that seems to involve just the SANT1 and not the SANT2.

The analysis of the electrostatic surface revealed the most interesting feature of the SANT2 domain: the charge distribution resembles the one of the Myb domains, not of the SANT ones. In fact, the third helix has a positive charge, while the first one has a negative one (figure 60). If it has DNA-binding properties, currently there are no clue about the possible consensus sequence. Some binding tests were performed with the same sequences used for the R1 repeat of hDmp1, with negative results (data not reported). The histone tail binding properties was also tested and the preliminary results are reported in the dedicated section.

Figure 55: Structural ensemble of the SANT2 domain of NCoR2 (PDB: 2LTP).

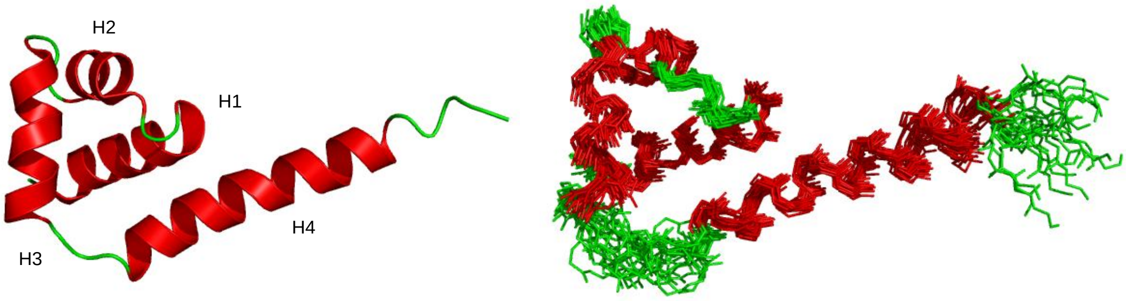


Table 6: Statistics of the SANT2 (615-685) NMR structure[a].

Conformationally-Restricting Distance Constraints	
Total	992
Intraresidue [i = j]	181
Sequential [|i – j| =1]	279
Medium Range [1< |i – j| <5]	335
Long Range [|i – j| >5]	197
NOE constraints per restrained residue[b]	15.0
Dihedral angle constraints (φ and ψ)	107
Total number of restricting constraints[b]	1099
Total number of restricting constraints per restrained residue[b]	16.7
Restricting long-range constraints per restrained residue[b]	3.0
Number of structures used	20
Residual constraints violations[a,c]	
NOE distance violations/structure	
0.1-0.2 Å	4.25
0.2-0.5 Å	0.85
>0.5 Å	0
Dihedral Angle Violations/Structure	
1-10°	9.35
>10°	0
RMSD Values (Å)	
All residues (backbone atoms)	1.3
Selected structure[d] (backbone atoms)	0.7
All residues (heavy atoms)	2.1
Selected structure[d] (heavy atoms)	1.1
Structure Quality Factors	
PROCHECK (129) G-factors[d] Z score (φ and ψ/all dihedral angles)	1.38/0.65
PROCHECK (129) G-factors[d] mean score (φ and ψ/all dihedral angles)	0.27/0.11
MOLPROBITY (130) clash score (mean/Z-score)	11.15/-0.39
R/P/DP scores	94.7%/86.1%/0.733
Ramachandran Plot Summary from Procheck[f] [%]	
Most favored regions	93.3
Additionally allowed regions	6.7
Generously allowed regions	0
Disallowed regions	0

a: Analysed for residue 615-685.

b: There are 66 residues with conformationally restricting constraints

c: Calculated for all constraints for the given residues, using sum over r^-6

d: Selected residues: with sum of φ and ψ order parameters > 1.8. Selected residue ranges: 615-631; 634-644; 647-655; 663-682.

e: Recall: percentage of peaks that are consistent with the structure, they are retrieved by the algorithm and are thus part of the query structure; Precision: percentage of expected peaks from the structure that are already included in the NOESY peak list; DP score: is a normalized F-measure statistic, DP > 0.7 correlates to structures having accuracies of < ~ 2 Å rmsd.

f: Residue selected based on: Dihedral angle parameter, with S(φ)+S(ψ)>=1.8. Selected residue ranges: 615-631; 634-644; 647-655; 663-682.

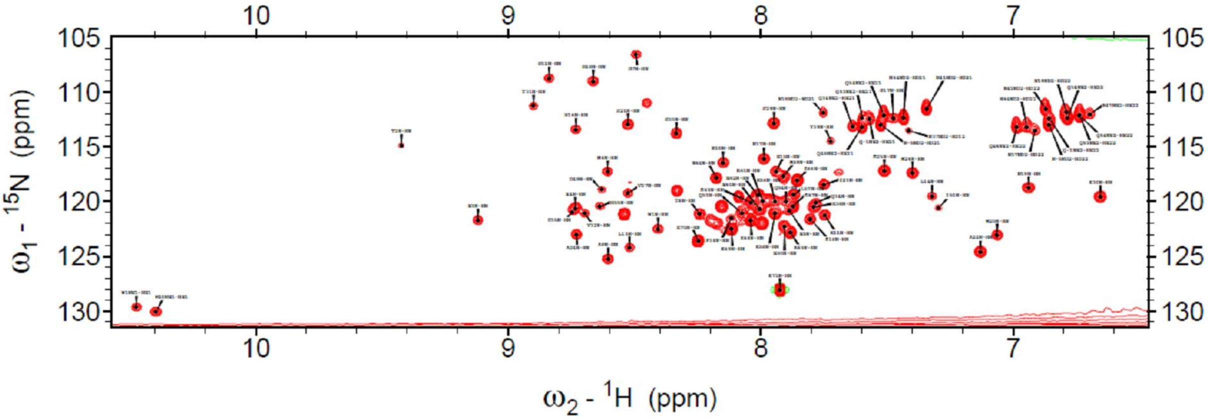


Figure 56: HSQC with the assignment of the SANT2 domain of NCoR2. The not assigned peaks correspond to peaks from the histag.

Figure 57: Superposition of the DAD domain (magenta, PDB: 1XC5) with the HID domain (yellow, PDB: 2LTP).

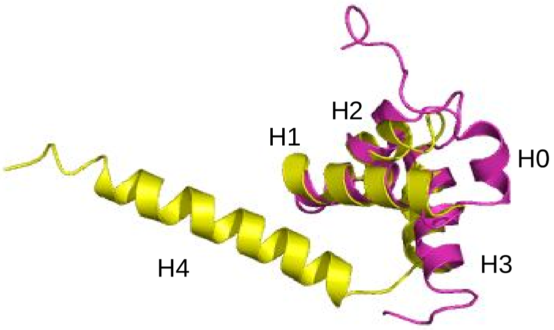


Figure 58: Comparison of the hydrophobic surfaces of the domains, coloured according to the Eisenberg's hydrophobicity scale (133). Top: secondary structure views, oriented in order to show the different position of the helix H3. Bottom: surface coloured based on the hydrophobicity; the yellow circle highlights the region between H3 and the loop H1-H2.

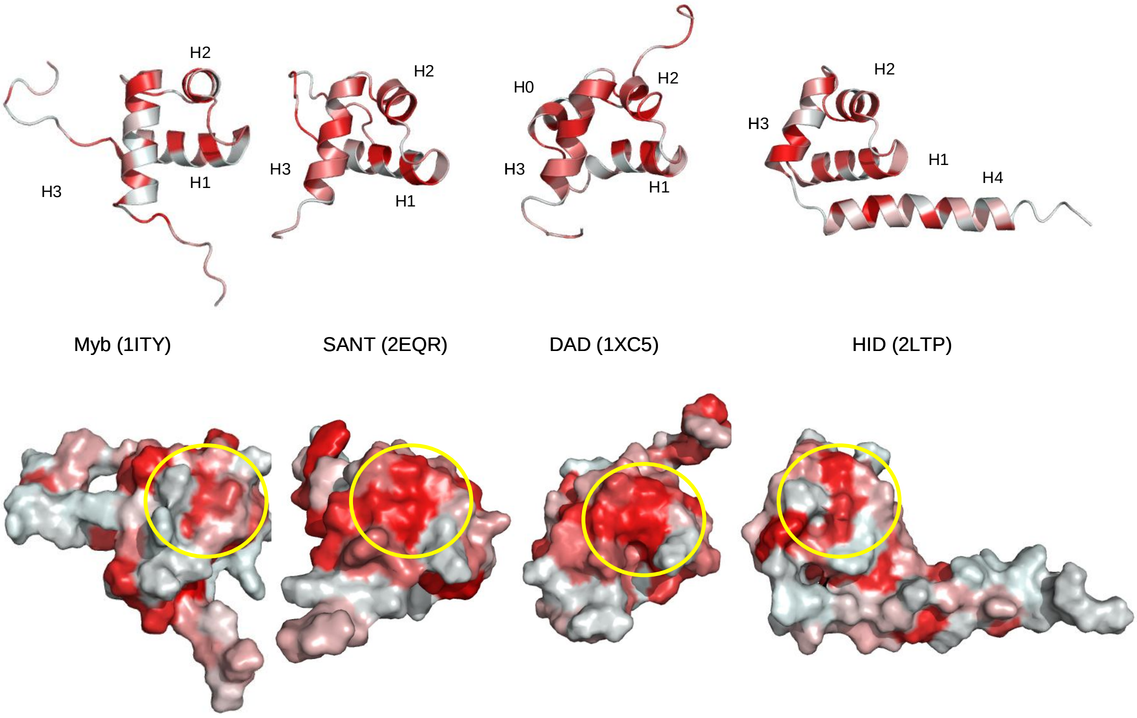


Figure 59: Superposition of a Myb domain (red; PDB: 1ITY), a SANT domain (cyan, PDB: 2EQR), DAD (magenta, PDB: 1XC5), HID (yellow, PDB: 2LTP).

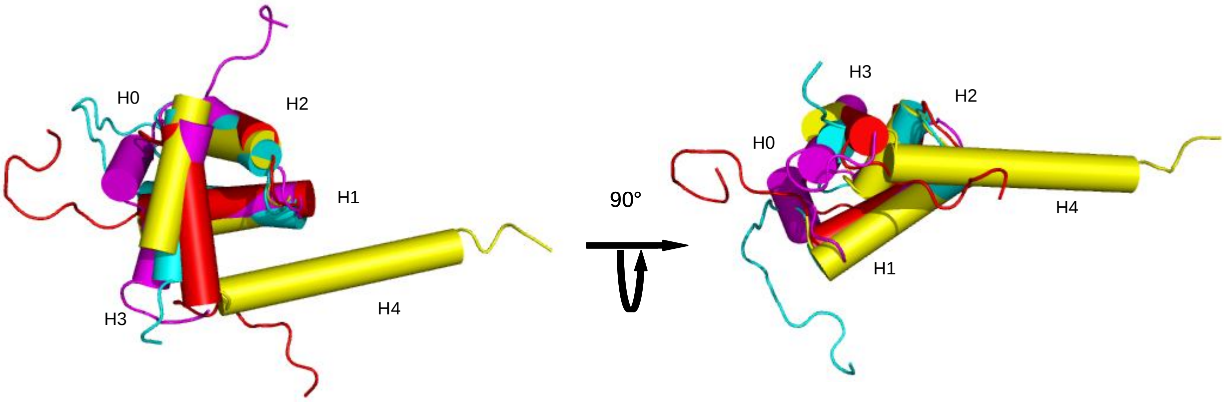
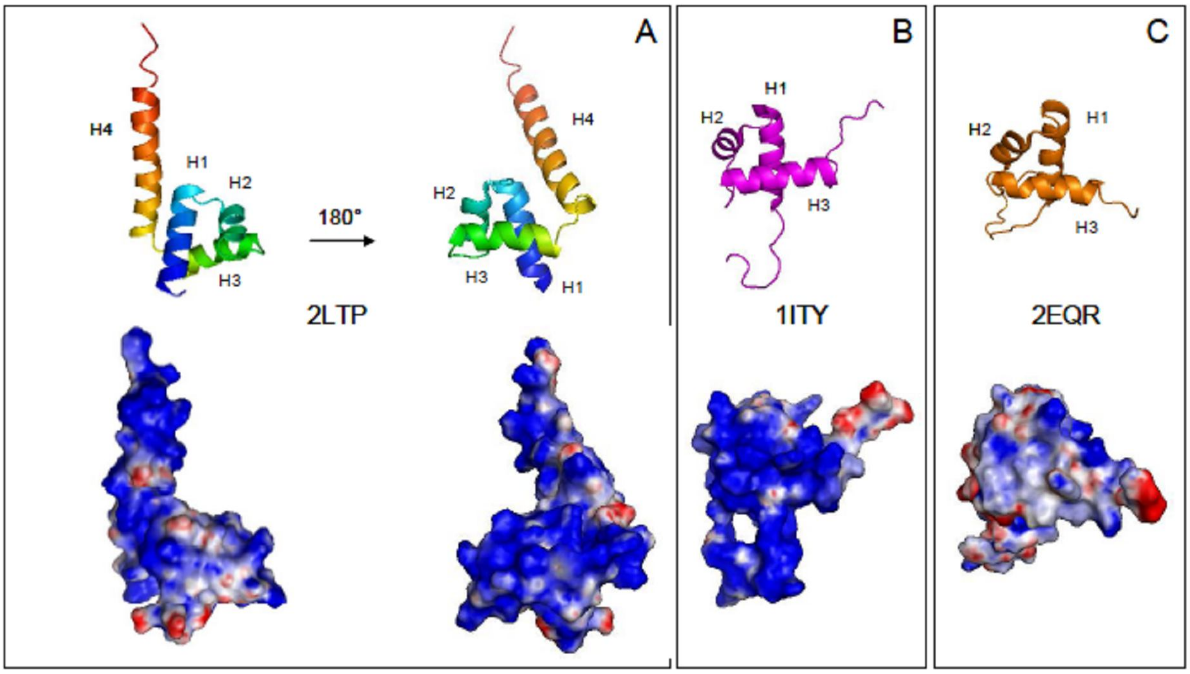


Figure 60(A) SANT2 of NCoR2; **(B)** Myb of hTRF1, DNA-binding; **(C)** SANT of NCoR1; histone tail binding.



9.4 ^{15}N relaxation measurement of the SANT2 domain of NCoR2

The SANT2 of NCoR2 is a 71 amino acid domain with a four helix structure, with the long most C-terminal helix protruding from the central core of the protein. Three residues were excluded from the analysis, because of overlapping/bad signal problems: T616, E617 (both at the N-terminal end of the first helix) and G631. Also in this case, the histag was not removed and so the analysis of the relaxation is just managed at a qualitative level. During the assignment, eight contiguous amino acids were not detected. The analysis of the relaxation data shows high T1/T2 values for the amino acids surrounding the disappeared region (figure 63): while the T1 values are not anomalous for these amino acids (figure 61), the T2 decrease as they come closer to the absent region (figure 62). This is an indication that this segment of the domain undergoes a phenomenon of conformational averaging that leads to an increasing line broadening, up to the complete disappearance of the peaks.

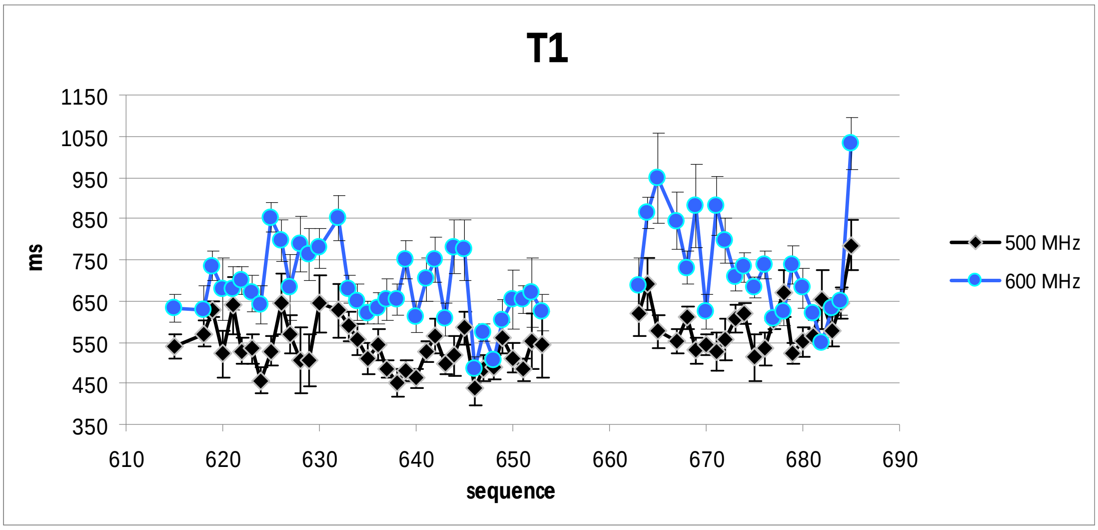


Figure 61: T1 of the SANT2 domain of NCoR2.

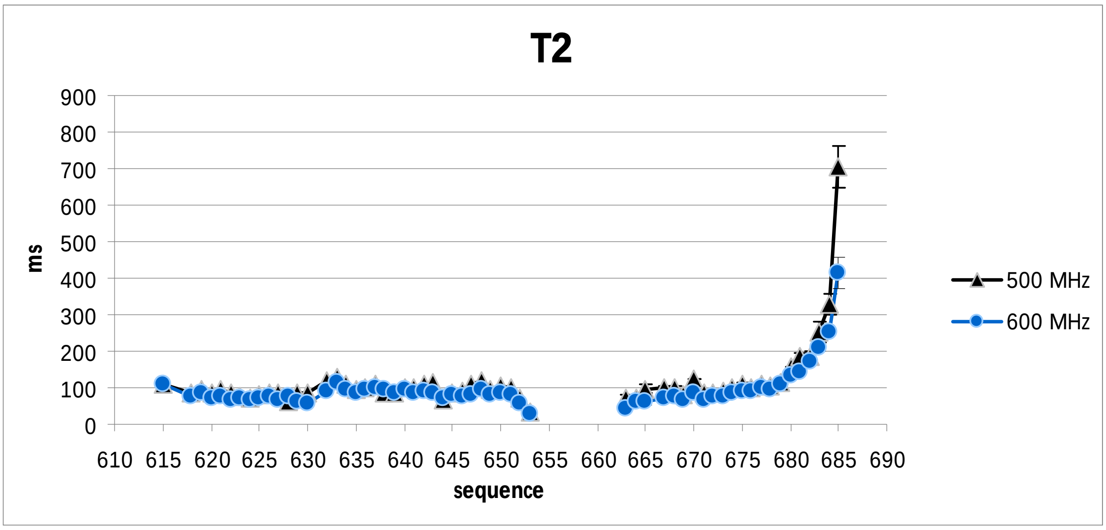


Figure 62: T2 of the SANT2 domain of NCoR2.

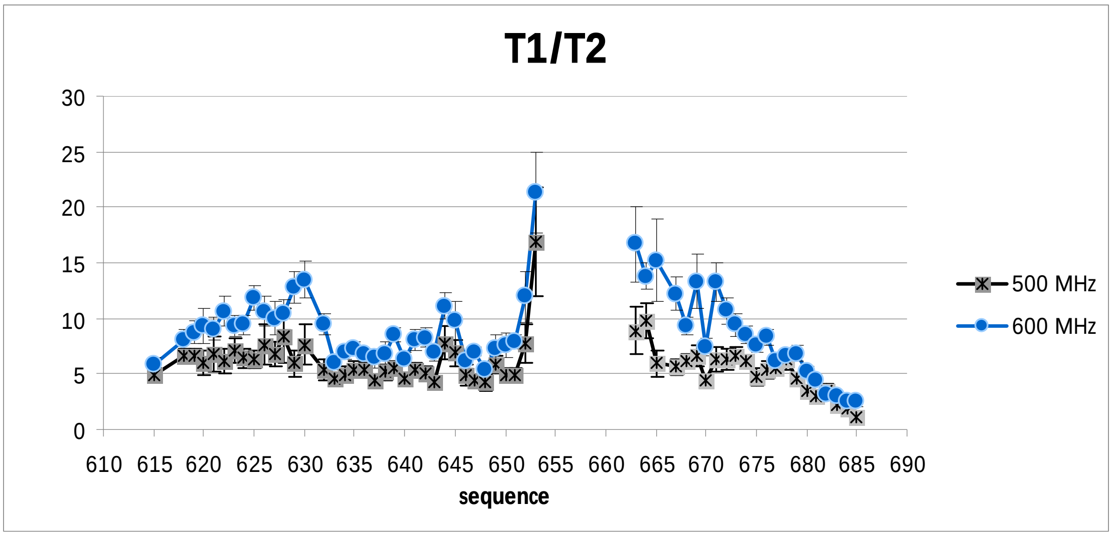


Figure 63: T1/T2 of the SANT2 domain of NCoR2.

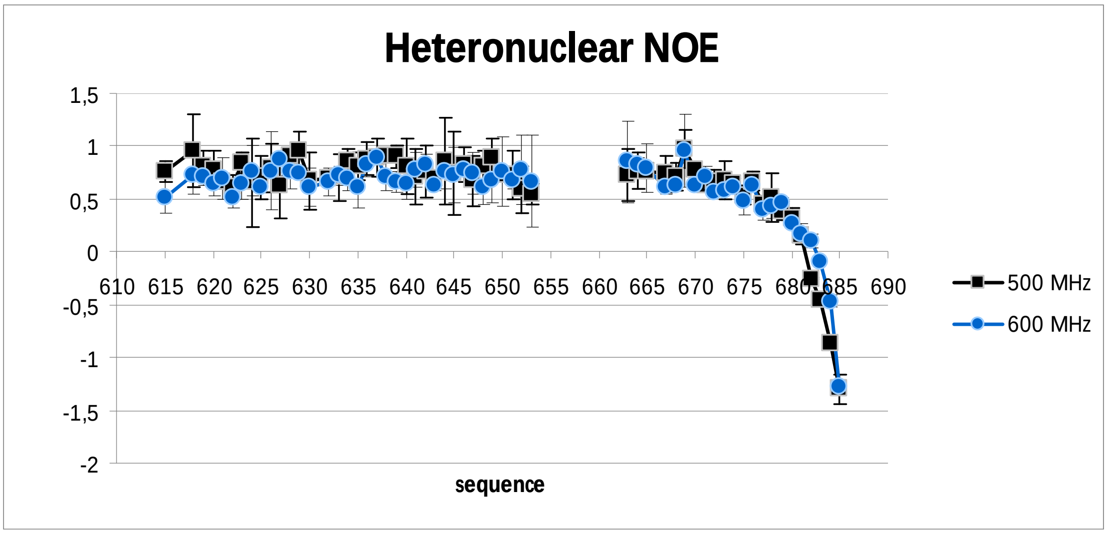


Figure 64: Heteronuclear NOE of the SANT2 domain of NCoR2.

9.5 Preliminary binding experiments monitored by ^{15}N-HSQC of the labelled SANT2 domain with the H4 histone tail peptide

The titrations experiments showed indeed that there is an interaction between SANT2 and the histone tails (figure 65).

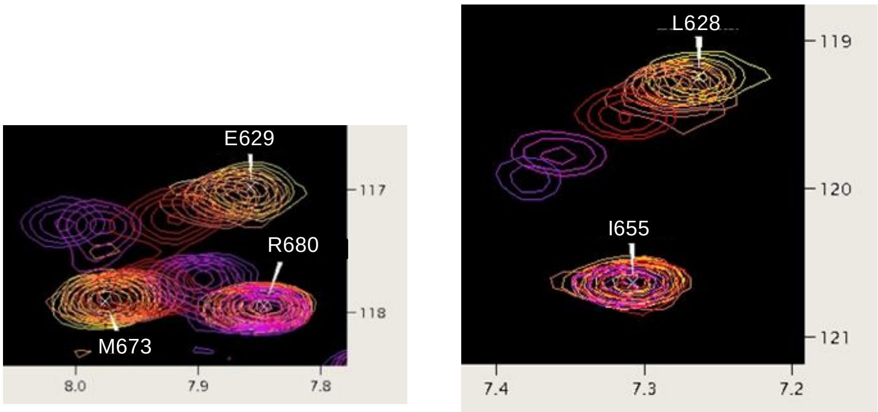


Figure 65: Details of the ^{15}N-HSQC spectra of the titrations, highlighting some of the highest chemical shift perturbations.

The experiment was repeated twice: the first time the molar ratio of H4 histone tails/SANT2 tested were: 0, 2, 5, 10, 20, 30, 40, 50; the second time: 0, 5, 14, 20, 35, with the same protein concentration. The two combined titrations experiments of the ^{15}N-labelled SANT2 with the H4 histone tails show a trend with a double step (figure 66): this is explained by the presence of a higher concentration of TFA (checked with the litmus paper) from the HPLC purification of the peptide from the molar ratio 30 and on. The lower molar ratios were not sensibly affected by the TFA because the concentrated histone tails solution was further diluted to achieve feasible volumes, lowering the concentration of TFA.

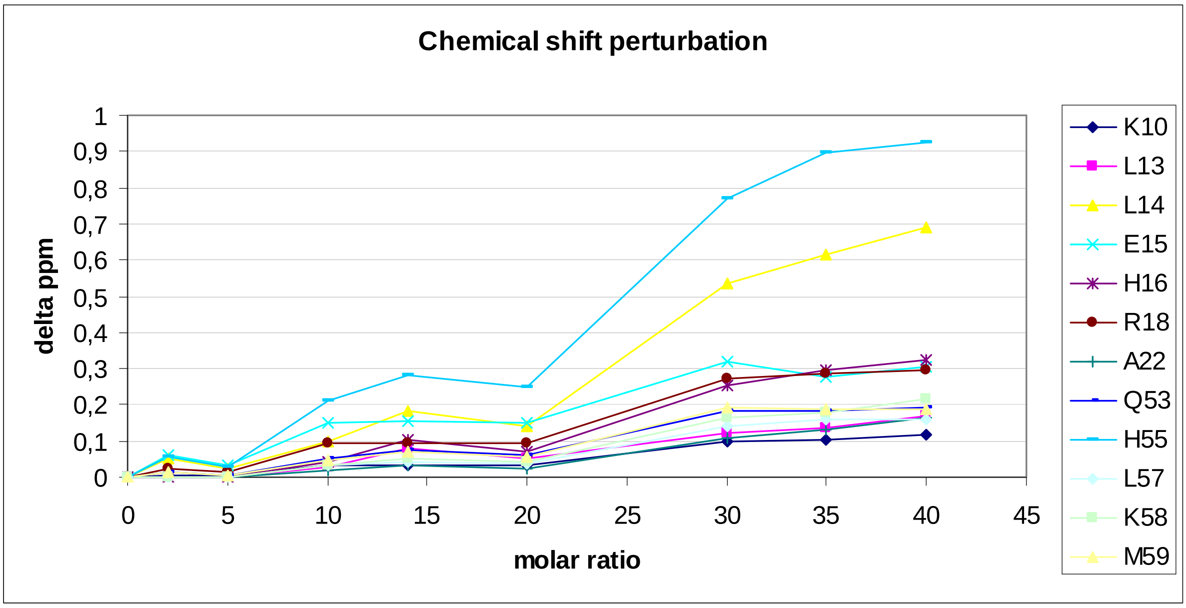


Figure 66: Chart of the two combined titration experiments.

Anyway, the trend seems to reach the saturation before the molar ratio 30, so the experiments were analyzed up to the molar ratio 20. The affected amino acids belong to the helices 1 and 4 and, in particular, five amino acids have a chemical shift perturbation close or higher than 0.1 (figure 67).

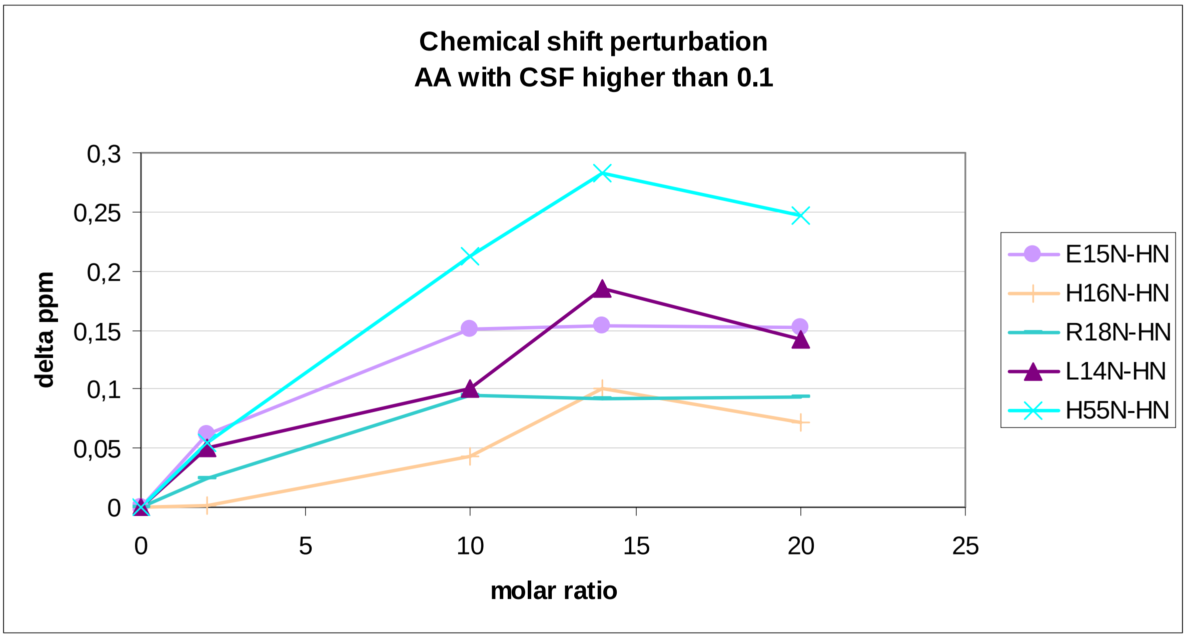


Figure 67: Chart of the five amino acids with the most significant CSP.

The area concerned by the major chemical shift perturbations is in close proximity to the mutation E621G and since the H4 histone tails are basic (pI 9.1), it is not excluded that this missing glutammic acid can be involved in the binding in the wild-type protein (figure 68). Moreover, there is a patch of three other glutammic acids close to the mutation: the missing glutammic acid could be a “bridge” between the area concerned by the titrations and this acid patch on the surface (figure 69).

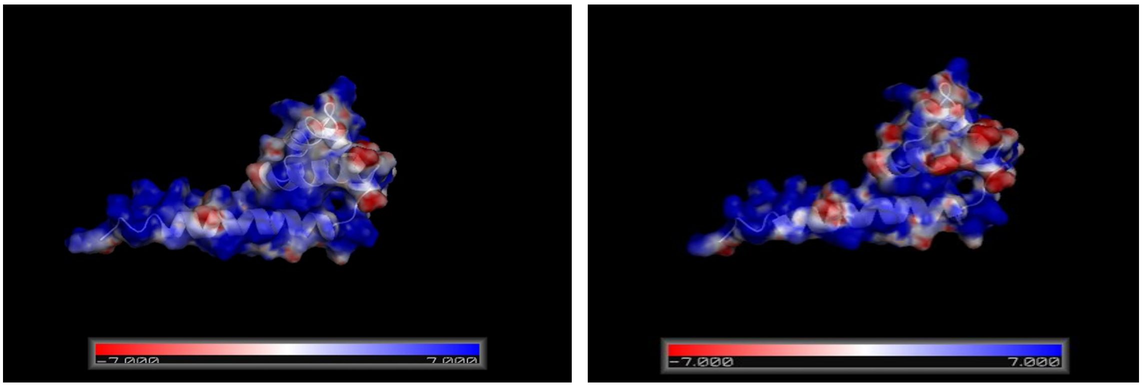


Figure 68: Electrostatic surface of 2LTP (left) and of the wild-type model (right) achieved with HOMER (134).

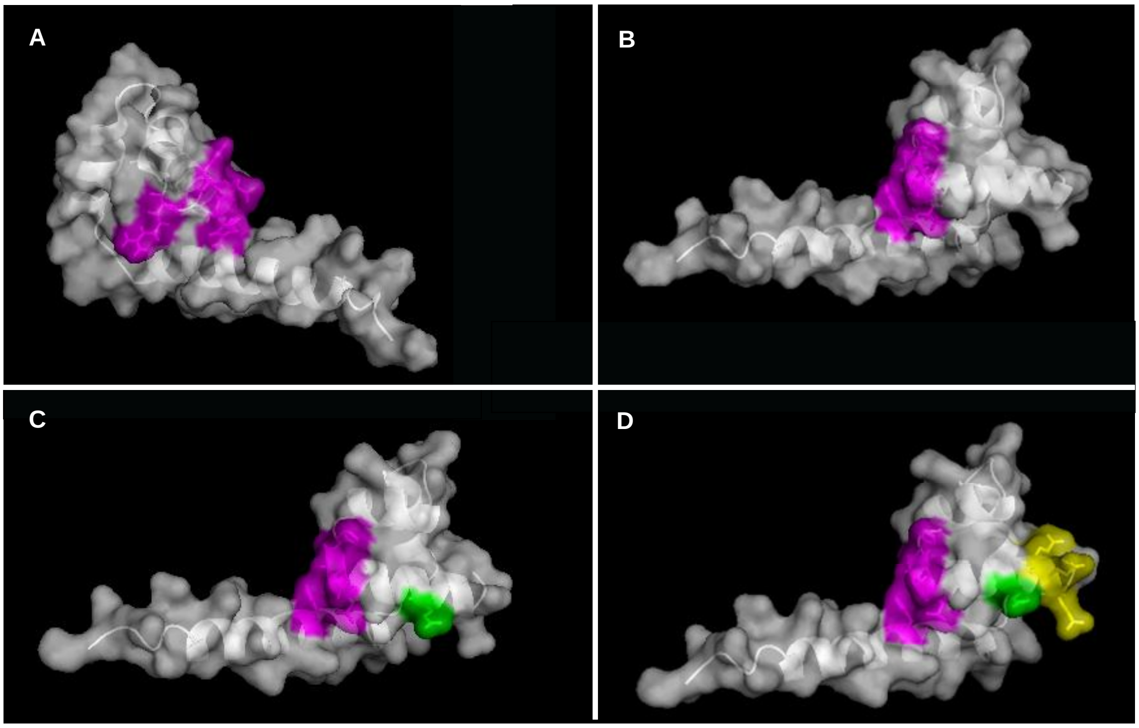


Figure 69: (A) Front and **(B)** back views of the amino acids with a chemical shift perturbation close or higher than 0.1 (magenta) in the SANT2 domain; **(C)** position of the mutation (green) (model achieved with HOMER (134)); **(D)** location of a negative patch of three glutamic acids (yellow).

10. CONCLUSIONS AND FUTURE PERSPECTIVES

The R1 repeat of hDmp1

The first repeat of hDmp1, a 55 amino acid domain, represented a typical Myb domain, with three α-helices where the second and the last one form a Helix-Turn-Helix. No additional elements are present. The structure was solved with 733 NOE based constraints, for an average of 13.3 NOE constraint per residue.

Since it belongs to a three imperfect Myb repeat DNA binding domain, it was tested for its ability to bind known and new sequences and it failed. The explanation about its incapacity of interacting with DNA could be found in the comparison with other well studied DNA-binding Myb domains: its third helix has approximately half of the characteristics necessary to bind the nucleic acid, in terms of charge and conservation of residues interacting with the DNA backbone.

The relaxation analysis did not show any particular feature: the protein did not present any disordered region, except for the highly mobile N- and C-terminal ends.

The SANT2 domain of NCoR2

The SANT2 domain of NCoR2 is a 71 amino acid domain that was solved with 992 NOE based constraints, for an average of 16.6 per residue. Compared with other SANT domains, it present an additional C-terminal long helix, protruding from the central core. This long helix is connect with the previous one with a loop of 8 unassigned amino acids that seem to undergo conformational exchange, as confirmed by the relaxation analysis.

The sample used to solve the structure had a mutation that was detected during the NMR assignment: this had not a repercussion on the backbone structure but it had an effect on the surface charge. In fact, binding experiments with the H4 histone tails were performed and monitored with ^{15}N-HSQC with a^{15}N-labelled SANT2 domain. Five amino acids were found to be affected by the binding and since their region is close to the mutation, we planned to repeat the same experiments with the wild-type protein. The histone tails in fact are basic and the mutated amino acid is an acid one (E621G); in addition, there is a patch constituted by three glutamic acids at the N-terminal end of the helix H1: the mutation is in the middle between this patch and the region affected by the binding. If a binding between the H4 histone tails and the WT SANT2, on its helix H1, will be confirmed, this will corroborate the fact that the SANT2 domain of NCoR2 has to be classified as a Myb domain. Indeed, the charge distribution on the different helices on the SANT2 domain was more similar to the Myb domains, rather than to the SANT domains. No DNA-binding activity is currently known.

REFERENCES

1. Duncan L, Alper S, Losick R. SpoIIAA governs the release of the cell-type specific transcription factor sigma F from its anti-sigma factor SpoIIAB. J. Mol. Biol. 1996 Jul 12;260(2):147–64.

2. Hecker M, Völker U. General stress response of Bacillus subtilis and other bacteria. Adv. Microb. Physiol. 2001;44:35–91.

3. Pané-Farré J, Lewis RJ, Stülke J. The RsbRST stress module in bacteria: a signalling system that may interact with different output modules. J. Mol. Microbiol. Biotechnol. 2005;9(2):65–76.

4. Aravind L, Koonin EV. The STAS domain â€” a link between anion transporters and antisigma-factor antagonists. Current Biology. 2000 Gennaio;10(2):R53–R55.

5. Jurk M, Dorn M, Schmieder P. Blue flickers of hope: secondary structure, dynamics, and putative dimerization interface of the blue-light receptor YtvA from Bacillus subtilis. Biochemistry. 2011 Sep 27;50(38):8163–71.

6. Sharma AK, Rigby AC, Alper SL. STAS domain structure and function. Cell. Physiol. Biochem. 2011;28(3):407–22.

7. Shibagaki N, Grossman AR. Probing the function of STAS domains of the Arabidopsis sulfate transporters. J. Biol. Chem. 2004 Jul 16;279(29):30791–9.

8. Rouached H, Berthomieu P, El Kassis E, Cathala N, Catherinot V, Labesse G, et al. Structural and functional analysis of the C-terminal STAS (sulfate transporter and anti-sigma antagonist) domain of the Arabidopsis thaliana sulfate transporter SULTR1.2. J. Biol. Chem. 2005 Apr 22;280(16):15976–83.

9. Loqué D, Lalonde S, Looger LL, Von Wirén N, Frommer WB. A cytosolic trans-activation domain essential for ammonium uptake. Nature. 2007 Mar 8;446(7132):195–8.

10. Kovacs H, Comfort D, Lord M, Campbell ID, Yudkin MD. Solution structure of SpoIIAA, a phosphorylatable component of the system that regulates transcription factor sigmaF of Bacillus subtilis. Proc. Natl. Acad. Sci. U.S.A. 1998 Apr 28;95(9):5067–71.

11. Sharma AK, Ye L, Baer CE, Shanmugasundaram K, Alber T, Alper SL, et al. Solution structure of the guanine nucleotide-binding STAS domain of SLC26-related SulP protein Rv1739c from Mycobacterium tuberculosis. J. Biol. Chem. 2011 Mar 11;286(10):8534–44.

12. Pasqualetto E, Aiello R, Gesiot L, Bonetto G, Bellanda M, Battistutta R. Structure of the cytosolic portion of the motor protein prestin and functional role of the STAS domain in SLC26/SulP anion transporters. J. Mol. Biol. 2010 Jul 16;400(3):448–62.

13. Dorwart MR, Shcheynikov N, Yang D, Muallem S. The solute carrier 26 family of proteins in epithelial ion transport. Physiology (Bethesda). 2008 Apr;23:104–14.

14. Dawson PA, Markovich D. Pathogenetics of the human SLC26 transporters. Curr. Med. Chem. 2005;12(4):385–96.

15. Detro-Dassen S, Schänzler M, Lauks H, Martin I, Zu Berstenhorst SM, Nothmann D, et al. Conserved dimeric subunit stoichiometry of SLC26 multifunctional anion exchangers. J. Biol. Chem. 2008 Feb 15;283(7):4177–88.

16. Mount DB, Romero MF. The SLC26 gene family of multifunctional anion exchangers. Pflugers Arch. 2004 Feb;447(5):710–21.

17. Gouaux E, Mackinnon R. Principles of selective ion transport in channels and pumps. Science. 2005 Dec 2;310(5753):1461–5.

18. Saier MH Jr, Eng BH, Fard S, Garg J, Haggerty DA, Hutchinson WJ, et al. Phylogenetic characterization of novel transport protein families revealed by genome analyses. Biochim. Biophys. Acta. 1999 Feb 25;1422(1):1–56.

19. Songyang Z, Fanning AS, Fu C, Xu J, Marfatia SM, Chishti AH, et al. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. Science. 1997 Jan 3;275(5296):73–7.

20. Toure A, Morin L, Pineau C, Becq F, Dorseuil O, Gacon G. Tat1, a novel sulfate transporter specifically expressed in human male germ cells and potentially linked to rhogtpase signaling. J. Biol. Chem. 2001 Jun 8;276(23):20309–15.

21. Ko SBH, Zeng W, Dorwart MR, Luo X, Kim KH, Millen L, et al. Gating of CFTR by the STAS domain of SLC26 transporters. Nat. Cell Biol. 2004 Apr;6(4):343–50.

22. Alvarez BV, Vilas GL, Casey JR. Metabolon disruption: a mechanism that regulates bicarbonate transport. EMBO J. 2005 Jul 20;24(14):2499–511.

23. Evans BN, Dallos P. Stereocilia displacement induced somatic motility of cochlear outer hair cells. Proc. Natl. Acad. Sci. U.S.A. 1993 Sep 15;90(18):8347–51.

24. Kachar B, Brownell WE, Altschuler R, Fex J. Electrokinetic shape changes of cochlear outer hair cells. Nature. 1986 Jul 24;322(6077):365–8.

25. Ashmore JF. Forward and reverse transduction in the mammalian cochlea. Neurosci. Res. Suppl. 1990;12:S39–50.

26. Zheng J, Shen W, He DZ, Long KB, Madison LD, Dallos P. Prestin is the motor protein of cochlear outer hair cells. Nature. 2000 May 11;405(6783):149–55.

27. Mistrík P, Daudet N, Morandell K, Ashmore JF. Mammalian prestin is a weak Cl-/HCO3- electrogenic antiporter. J. Physiol. (Lond.). 2012 Nov 15;590(Pt 22):5597–610.

28. Holley MC. Outer Hair Cell Motility. In: Dallos P, Popper AN, Fay RR, editors. The Cochlea [Internet]. Springer New York; 1996 [cited 2013 Jan 14]. page 386–434. Available from: http://link.springer.com/chapter/10.1007/978-1-4612-0757-3_7

29. Schaechinger TJ, Oliver D. Nonmammalian orthologs of prestin (SLC26A5) are electrogenic divalent/chloride anion exchangers. Proc. Natl. Acad. Sci. U.S.A. 2007 May 1;104(18):7693–8.

30. Oliver D, He DZ, Klöcker N, Ludwig J, Schulte U, Waldegger S, et al. Intracellular anions as the voltage sensor of prestin, the outer hair cell motor protein. Science. 2001 Jun 22;292(5525):2340–3.

31. Rybalchenko V, Santos-Sacchi J. Anion control of voltage sensing by the motor protein prestin in outer hair cells. Biophys. J. 2008 Nov 1;95(9):4439–47.

32. Gale JE, Ashmore JF. The outer hair cell motor in membrane patches. Pflugers Arch. 1997 Jul;434(3):267–71.

33. Bai J-P, Surguchev A, Ogando Y, Song L, Bian S, Santos-Sacchi J, et al. Prestin surface expression and activity are augmented by interaction with MAP1S, a microtubule-associated protein. J. Biol. Chem. 2010 Jul 2;285(27):20834–43.

34. Dallos P, Fakler B. Prestin, a new type of motor protein. Nat. Rev. Mol. Cell Biol. 2002 Feb;3(2):104–11.

35. Höfgen R, Kreft O, Willmitzer L, Hesse H. Manipulation of thiol contents in plants. Amino Acids. 2001;20(3):291–9.

36. Takahashi H, Kopriva S, Giordano M, Saito K, Hell R. Sulfur Assimilation in Photosynthetic Organisms: Molecular Functions and Regulations of Transporters and Assimilatory Enzymes. Annual Review of Plant Biology. 2011;62(1):157–84.

37. Hawkesford MJ. Transporter gene families in plants: the sulphate transporter gene family — redundancy or specialization? Physiologia Plantarum. 2003;117(2):155–63.

38. Yoshimoto N, Takahashi H, Smith FW, Yamaya T, Saito K. Two distinct high□affinity sulfate transporters with different inducibilities mediate uptake of sulfate in Arabidopsis roots. The Plant Journal. 2002 Feb 1;29(4):465–73.

39. Maruyama-Nakashita A, Inoue E, Watanabe-Takahashi A, Yamaya T, Takahashi H. Transcriptome profiling of sulfur-responsive genes in Arabidopsis reveals global effects of sulfur nutrition on multiple metabolic pathways. Plant Physiol. 2003 Jun;132(2):597–605.

40. Shibagaki N, Grossman AR. Binding of cysteine synthase to the STAS domain of sulfate transporter and its regulatory consequences. J. Biol. Chem. 2010 Aug 6;285(32):25094–102.

41. SULFUR in South-East Asian Y South Pacific Agriculture, Research for Development Seminar, Ciawi, Indonesia, 1983. Australian Development Assistance Bureau; 19??

42. Deshpande SS. Food legumes in human nutrition: a personal perspective. Crit Rev Food Sci Nutr. 1992;32(4):333–63.

43. Rouached H, Secco D, Arpat AB. Getting the most sulfate from soil: Regulation of sulfate uptake transporters in Arabidopsis. J. Plant Physiol. 2009 Jun 1;166(9):893–902.

44. Xie Q, Welch R, Mercado A, Romero MF, Mount DB. Molecular characterization of the murine Slc26a6 anion exchanger: functional comparison with Slc26a1. Am. J. Physiol. Renal Physiol. 2002 Oct;283(4):F826–838.

45. Shelden MC, Howitt SM, Price GD. Membrane topology of the cyanobacterial bicarbonate transporter, BicA, a member of the SulP (SLC26A) family. Mol. Membr. Biol. 2010 Jan;27(1):12–23.

46. Hautbergue GM, Golovanov AP. Increasing the sensitivity of cryoprobe protein NMR experiments by using the sole low-conductivity arginine glutamate salt. J. Magn. Reson. 2008 Apr;191(2):335–9.

47. Boyer LA, Latek RR, Peterson CL. The SANT domain: a unique histone-tail-binding module? Nat Rev Mol Cell Biol. 2004 Feb;5(2):158–63.

48. Ogata K, Hojo H, Aimoto S, Nakai T, Nakamura H, Sarai A, et al. Solution structure of a DNA-binding unit of Myb: a helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core. Proc. Natl. Acad. Sci. U.S.A. 1992 Jul 15;89(14):6428–32.

49. Saikumar P, Murali R, Reddy EP. Role of tryptophan repeats and flanking amino acids in Myb-DNA interactions. Proc. Natl. Acad. Sci. U.S.A. 1990 Nov;87(21):8452–6.

50. Zargarian L, Le Tilly V, Jamin N, Chaffotte A, Gabrielsen OS, Toma F, et al. Myb−DNA Recognition: Role of Tryptophan Residues and Structural Changes of the Minimal DNA Binding Domain of c-Myb. Biochemistry. 1999;38(6):1921–9.

51. Kanei-Ishii C, Sarai A, Sawazaki T, Nakagoshi H, He DN, Ogata K, et al. The tryptophan cluster: a hypothetical structure of the DNA-binding domain of the myb protooncogene product. J. Biol. Chem. 1990 Nov 15;265(32):19990–5.

52. Oda M, Furukawa K, Ogata K, Sarai A, Ishii S, Nishimura Y, et al. Identification of indispensable residues for specific DNA-binding in the imperfect tandem repeats of c-Myb R2R3. Protein Eng. 1997 Dec;10(12):1407–14.

53. De la Cruz X, Lois S, Sánchez-Molina S, Martínez-Balbás MA. Do protein motifs read the histone code? Bioessays. 2005 Feb;27(2):164–75.

54. Ko ER, Ko D, Chen C, Lipsick JS. A conserved acidic patch in the Myb domain is required for activation of an endogenous target gene and for chromatin binding. Mol. Cancer. 2008;7:77.

55. Wei S-Y, Lou Y-C, Tsai J-Y, Ho M-R, Chou C-C, Rajasekaran M, et al. Structure of the Trichomonas vaginalis Myb3 DNA-binding domain bound to a promoter sequence reveals a unique C-terminal β-hairpin conformation. Nucl. Acids Res. [Internet]. 2011 Sep 8 [cited 2012 Aug 31]; Available from: http://nar.oxfordjournals.org/content/early/2011/09/08/nar.gkr707

56. Sarai A, Uedaira H, Morii H, Yasukawa T, Ogata K, Nishimura Y, et al. Thermal stability of the DNA-binding domain of the Myb oncoprotein. Biochemistry. 1993 Aug 1;32(30):7759–64.

57. Ogata K, Morikawa S, Nakamura H, Hojo H, Yoshimura S, Zhang R, et al. Comparison of the free and DNA-complexed forms of the DMA-binding domain from c-Myb. Nature Structural & Molecular Biology. 1995 Apr 1;2(4):309–20.

58. Sasaki M, Ogata K, Hatanaka H, Nishimura Y. Backbone dynamics of the c-Myb DNA-binding domain complexed with a specific DNA. J. Biochem. 2000 Jun;127(6):945–53.

59. Ogata K, Kanei-Ishii C, Sasaki M, Hatanaka H, Nagadoi A, Enari M, et al. The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation. Nat Struct Mol Biol. 1996 Feb;3(2):178–87.

60. Hanaoka S, Nagadoi A, Yoshimura S, Aimoto S, Li B, De Lange T, et al. NMR structure of the hRap1 Myb motif reveals a canonical three-helix bundle lacking the positive surface charge typical of Myb DNA-binding domains. J. Mol. Biol. 2001 Sep 7;312(1):167–75.

61. Bergholtz S, Andersen TO, Andersson KB, Borrebaek J, Lüscher B, Gabrielsen OS. The highly conserved DNA-binding domains of A-, B- and c-Myb differ with respect to DNA-binding, phosphorylation and redox properties. Nucleic Acids Res. 2001 Sep 1;29(17):3546–56.

62. Tanikawa J, Yasukawa T, Enari M, Ogata K, Nishimura Y, Ishii S, et al. Recognition of specific DNA sequences by the c-myb protooncogene product: role of three repeat units in the DNA-binding domain. Proc Natl Acad Sci U S A. 1993 Oct 15;90(20):9320–4.

63. Codina A, Love JD, Li Y, Lazar MA, Neuhaus D, Schwabe JWR. Structural insights into the interaction and activation of histone deacetylase 3 by nuclear receptor corepressors. Proc. Natl. Acad. Sci. U.S.A. 2005 Apr 26;102(17):6009–14.

64. Hirai H, Sherr CJ. Interaction of D-type cyclins with a novel myb-like transcription factor, DMP1. Mol. Cell. Biol. 1996 Nov;16(11):6457–67.

65. Bodner SM, Naeve CW, Rakestraw KM, Jones BG, Valentine VA, Valentine MB, et al. Cloning and chromosomal localization of the gene encoding human cyclin D-binding Myb-like protein (hDMP1). Gene. 1999 Mar 18;229(1-2):223–8.

66. Tschan MP, Fischer KM, Fung VS, Pirnia F, Borner MM, Fey MF, et al. Alternative splicing of the human cyclin D-binding Myb-like protein (hDMP1) yields a truncated protein isoform that alters macrophage differentiation patterns. J. Biol. Chem. 2003 Oct 31;278(44):42750–60.

67. Sugiyama T, Frazier DP, Taneja P, Kendig RD, Morgan RL, Matise LA, et al. Signal transduction involving the dmp1 transcription factor and its alteration in human cancer. Clin Med Oncol. 2008;2:209–19.

68. Inoue K, Roussel MF, Sherr CJ. Induction of ARF tumor suppressor gene expression and cell cycle arrest by transcription factor DMP1. Proc. Natl. Acad. Sci. U.S.A. 1999 Mar 30;96(7):3993–8.

69. Inoue K, Sherr CJ, Shapiro LH. Regulation of the CD13/aminopeptidase N gene by DMP1, a transcription factor antagonized by D-type cyclins. J. Biol. Chem. 1998 Oct 30;273(44):29188–94.

70. Inoue K, Sherr CJ. Gene Expression and Cell Cycle Arrest Mediated by Transcription Factor DMP1 Is Antagonized by D-Type Cyclins through a Cyclin-Dependent-Kinase-Independent Mechanism. Mol. Cell. Biol. 1998 Mar 1;18(3):1590–600.

71. Mallakin A, Taneja P, Matise L, Willingham M, Inoue K. Expression of Dmp1 in specific differentiated, nonproliferating cells and its regulation by E2Fs. Oncogene. 2006 Dec 14;25(59):7703–13.

72. Taneja P, Mallakin A, Matise LA, Frazier DP, Choudhary M, Inoue K. Repression of Dmp1 and Arf transcription by anthracyclins: critical roles of the NF-kappaB subunit p65. Oncogene. 2007 Nov 22;26(53):7457–66.

73. Sreeramaneni R, Chaudhry A, McMahon M, Sherr CJ, Inoue K. Ras-Raf-Arf Signaling Critically Depends on the Dmp1 Transcription Factor. Mol Cell Biol. 2005 Jan;25(1):220–32.

74. Taneja P, Maglic D, Kai F, Sugiyama T, Kendig RD, Frazier DP, et al. Critical roles of DMP1 in human epidermal growth factor receptor 2/neu-Arf-p53 signaling and breast cancer development. Cancer Res. 2010 Nov 15;70(22):9084–94.

75. Trovato M, Ulivieri A, Dominici R, Ruggeri RM, Vitarelli E, Benvenga S, et al. Clinico-pathological significance of cell-type-specific loss of heterozygosity on chromosome 7q21: analysis of 318 microdissected thyroid lesions. Endocr. Relat. Cancer. 2004 Jun;11(2):365–76.

76. Mallakin A, Sugiyama T, Taneja P, Matise LA, Frazier DP, Choudhary M, et al. Mutually exclusive inactivation of DMP1 and ARF/p53 in lung cancer. Cancer Cell. 2007 Oct;12(4):381–94.

77. Inoue K, Sugiyama T, Taneja P, Morgan RL, Frazier DP. Emerging roles of DMP1 in lung cancer. Cancer Res. 2008 Jun 15;68(12):4487–90.

78. Inoue K, Wen R, Rehg JE, Adachi M, Cleveland JL, Roussel MF, et al. Disruption of the ARF transcriptional activator DMP1 facilitates cell

immortalization, Ras transformation, and tumorigenesis. Genes Dev. 2000 Jul 15;14(14):1797–809.

79. Inoue K, Zindy F, Randle DH, Rehg JE, Sherr CJ. Dmp1 is haplo-insufficient for tumor suppression and modifies the frequencies of Arf and p53 mutations in Myc-induced lymphomas. Genes Dev. 2001 Nov 15;15(22):2934–9.

80. Frazier DP, Kendig RD, Kai F, Maglic D, Sugiyama T, Morgan RL, et al. Dmp1 physically interacts with p53 and positively regulates p53's stability, nuclear localization, and function. Cancer Res. 2012 Apr 1;72(7):1740–50.

81. Mallakin A, Sugiyama T, Kai F, Taneja P, Kendig RD, Frazier DP, et al. The Arf-inducing transcription factor Dmp1 encodes a transcriptional activator of amphiregulin, thrombospondin-1, JunB and Egr1. Int. J. Cancer. 2010 Mar 15;126(6):1403–16.

82. Inoue K, Mallakin A, Frazier DP. Dmp1 and tumor suppression. Oncogene. 2007 Jun 28;26(30):4329–35.

83. Xu L, Glass CK, Rosenfeld MG. Coactivator and corepressor complexes in nuclear receptor function. Current Opinion in Genetics & Development. 1999 Apr;9(2):140–7.

84. Privalsky ML. The Role of Corepressors in Transcriptional Regulation by Nuclear Hormone Receptors. Annual Review of Physiology. 2004;66(1):315–60.

85. Hörlein AJ, Näär AM, Heinzel T, Torchia J, Gloss B, Kurokawa R, et al. Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. Nature. 1995 Oct 5;377(6548):397–404.

86. Chen JD, Evans RM. A transcriptional co-repressor that interacts with nuclear hormone receptors. Nature. 1995 Oct 5;377(6548):454–7.

87. Dowell P, Ishmael JE, Avram D, Peterson VJ, Nevrivy DJ, Leid M. Identification of nuclear receptor corepressor as a peroxisome proliferator-activated receptor alpha interacting protein. J. Biol. Chem. 1999 May 28;274(22):15901–7.

88. Dwivedi PP, Muscat GE, Bailey PJ, Omdahl JL, May BK. Repression of basal transcription by vitamin D receptor: evidence for interaction of unliganded vitamin D receptor with two receptor interaction domains in RIP13delta1. J. Mol. Endocrinol. 1998 Jun;20(3):327–35.

89. Jackson TA, Richer JK, Bain DL, Takimoto GS, Tung L, Horwitz KB. The partial agonist activity of antagonist-occupied steroid receptors is controlled by a novel hinge domain-binding coactivator L7/SPA and the corepressors N-CoR or SMRT. Mol. Endocrinol. 1997 Jun;11(6):693–705.

90. Hsia EY, Goodson ML, Zou JX, Privalsky ML, Chen H-W. Nuclear receptor coregulators as a new paradigm for therapeutic targeting. Adv. Drug Deliv. Rev. 2010 Oct 30;62(13):1227–37.

91. Tsai S, Bartelmez S, Heyman R, Damm K, Evans R, Collins SJ. A mutated retinoic acid receptor-alpha exhibiting dominant-negative activity alters the lineage development of a multipotent hematopoietic cell line. Genes Dev. 1992 Dec;6(12A):2258–69.

92. Breems-de Ridder MC, Löwenberg B, Jansen JH. Retinoic acid receptor fusion proteins: friend or foe. Molecular and Cellular Endocrinology. 2000 Jul 25;165(1–2):1–6.

93. Perissi V, Jepsen K, Glass CK, Rosenfeld MG. Deconstructing repression: evolving models of co-repressor action. Nat. Rev. Genet. 2010 Feb;11(2):109–23.

94. Zhang J, Kalkum M, Chait BT, Roeder RG. The N-CoR-HDAC3 nuclear receptor corepressor complex inhibits the JNK pathway through the integral subunit GPS2. Mol. Cell. 2002 Mar;9(3):611–23.

95. Guenther MG, Lane WS, Fischle W, Verdin E, Lazar MA, Shiekhattar R. A core SMRT corepressor complex containing HDAC3 and TBL1, a WD40-repeat protein linked to deafness. Genes Dev. 2000 May 1;14(9):1048–57.

96. Yoon H-G, Chan DW, Huang Z-Q, Li J, Fondell JD, Qin J, et al. Purification and functional characterization of the human N-CoR complex: the roles of HDAC3, TBL1 and TBLR1. EMBO J. 2003 Mar 17;22(6):1336–46.

97. Heinzel T, Lavinsky RM, Mullen TM, Söderstrom M, Laherty CD, Torchia J, et al. A complex containing N-CoR, mSin3 and histone deacetylase mediates transcriptional repression. Nature. 1997 May 1;387(6628):43–8.

98. Oberoi J, Fairall L, Watson PJ, Yang J-C, Czimmerer Z, Kampmann T, et al. Structural basis for the assembly of the SMRT/NCoR core transcriptional repression machinery. Nat. Struct. Mol. Biol. 2011 Feb;18(2):177–84.

99. Watson PJ, Fairall L, Schwabe JWR. Nuclear hormone receptor co-repressors: structure and function. Mol. Cell. Endocrinol. 2012 Jan 30;348(2):440–9.

100. Guenther MG, Barak O, Lazar MA. The SMRT and N-CoR corepressors are activating cofactors for histone deacetylase 3. Mol. Cell. Biol. 2001 Sep;21(18):6091–101.

101. Yu J, Li Y, Ishizuka T, Guenther MG, Lazar MA. A SANT motif in the SMRT corepressor interprets the histone code and promotes histone deacetylation. EMBO J. 2003 Jul 1;22(13):3403–10.

102. Hartman HB, Yu J, Alenghat T, Ishizuka T, Lazar MA. The histone-binding code of nuclear receptor co-repressors matches the substrate specificity of histone deacetylase 3. EMBO Rep. 2005 May;6(5):445–51.

103. 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. J. Biomol. NMR. 1991 May;1(1):99–104.

104. Grzesiek S, Döbeli H, Gentz R, Garotta G, Labhardt AM, Bax A. 1H, 13C, and 15N NMR backbone assignments and secondary structure of human interferon-gamma. Biochemistry. 1992 Sep 8;31(35):8180–90.

105. Ikura M, Kay LE, Bax A. A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. Biochemistry. 1990 May 15;29(19):4659–67.

106. Kay LE, Ikura M, Tschudin R, Bax A. Three-dimensional triple-resonance NMR Spectroscopy of isotopically enriched proteins. 1990. J. Magn. Reson. 2011 Dec;213(2):423–41.

107. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR. 1995 Nov;6(3):277–93.

108. Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A. Use of dipolar 1H-15N and 1H-13C couplings in the structure determination of magnetically oriented macromolecules in solution. Nat. Struct. Biol. 1997 Sep;4(9):732–8.

109. Kay LE, Clore GM, Bax A, Gronenborn AM. Four-dimensional heteronuclear triple-resonance NMR spectroscopy of interleukin-1 beta in solution. Science. 1990 Jul 27;249(4967):411–4.

110. Grzesiek S, Bax A. Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. J. Biomol. NMR. 1993 Mar;3(2):185–204.

111. Goddard TD, Kneller DG. SPARKY 3. San Francisco, CA, USA: University of California;

112. Güntert P. Automated NMR structure calculation with CYANA. Methods Mol. Biol. 2004;278:353–78.

113. Lemak A, Gutmanas A, Chitayat S, Karra M, Farès C, Sunnerhagen M, et al. A novel strategy for NMR resonance assignment and protein structure determination. J. Biomol. NMR. 2011 Jan;49(1):27–38.

114. Orekhov VY, Ibraghimov I, Billeter M. Optimizing resolution in multidimensional NMR by three-way decomposition. Journal of Biomolecular NMR. 2003;27(2):165–73.

115. Grishaev A, Steren CA, Wu B, Pineda-Lucena A, Arrowsmith C, Llinás M. ABACUS, a direct method for protein NMR structure computation via assembly of fragments. Proteins. 2005 Oct 1;61(1):36–43.

116. Bodenhausen G, Ruben DJ. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. Chemical Physics Letters. 1980 Jan 1;69(1):185–9.

117. Grishaev A, Llinás M. BACUS: A Bayesian protocol for the identification of protein NOESY spectra via unassigned spin systems. J. Biomol. NMR. 2004 Jan;28(1):1–10.

118. Lemak A, Steren CA, Arrowsmith CH, Llinás M. Sequence specific resonance assignment via Multicanonical Monte Carlo search using an ABACUS approach. J. Biomol. NMR. 2008 May;41(1):29–41.

119. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. 1998 Sep 1;54(Pt 5):905–21.

120. Stern AS, Li K-B, Hoch JC. Modern spectrum analysis in multidimensional NMR spectroscopy: comparison of linear-prediction extrapolation and maximum-entropy reconstruction. J. Am. Chem. Soc. 2002 Mar 6;124(9):1982–93.

121. Barna JC., Laue E., Mayger M., Skilling J, Worrall SJ. Exponential sampling, an alternative method for sampling in two-dimensional NMR experiments. Journal of Magnetic Resonance (1969). 1987 Jun 1;73(1):69–77.

122. Gutmanas A, Arrowsmith CH. MDDGUI. University of Toronto;

123. Koradi R, Billeter M, Wüthrich K. MOLMOL: A program for display and analysis of macromolecular structures. Journal of Molecular Graphics. 1996 Feb;14(1):51–5.

124. DeLano WL. The PyMOL Molecular Graphics System [Internet]. San Carlos, CA, USA: DeLano Scientific; 2002. Available from: http://www.pymol.org

125. Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. Proteins. 2007 Mar 1;66(4):778–95.

126. Kay LE, Torchia DA, Bax A. Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry. 1989 Nov 14;28(23):8972–9.

127. Studier FW. Protein production by auto-induction in high density shaking cultures. Protein Expr. Purif. 2005 May;41(1):207–34.

128. Hanaoka S, Nagadoi A, Nishimura Y. Comparison between TRF2 and TRF1 of their telomeric DNA-bound structures and DNA-binding activities. Protein Sci. 2005 Jan;14(1):119–30.

129. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography. 1993 Apr 1;26(2):283–91.

130. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography.

Acta Crystallographica Section D Biological Crystallography. 2009 Dec 21;66(1):12–21.

131. Cole C, Barber JD, Barton GJ. The Jpred 3 secondary structure prediction server. Nucleic Acids Research. 2008 May 19;36(Web Server):W197–W201.

132. Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. Nucleic Acids Research. 2010 May 10;38(Web Server):W545–W549.

133. Eisenberg D, Schwarz E, Komaromy M, Wall R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. Journal of Molecular Biology. 1984 Oct 15;179(1):125–42.

134. Tosatto. BioComputing UP: Victor [Internet]. [cited 2013 Jan 17]. Available from: http://protein.bio.unipd.it/victor.html

ACKNOWLEDGMENTS

I would like to thank...

My supervisor Dr. Massimo Bellanda, since this work would not have been possible without his support;

Prof. Stefano Mammi, for welcoming me in his lab;

My lab colleagues: Dr. Lorenzo Gesiot, Dr. Mattia Sturlese, Dr. Carlo Pavan;

Prof. Roberto Battistutta for the collaboration in the STAS project and Dr. Elisa Pasqualetto for helping me with the cloning.

The year I spent at the SGC Toronto, Canada, has had an inestimable value for my work and I would like to thank...

Prof. Cheryl Arrowsmith,

Dr. Alexander Lemak,

Dr. Scott Houliston,

Dr. Adelinda Yee