



Université de Strasbourg

Ecole Doctorale des Sciences de la Vie et de la Santé

Thèse présentée par

**Ying-Hui WANG**

Pour obtenir le grade de

Docteur de l'Université de Strasbourg

Sciences du Vivant

Aspects Moléculaires et Cellulaires de la Biologie  
UNIVERSITÉ DE STRASBOURG

**Etude par RMN des domaines DBD du récepteur nucléaire des  
hormones androgènes et des domaines SCA7 de l'Ataxine 7**

Soutenue publiquement le 5 mars 2010

Membres du jury

Directeur de thèse : Pr. Bruno KIEFFER

Rapporteur Interne : Pr. Patrick SCHULTZ

Rapporteur Externe : Pr. Rolf BOELENS

Rapporteur Externe : Pr. Christian ROUMESTAND

Examineur : Dr. Jocelyn CERALINE

Examineur : Dr. Andrew ATKINSON

The logo of the University of Strasbourg, featuring two blue curved segments that form a stylized 'S' shape, positioned behind the text.

UNIVERSITÉ DE STRASBOURG

## Table of contents

<b>Acknowledgements .....</b>	<b>1</b>
<b>Résumé .....</b>	<b>3</b>
<b>Chapter 1: The zinc finger proteins: A big family .....</b>	<b>7</b>
1. The discovery of zinc fingers .....	7
1.1 A brief history of zinc finger discovery .....	7
1.2 Zinc fingers today: Survey of the PDB .....	9
2. Current view of zinc finger evolution .....	14
2.1 Gene duplication of zinc fingers is a hint for evolution .....	14
2.2 Zinc finger associated domains .....	14
3. Functions of zinc finger proteins .....	17
3.1 Zinc finger involved in nucleic acid binding .....	17
3.2 Zinc fingers involved in protein-protein interaction .....	25
3.3 Zinc finger proteins involved in membrane trafficking: FYVE domain .....	38
4. Conclusion .....	39
References.....	40
<b>Chapter 2: Production of zinc finger proteins .....</b>	<b>49</b>
1. Strategy for the production of zinc finger proteins .....	49
1.1 Expression of zinc finger proteins.....	50
2. Materials and methods .....	55
2.1. Culture media.....	55
2.2 Plasmid transformation .....	56
2.3 Culture growth and protein expression .....	56
2.4 Protein purification .....	56
2.5 Expression and purification of recombinant TEV protease.....	58
2.6 Methods for the removal of a tag by TEV protease .....	59
3. Results and discussion .....	60
3.1 Expression and purification of AR-DBD wild type and T575A mutant.....	60
3.2 Comparison of sample quality between ATXN7 fused with His-tag and GST tag.....	61
3.3 Removal of affinity tags from recombinant AR-DBD proteins by TEV protease.....	64
4. Conclusions.....	67
References.....	67
<b>Chapter 3: Metal exchange experiments of zinc finger proteins.....</b>	<b>73</b>
<b>1. Introduction.....</b>	<b>73</b>
1.1 Investigation of zinc binding sites by metal exchange experiments .....	73
1.2 Kinetics of metal exchange .....	74
2. Materials and methods .....	74

2.1 Replacement of Zn <sup>2+</sup> by Cd <sup>2+</sup> .....	74
2.2 Measurements of kinetics.....	75
2.3 <sup>15</sup> N relaxation rate measurement for AR-DBD WT and T575A mutant .....	75
<b>3. Results and discussion .....</b>	<b>76</b>
3.1 Metal exchange experiments for AR-DBD wild type and T575A mutant .....	76
3.2 Metal exchange experiments on ATXN7 and ATXN7 L3 .....	84
References.....	90
<b>Chapter 4: Dynamic studies of AR-DBD and SCA7 domains of ATXN7 and ATXN7 L3.....</b>	<b>91</b>
1. Introduction.....	91
1.1 Protein dynamics time-scales.....	91
1.2 General principles of protein motion studies using nuclear spin relaxation.....	94
1.3 Interpretation of relaxation parameters .....	94
1.4 Alternative methods to probe protein dynamics .....	98
2. Materials and methods .....	100
2.1 Relaxation measurements and <sup>1</sup> H- <sup>15</sup> N NOE experiments .....	100
2.2 HET-SOFAST experiments.....	101
2.3 Residue dipolar couplings (RDC) measurements .....	102
3. Results.....	103
3.1 Dynamics of ARDBD wild type and T575A mutant.....	103
3.2 Dynamics of ATXN7 zinc finger domains .....	113
3.3 A view on dynamics from the HET-SOFAST experiments .....	127
3.4 Between order and disorder: combining RDC with dynamics data .....	128
3.5 Conclusions.....	130
References.....	131
<b>Chapter 5: Altered specificity of a mutated androgen receptor in prostate cancer is associated with a change in the surface potential of the DNA-binding domain .....</b>	<b>133</b>
<b>Chapter 6: Histone deubiquitination by SAGA is modulated by an atypical zinc finger domain of Ataxin-7.....</b>	<b>147</b>
<b>Conclusion and Perspectives.....</b>	<b>166</b>
<b>Appendix I. NMR spectroscopy and data analysis.....</b>	<b>170</b>
1. Spectral acquisition for backbone assignments .....	170
2. Structural Calculation .....	170
2.1 Initial calculation using ATNOS-CANDID.....	170
2.2 Structure refinement using X-PLOR.....	173
2.3 Water refinement .....	174
<b>Appendix II: DNA interaction of AR-DBD wild type and T575A</b>	



<b>mutant .....</b>	<b>175</b>
1. Introduction.....	175
2. Methods.....	177
3. Results.....	177
4. Conclusions and perspectives .....	178
<b>Appendix III: Investigation of the dynamic behavior of the C-terminal extension region (CTE) of AR-DBD.....</b>	<b>181</b>
1. Introduction.....	181
2. Results and discussion .....	182
2.1 Comparison of $^1\text{H}$ - $^{15}\text{N}$ HSQC spectra of AR-DBD and AR-DBD-CTE.....	182
2.2 Investigation of the dynamic behavior in AR-DBD-CTE.....	183
2.3 Conclusions and perspectives .....	184
<b>Appendix IV: Structural role of ATXN7 in histone 2A 2B heterodimer binding .....</b>	<b>185</b>
<b>Appendix V: ATXN7 lacks DNA binding activity .....</b>	<b>189</b>



The logo of the University of Strasbourg, featuring two blue curved segments that form a stylized 'S' shape, positioned behind the text.

UNIVERSITÉ DE STRASBOURG

## **Acknowledgements**

*Since 2005 I arrived to Strasbourg for my PhD study, a big adventure started. I admit that this would be a very special experience in my life.*

*I would like to thank Comité du Bas-Rhin de LA LIGUE CONTRE LE CANCER provide me three-year PhD fellowship.*

*I am deeply indebted to my thesis advisor, Professor Bruno Kieffer. His wide knowledge of NMR and logical way for thinking inspire me. His understanding, encouraging and personal guidance have provided fundamentals of this thesis. I appreciate his patience with discussion of scientific questions and constructive suggestions for my study.*

*I would like to express my warm and sincere to thank to Dr. Andrew Atkinson for English corrections of this thesis. Andrew provides me valuable instruction for structure calculation, in particular, helps me set up and complete the structure determination of AR-DBD and SCA7 domains in ATXN7 and ATXN7 L3. Moreover, I thank him sincerely to help me settle down when I arrived in Strasbourg.*

UNIVERSITÉ DE STRASBOURG

*For my first topic in the thesis, structure studies of androgen receptor DNA binding domains, I would like to thank our collaborator, Dr. Jocelyn Céraline's group in Equipe Signalisation et Cancer de la prostate, Université de Strasbourg, Faculté de Médecine. Dr. Audrey Monge studied the relations of mutating effect on androgen receptor and the prostate cancer. She provided me the constructs of androgen receptor DNA binding domain and gave me valuable information of biological studies. I also thank to people in the Plate-form of Structural biology and genomic technology of IGBMC, Dr. Didier Busso and Pierre Poussin, provide me technical instruction for protein production.*

*For my second topic in the thesis, structural studies of SCA7 domains in ATXN7 and ATXN7 L3, I would like to thank our collaborator, Dr. Laszlo Tora's group in IGBMC. Dr. Didier Devys directs this topic and provides valuable information for our structural studies. Jacques Bonnet devotes his effort in studies of the interaction of ATXN7 with histone and investigates other possible functions of SCA7 domains. He also provided me technical suggestions for production of SCA7 domains in ATXN7 and ATXN7 L3. I also thank to Dr. Christophe Romier for providing the constructs of SCA7 domains of ATXN7 and ATXN7 L3.*

*In addition, I also would like to thank to Dr. Dominique Desplancq for instruction of purification of TEV protease and thus I could work on the protein purification independently.*

*During my PhD study, I would like deeply thank to the members in biomolecular NMR group: Claude gives me the technical support for NMR spectrometer. He also helps me deal with many problems which the foreigners could meet and may be difficult to overcome in France. Isabelle helps me set up the DNA titration of AR-DBD. Gilles, Juan, Katia, Katja, Marc-André, Marc, Matthieu and Sebastian.*

*I would like to extend my thanks to the library in IGBMC that provides a quiet environment for thesis writing.*

*I would like to express gratitude to all those who help adapt the environment since I arrived to Strasbourg.*

*Finally, I would like to thank deeply to my family for the strongest support in my life of study abroad.*

The logo of the University of Strasbourg, featuring a stylized 'S' composed of two blue curved segments. The text 'UNIVERSITÉ DE STRASBOURG' is centered horizontally across the middle of the 'S' in a light blue, sans-serif font.

UNIVERSITÉ DE STRASBOURG

## Résumé

### I. Introduction

Les domaines à doigts de zinc sont extrêmement abondants chez les eucaryotes supérieurs. Découverts dans un premier temps dans un facteur de transcription (TFIIIA) du xénope, les domaines à doigts de zinc ont d'abord été associés aux fonctions de reconnaissance des acides nucléiques. Depuis quelques années, le répertoire des fonctions associées à ces domaines s'est considérablement élargi, en particulier avec la découverte que de nombreuses interactions protéine protéine impliquaient également des domaines à doigts de zinc. La diversité des fonctions assurées par ce type de domaines a augmenté au fur et à mesure que de nouveaux types de repliements étaient découverts. L'analyse des structures de domaines à doigts de zinc répertoriés dans la banque PDB montre que la Résonance Magnétique Nucléaire (RMN) a contribué de façon très importante à l'étude de ces domaines. Cette observation s'explique par leur taille modeste (souvent inférieure à 100 acides aminés) qui facilite la détermination de la structure en solution mais également par le caractère dynamique de ces structures, qui les rend réfractaires au processus de cristallisation. Comme ces domaines interviennent très souvent dans des phénomènes de reconnaissance moléculaire, il est tentant de penser que ces comportements dynamiques sont étroitement associés à leur rôle de médiateurs d'interactions moléculaires. En donnant accès à la fois aux informations sur la structure et la dynamique de ces domaines, sous la forme isolée ou en interaction avec leur partenaire, la RMN constitue un outil très puissant pour leur étude. Au cours de ma thèse, je me suis intéressée à l'étude de deux types de domaines à doigts de zinc: (i) le domaine d'interaction à l'ADN du récepteur des hormones androgènes (AR DBD), (ii) une nouvelle famille de doigts de zinc, SCA7, que l'on trouve spécifiquement dans les ataxines, des protéines impliquées dans la régulation de la transcription.

### II. Etude du récepteur des hormones androgènes

La voie de signalisation du récepteur des androgènes (AR) est impliquée dans la progression du cancer de la prostate, et il a été montré que des mutations dans ce domaine étaient responsables de l'activation constitutive des gènes placés sous le contrôle des hormones androgènes. Une de ces mutations transforme un résidu thréonine du DBD en alanine (T575A). Des expériences permettant de mesurer l'activité de transcription ont permis à l'équipe du Dr. Ceraline à l'IRCAD de montrer que la mutation T575A induit un changement de spécificité du récepteur. Alors que l'activité de promoteurs placés sous le contrôle d'éléments de réponse spécifique de AR diminue, celle des promoteurs placés sous le contrôle d'éléments non spécifique

augmente. Ce changement de spécificité est corrélé à une modification de l'affinité du récepteur pour les éléments de réponse spécifiques et non spécifiques. Afin de comprendre le mécanisme de cette "reprogrammation" à l'échelle moléculaire, l'étude structurale des domaines DBD des récepteurs sauvage et muté a été entreprise par RMN. La comparaison des deux structures en solution a montré que la mutation n'altère pas le repliement du domaine et donc que la différence de reconnaissance des éléments de réponse n'est pas liée directement à la structure tridimensionnelle du domaine. Nous avons ensuite cherché à déterminer si l'altération de la fonction n'était pas due à une différence de dynamique de la chaîne peptidique. Afin d'étudier les mouvements moléculaires le long de la chaîne, des mesures de relaxation hétéronucléaire ont été effectuées et ont montré également une grande similarité dans le comportement dynamique des deux domaines, à l'exception d'une région située dans le premier doigt de zinc à proximité d'une histidine (H570), qui est conservée dans l'ensemble de la famille des domaines DBD des récepteurs nucléaires. Cette différence nous a conduit à mesurer, par RMN, le pKa de cette histidine pour les deux protéines. Nous avons ainsi montré que la mutation T575A induit une diminution de 0,5 unité de pH par rapport à la même histidine dans le domaine sauvage. L'analyse de la structure a permis de montrer que cette différence de pKa est liée à la perte d'une interaction entre le groupe hydroxyle de la thréonine 575 et le cycle imidazole de l'histidine. L'effet de la mutation sur le mécanisme de reconnaissance s'explique donc par un effet indirect dans lequel un acide aminé situé à distance de la région d'interaction modifie la surface électrostatique du domaine DBD. L'effet de la charge positive en position 570 sur la spécificité de reconnaissance de l'élément de réponse a ensuite été étudiée en construisant plusieurs mutants portant ou non une charge à cette position (mutants H570R et H570A). Ces études ont permis de confirmer l'importance de cette charge et l'ensemble de nos travaux fournissent un éclairage inédit sur les mécanismes de reconnaissance de l'ADN par les récepteurs nucléaires.

### **III. La famille des doigts de zinc SCA7**

L'ataxie cérébelleuse autosomique dominante (en anglais : spinocerebellar ataxia SCA) est une maladie génétique qui conduit à une perte progressive de la coordination des mouvements. Cette maladie est due à une expansion de trinuécléotides CAG codant pour une polyglutamine dans la région N-terminale du gène ATXN7, une protéine qui fait partie de SAGA, un complexe macromoléculaire impliqué dans la régulation de la transcription. Plus spécifiquement, la protéine ATXN7 forme, avec 3 autres protéines (ATXN7L3, USP22 et Eny2) le complexe de déubiquitination (DUB) au sein de SAGA. Ce sous-complexe intervient dans la modification post-traductionnelle des histones. Les deux protéines ATXN7 et ATXN7L3 sont composées de plusieurs domaines

similaires entre eux. Ainsi, le deuxième domaine de ces deux protéines est caractérisé par la présence de trois cystéines et d'une histidine conservées dans les séquences d'eucaryotes inférieurs et supérieurs. Des expériences de spectrométrie de masse ont permis de montrer que ces domaines lient un atome de zinc. La disposition des acides aminés coordonnant le zinc au sein de la séquence primaire de la protéine suggère la présence de doigts de zinc d'un type nouveau, baptisé SCA7 et caractérisé par une insertion de séquence importante entre les deux premiers coordonnants du zinc. Des expériences d'interaction avec le nucléosome ont été entreprises en utilisant la technique "GST pull-down" et ont permis de montrer que le domaine SCA7 de ATXN7 était capable d'interagir fortement avec les nucléosomes alors que celui de ATXN7L3 avait perdu cette capacité. L'interaction entre ATXN7 et les nucléosomes fait intervenir une reconnaissance spécifique entre le dimère d'histones H2A/H2B et le domaine SCA7. Afin de déterminer les bases structurales de cette interaction, nous avons déterminé la structure tridimensionnelle en solution des domaines SCA7 de ATXN7 et ATXN7L3. Cette étude a permis d'obtenir les premières structures de ce type de domaine à zinc. Les sites de coordination du zinc, de type CCCH sont conservés dans les deux protéines et correspondent à un repliement original, qui n'a pour le moment aucun équivalent dans la banque de structure PDB. L'extension de séquence observée entre les deux premières cystéines qui caractérise cette famille de doigts de zinc correspond à une structure de type "épingle bêta" qui pointe à l'extérieur de la structure. Les domaines SCA7 des protéines ATXN7 et ATXN7L3 comprennent également deux hélices alpha dont la position au sein du domaine diffère pour les deux protéines. Dans ATXN7, les deux hélices sont situées dans la partie C-terminale du doigt de zinc, elles se suivent en formant un coude à angle droit. Les deux hélices de ATXN7L3 sont quant à elles localisées en amont et en aval du doigt de zinc, et forment un faisceau de deux hélices anti-parallèles, stabilisé par un réseau conservé d'interactions hydrophobes. L'interaction entre le dimère d'histone H2A-H2B a également été étudié par RMN. La mise en présence d'un échantillon marqué de protéine ATXN7 avec les histones non marquées conduit à la disparition des signaux de ATXN7, confirmant ainsi l'interaction observée par voie biochimique. Cependant, cette disparition du signal, qui correspond à une constante d'affinité d'une dizaine de micro molaire ne permet pas d'envisager de faire la structure du complexe en solution. Par contre, il a été possible d'observer un phénomène de transfert d'aimantation entre les histones et le domaine à zinc de ATXN7 qui suggère l'implication des acides aminés situés aux extrémités du domaine. Des expériences de mutagenèse dirigée, basées sur ces observations ont permis par la suite de confirmer l'implication de la partie C-terminale de ATXN7 dans l'interaction avec le dimère d'histones.

## **PUBLICATIONS:**

1. J. Bonnet, Y.-H. Wang, R. A. Atkinson, J. Koffler, C. Romier, A. Hamiche1, L. Tora, D. Devys, and B. Kieffer

Histone deubiquitination by SAGA is modulated by an atypical zinc finger domain of Ataxin-7 (under revision in EMBO report)

2. Y.-H. Wang, A. Monge, J. Céraline, B. Kieffer and R. A. Atkinson

Altered specificity of a mutated androgen receptor in prostate cancer is associated with a change in the surface potential of the DNA-binding domain

## **Communications:**

2007 (June): 8th Course: Biophysics and the Challenges of Emerging Threats.

(NATO Advanced Study Institute and part of the International School of Biological Magnetic Resonance)

Ettore Majorana Center for Scientific Culture in Erice, Sicily (EMFCSC)

Présentation de poster: *Investigation DNA response element by androgen receptor DNA binding domain using NMR.*





## Chapter 1: The zinc finger proteins: A big family

### 1. The discovery of zinc fingers

#### 1.1 A brief history of zinc finger discovery

Zinc is an important metal for the growth and development in biological systems. Research into zinc deficiency was first studied in microorganisms in 1869 and showed the importance of zinc in nutrition for maintenance of life.[1] With advances in technology, various methods have been established to investigate the functions of zinc in living organisms. Zinc is involved in a wide variety of metabolic processes including carbohydrate, protein, nucleic acid and lipid synthesis and degradation.[2] In the past forty years, studies of zinc at the molecular level in enzymes have demonstrated that zinc participates in numerous catalytic reactions.[3] Moreover, zinc can also be incorporated into macromolecules to stabilize the overall fold. Structural studies characterizing zinc binding proteins using X-ray crystallography and NMR have provided detailed information about the metal binding geometry of coordination. The physical basis of zinc binding has also been studied in detail by computational simulations.[4]

The first study to suggest that zinc-containing proteins are able to regulate gene transcription by interacting with nucleic acids was by Miller *et. al.* [5] They demonstrated that transcription factor IIIA (TFIIIA) of *Xenopus oocytes*, which binds to 5S RNA and regulates transcription, binds 7 to 11 zinc ions and contains about 30 amino acids in each zinc binding unit. These repeats were characterized by a set of strictly conserved cysteine and histidine residues, with the  $X_{1-5}$ -h $X_1$ -Cys- $X_{2-4}$ -Cys- $X_3$ -h $X_5$ -h $X_{2-3}$ -His- $X_{3-4}$ -His sequence pattern where h represents a conserved hydrophobic residue. This repeating unit forms a small independent structure domain which was termed “zinc finger” (ZF). A similar motif of conserved cysteine and histidine residues was later found in the *gag* polyprotein from HIV which was shown to form a complex with RNA in the budding process of the retroviral life cycle.[6] Nucleocapsid protein type 7 (NCp7), which is a component of *gag* polyprotein, contains two conserved sequence Cys- $X_2$ -Cys- $X_4$ -His- $X_4$ -Cys, which suggests an ability to bind zinc.[7] To elucidate the function of zinc fingers in nucleic acid binding, Berg proposed a structural model for the metal-binding domain by searching systematically for the known structural motifs in other proteins.[8] He classified zinc-binding proteins involved in nucleic acid binding or gene regulation according to their sequence similarity. The first structural information was provided by the solution structures of two synthetic peptides and explained the use of sequence patterns in the two peptides.[9, 10] The first natural structure of zinc finger proteins, the HIV zinc finger like domain, which contains the following sequence

C-X<sub>2</sub>-C-X<sub>4</sub>-H-X<sub>4</sub>-C and represents the CCHC type zinc finger motif was determined by NMR one year later[11]. At the same time the structure of another zinc finger, the glucocorticoid receptor DNA-binding domain (GR-DBD), was solved by NMR.[12] Unlike the CCHC type zinc finger motif, GR-DBD was found to contain two zinc-binding sites with CCCC motifs. Similar to the involvement of TFIIIA in interaction with DNA, GR-DBD is able to regulate transcriptional activity via DNA binding. However, molecular insight into the interaction of this zinc-binding domain with nucleic acids was still unclear.

In 1991, the first complex of DNA bound to the zinc finger protein Zif268 was solved by X-ray crystallography.[13] Zif268 contains three C<sub>2</sub>H<sub>2</sub> zinc fingers and the structure revealed the origin of DNA binding specificity. Each finger binds to the major groove of DNA through its N-terminal  $\alpha$ -helix. The DNA binding contacts involve hydrogen bonding interactions between residues in the  $\alpha$ -helix and three successive bases on one strand of the DNA. The complex suggested a strategy for the design of artificial zinc fingers with specific DNA binding properties. At the same time, another zinc finger family, the RING finger, was characterized by Freemont *et al* .[14] Unlike the nuclear receptor DNA-binding domains, where the sets of residues involved in each of the two zinc coordination sites are arranged sequentially, the RING finger showed for the first time a cross-braced arrangement of the pairs of zinc binding residues.[15] Moreover, the RING finger was found to mediate protein-protein interactions rather than protein-DNA interactions associated with the typical C<sub>2</sub>H<sub>2</sub> zinc finger.[16] Since then, more and more different types of zinc fingers have been characterized and shown to be involved in protein-protein interactions, including LIM domains and PHD fingers. LIM domains contain a sequential-braced zinc-binding topology whereas PHD fingers are similar to RING finger, with a cross-braced zinc-binding topology. The diversity of zinc finger proteins is reflected in the number of genes encoding zinc finger proteins. In 2000, the availability of the human genome allowed another step forward in our knowledge of zinc fingers. Screening the sequence database for zinc binding motifs revealed a previously underestimated level of diversity and a large number of genes potentially encoding zinc fingers. It is estimated that about 10% of the human genome encodes zinc finger containing proteins.[17] In 2001, The FYVE domain which named after the first letter of the first four proteins in this class **F**ab1p, **Y**OTB, **V**ac1p and **E**EA1 was characterized in proteins that regulate lipid signaling and membrane trafficking.[18] The TAZ2 domain in CREB-binding protein (CBP) was shown to regulate p53 activity.[19] With the development of high field NMR spectroscopy and high-throughput crystallization technology, a growing number of zinc finger structures have been determined.

Despite the growing amount of experimental data, a number of questions remain to

be answered. These questions include the relationship between the fold and recognition properties, the variability in the dynamics of metal binding and the evolutionary mechanisms of the folds stabilized by metal ions. Functional exploration of zinc fingers related to diseases including cancer or genetic disorders will be the trends in future.

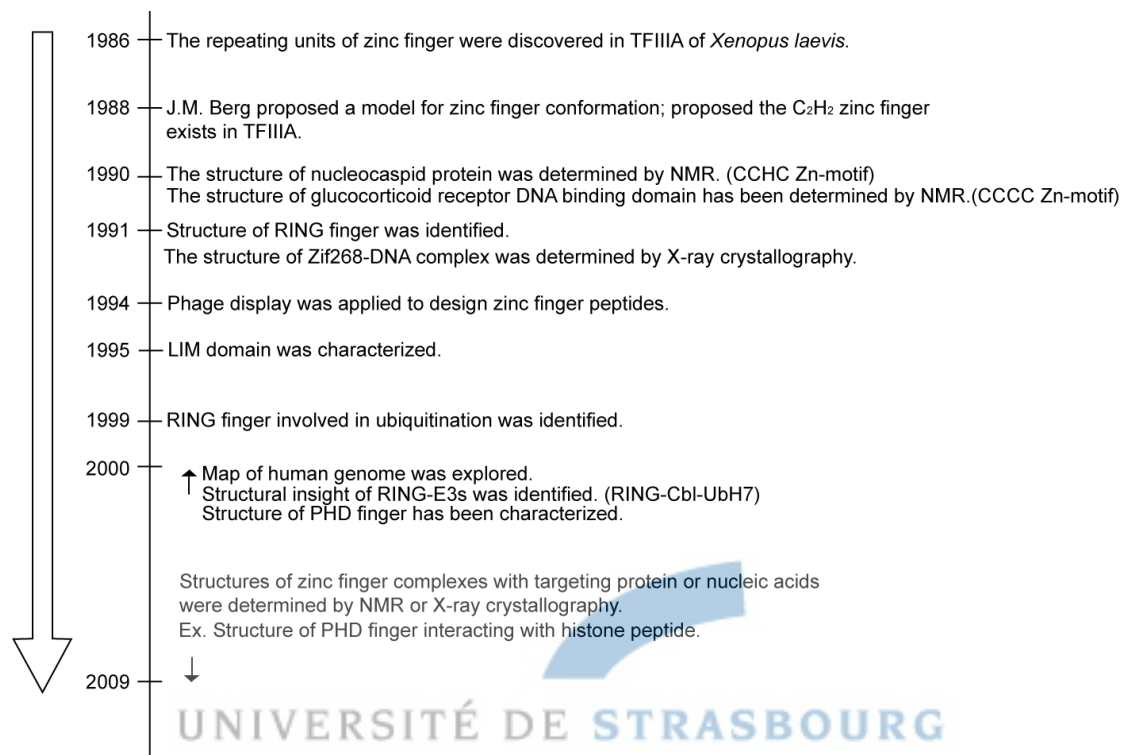


Fig.1 Timetable of zinc finger development over the past twenty years.

## 1.2 Zinc fingers today: Survey of the PDB

The growing number of zinc fingers reveals the diversity of folds stabilized by coordination of a zinc ion that was not originally predicted. This diversity is derived from the wide range of sequence features that satisfy the requirement of zinc coordinating residues but vary in the organization of protein chains around the zinc coordinating sites. The diversity contrasts with a previous underestimate of the role of zinc fingers in the evolution of proteins.[17] In addition, the discovery of different zinc fingers folds over the past twenty years has stimulated our interest in understanding the molecular interactions mediated by protein domains. The term “zinc fingers”, which was first proposed by Jeremy Berg after comparing pairwise positions of zinc-coordinating residues in primary sequences, became inappropriate after alternative zinc coordination sequence patterns such as that of RING domains were discovered. As a result, a survey based on the information of sequences and structures is absolutely necessary. A survey can be achieved by searching according to the strict conservation of the zinc coordinating residues in the sequence database, and the presence of a zinc atom in the PDB structural database. The survey presented below

took advantage of the links that exist between databases gathering sequences motif and multiple alignments (PROSITE and Pfam) and the PDB.

The keyword “zinc” gave 203 entries in Pfam including catalytic and structural zinc binding proteins. Excluding catalytic zinc binding proteins reduced this number to 99 entries. An independent search of the keyword “zinc” in PROSITE search gave 91 entries including some catalytic zinc proteins. The intersection of these two sets of proteins contains 52 families belonging to the zinc finger family. The associated structural information was retrieved using the direct link that exists between PROSITE and PDB entries. The initial analysis gave about 600 structures of zinc fingers determined by NMR or X-ray crystallography. Of these, 72% were determined by NMR, reflecting wide accessibility of structural determination using NMR. It may be because the NMR facilitates the structural study of proteins with low molecular weight and proteins that exist stably in solution. A timetable of the publication of zinc finger structures (Fig.2) provides interesting insight into our understanding of zinc fingers. During 1990 to 2000, the number of structures of zinc fingers was quite low and began to grow after 2000. This trend can be explained in part by initiation of several structural genomics programs in various parts of the world. The number of structures of zinc finger reached a maximum in 2007 when more than 200 structures were released. An analysis of the corresponding sequences shows that most structures belong to the C<sub>2</sub>H<sub>2</sub> DNA binding family (Fig.2B). The structural information available on C<sub>2</sub>H<sub>2</sub> zinc fingers allows the rules of molecular recognition for this system to be deduced, allowing the design of synthetic C<sub>2</sub>H<sub>2</sub> zinc fingers targeting specific DNA sequences.

Fig.3 presents the appearance of diverse zinc finger families in the PDB since the discovery of C<sub>2</sub>H<sub>2</sub> zinc fingers. The solution structure of the N-terminal “zinc knuckle” of the retrovirus provides the first structural insight into zinc coordination by cysteine and histidine residues.[11] At the same time, the structures of nuclear receptor DNA binding domains and GATA-like proteins were also released, resulting in the erroneous association of zinc fingers exclusively with DNA binding properties. Evidence that zinc finger domains were also involved in protein-protein interaction emerged only with the discovery of LIM domains, despite the fact that the first structure of the RING domain had been solved earlier.[20-22] The zinc binding motif in RING domains involves a cross-braced arrangement of the eight zinc coordinating residues in the primary sequence which is quite distinct from the sequential-braced topology of that in LIM domains.[23] These findings suggested that the structural organization of the zinc binding sphere could be more diverse than originally predicted.

Before 2000, in addition to the discovery of new folds of zinc fingers, studies of individual zinc finger proteins focused largely on structural and functional characterization and identification of the targeting function of each protein individually.

The first example of the post-genomic zinc finger research was provided by the structure determination of the complex of the human bromodomain and the PHD domain transcription factor (BPTF) complex with a histone peptide.[24] This should be not only explained the binding affinity of PHD interacting with a modified histone peptide but also suggested a function of the PHD domain in histone modification. As a result, the study of zinc fingers no longer emphasizes the “pure” zinc finger structure, but focuses on the cooperation of zinc fingers with other functional domains, for example, the roles of RING fingers in the regulation of E3 ligase and ubiquitination. Much remains to be done to fully characterize the diverse roles of zinc fingers. For example, domains such as Dof-type and DHHC-type zinc fingers have been studied at the gene level but the structures have not yet been solved. Dof-type (DNA-binding with one finger) zinc fingers are found in plants and are known to function as transcription factors.[25] DHHC-type zinc fingers regulate palmitoyl transferase activity.[26]



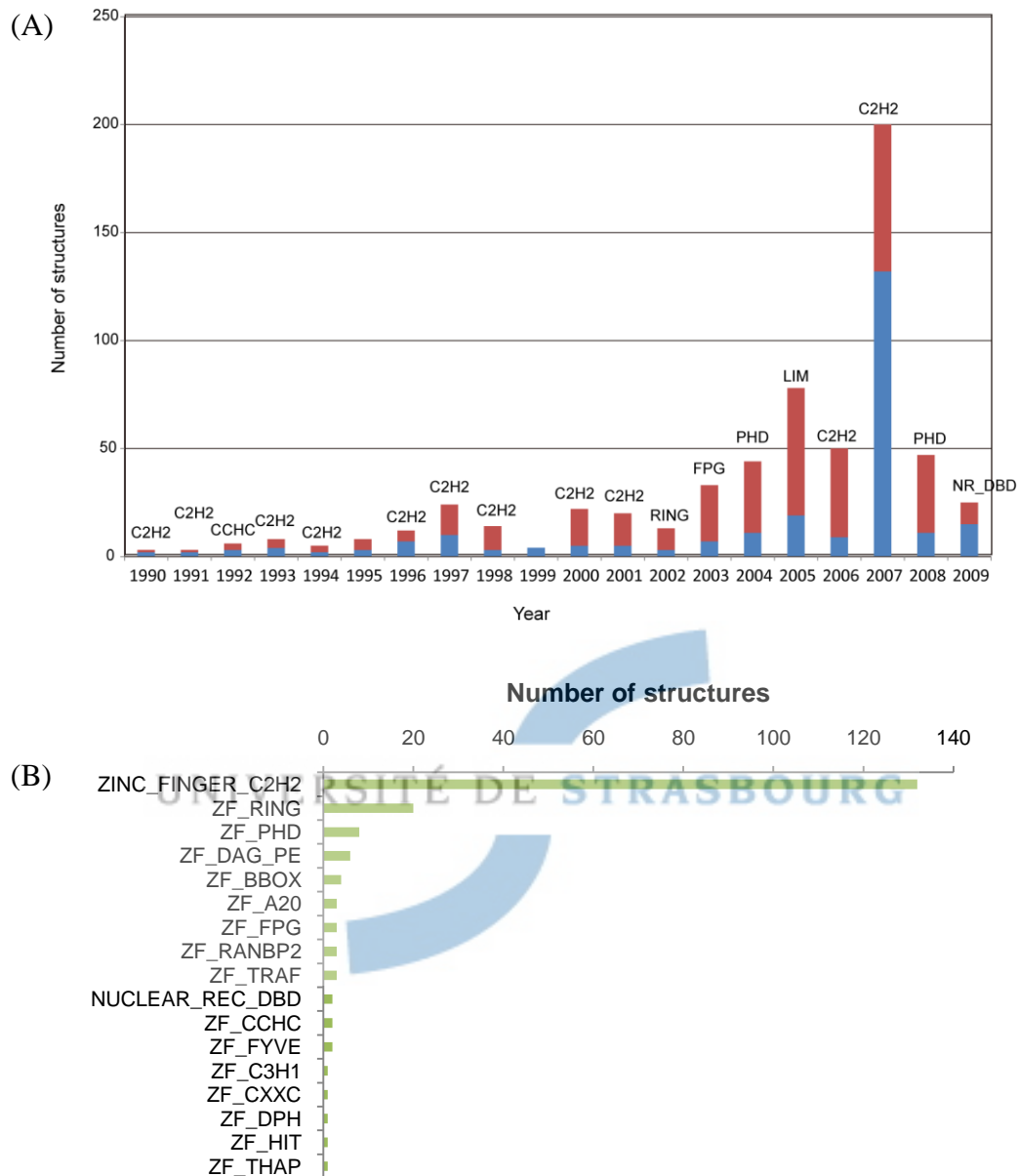


Fig.2 (A) A timetable of zinc finger protein structure determination over the past twenty years. The family for which the largest number of structures were determined in a particular year is labeled on the top and represented by the red bar. In 2007, the number reached a maximum due to development of zinc finger engineering. (B) Comparison of different types of zinc finger for which the structures were solved in 2007. The number of classical  $C_2H_2$  zinc finger takes first place in 2007 due to progress in zinc finger engineering. In second place is the RING finger, which reflects the breakthrough in the exploration of the function of RING fingers.

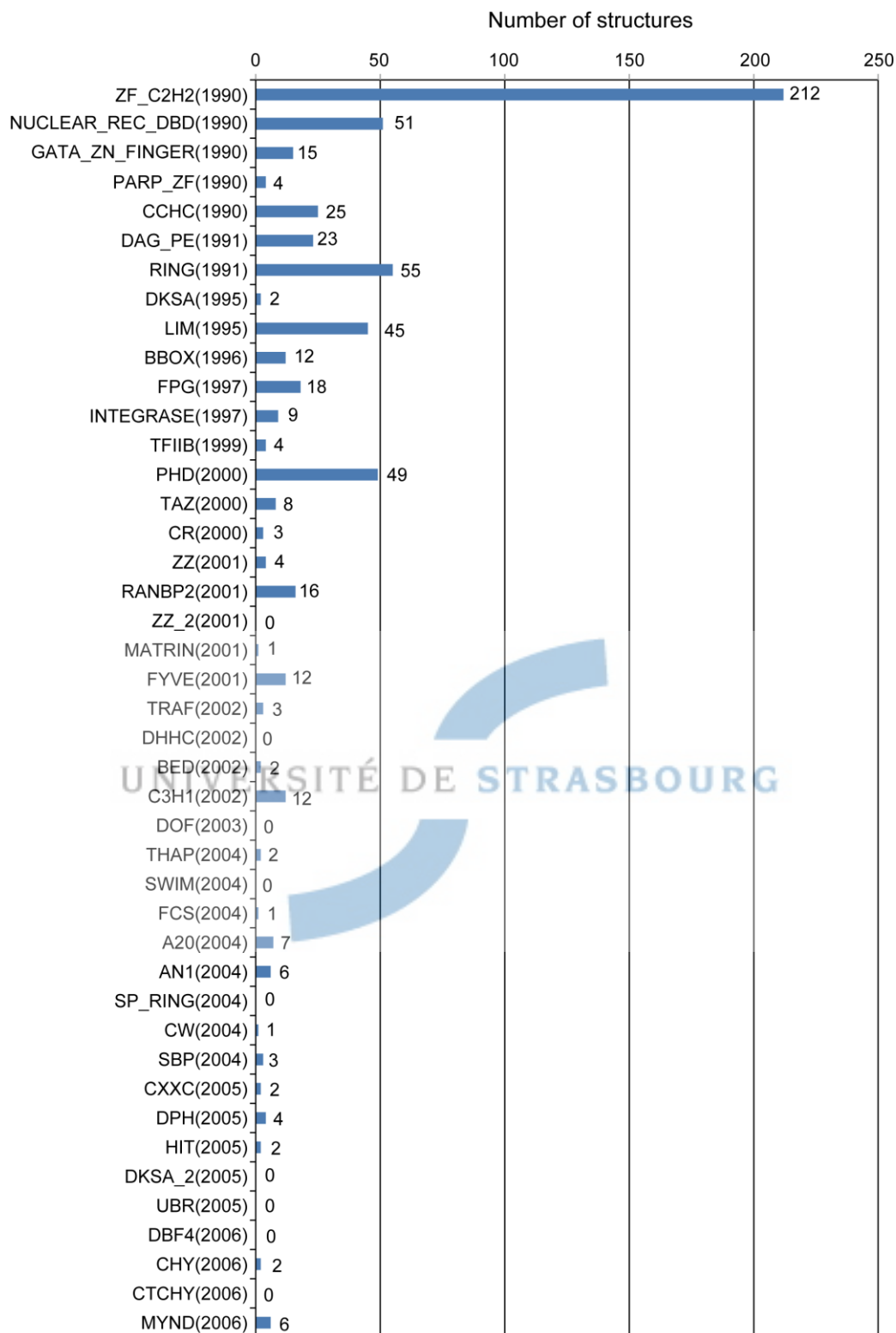


Fig.3 An estimate of the number of zinc finger protein. 52 zinc finger families acquire from Pfam and sorted by the year. The classical  $C_2H_2$  zinc finger is the most abundant structure, constituting 34% of all zinc finger structures.[27] The nuclear receptor DNA-binding domain is the second followed by RING finger, PHD, LIM domain, etc.



## **2. Current view of zinc finger evolution**

### **2.1 Gene duplication of zinc fingers is a hint for evolution**

As observed in the survey of zinc finger proteins in the PDB, the structures of many zinc finger proteins have been determined and new folds have been identified. The diversity of zinc finger proteins may be related to evolutionary pathways. The repeated duplication and diversification of existing genes are common features of the genomes of higher organisms.[28] Zinc finger proteins are found not only in eukaryotes but also in prokaryotes.[29] In prokaryotes, they are distributed in bacterial species for which the ecological relation is close to eukaryotes. The genes encoding zinc fingers in bacterial species derive from eukaryotic genomes through horizontal gene transfer.[30] They belong to the  $\alpha$ -subdivision of proteobacteria which originated from mitochondria.[31]

In the human genome, about 700 genes encoding for zinc fingers are estimated to code for the transcription factors.[32] The number of these genes in human is significantly higher than that in lower eukaryotes. 37% of human zinc finger genes are located on chromosome 19.[33] Analysis of the gene family indicates that these genes arose from an ancestral group of eukaryotic zinc finger transcriptional factors through a series of gene duplications. The ancestral gene family may be extended independently in several lineages, including mammals and fish. This expansion in divergent lineages suggests that the zinc finger gene family provides a framework for rapid transcriptional evolution that has been utilized repeatedly. In addition, changes in physical functions may also contribute to the evolution of zinc finger genes. For example, changes in DNA binding specificity of zinc fingers may be the major selective pressure for these genes to adapt to changes in the environment. [34] This term “adaptive evolution” of zinc fingers is not only able to contribute to the expansion and duplication of the numbers of zinc fingers but also to increase the diversity of zinc finger types.

### **2.2 Zinc finger associated domains**

To confirm the phenomena of gene duplication and expansion, the route of evolution of genes encoding the zinc fingers has been studied. Genetic analysis has shown that nearly half of zinc fingers were  $C_2H_2$  zinc fingers, representing 2% of all human genes.[35] The remarkable feature of  $C_2H_2$  zinc finger proteins is that they contain additional protein domains located at their N-termini. These associated domains include Kruppel-associated box (KRAB), the poxvirus and zinc finger (POZ) domain also known as the BTB domain, SCAN domains and the insect zinc finger associated domain (ZAD). These domains define subgroups within the  $C_2H_2$  family and may provide insights into the functions of zinc finger transcription factors.

Of these domains, over half are associated with KRAB and SCAN domains in the



human genome. KRAB-associated proteins were discovered in 1991 by Bellefroid *et al.*[36] Notably, KRAB-associated domains exist only in tetrapod vertebrates. They contain multiple C<sub>2</sub>H<sub>2</sub> zinc fingers which bind to KRAB domains and to DNA. SCAN domains are leucine rich domains, located near the N-terminus of C<sub>2</sub>H<sub>2</sub> zinc finger proteins. SCAN domains are found in vertebrates but are not associated with C<sub>2</sub>H<sub>2</sub> zinc finger domains in lower vertebrates.[37] Some SCAN family members may also contain a KRAB domain which are present in zinc finger genes following the addition of KRAB domains to the zinc finger genes.[28] Unlike the KRAB domain which is a transcription repressor, the SCAN domain is not able to regulate directly transcriptional activation or repression. In contrast, SCAN domain is like an adaptor that binds to other SCAN proteins forming homo or heterodimers. Unlike KRAB and SCAN domains that exist only in mammals, BTB/POZ domains are found in all animal species. PLZF (promyelocytic leukemia zinc finger) is a typical member of the group of protein containing a BTB/POZ domain. It functions as a transcriptional repressor by binding to the promoter of target genes involved in the regulation of the cell cycle. ZAD is a recently discovered zinc finger domain,[38] the genes for ZAD proteins being specific to the genomes of insects. The largest group of ZAD domains was found in *Drosophila melanogaster*. However, only few ZAD-coding genes in *Deosophila melanogaster* that have been identified are conserved in evolution.[39]

The evolutionary pathways of these zinc finger-associated domains exhibit the species-specific properties which are found either in mammal or other vertebrates. The set of C<sub>2</sub>H<sub>2</sub> zinc finger-associated domains has expanded considerably through evolution, and independently in the manner referred to as lineage-specific expansion. Although the functions of many of family members are not clear, associations of these domains with zinc fingers provide “markers” for the identifications of zinc fingers and may facilitate zinc fingers binding to DNA and thus regulating transcription activity.

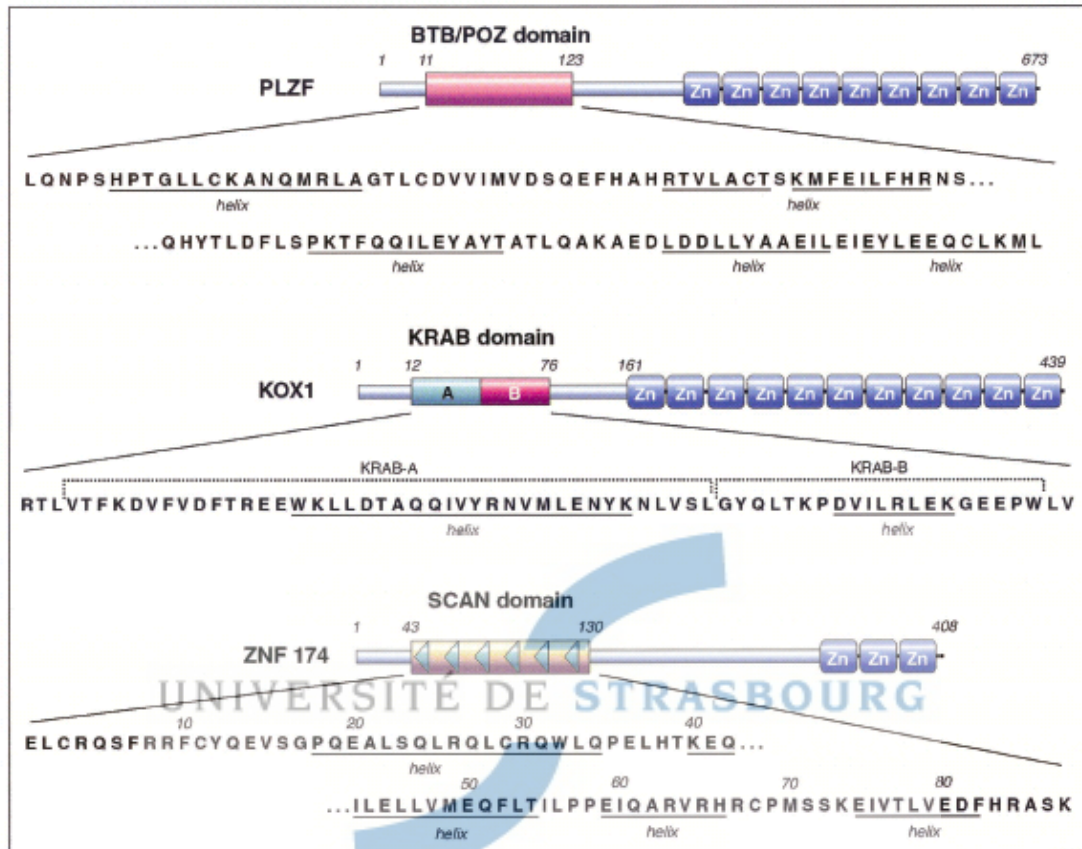


Fig.4 Three types of Cys<sub>2</sub>His<sub>2</sub>-type zinc finger-associated domains.[40] The structure of representative BTB/POZ (PLZF), KRAB (KOX1), SCAN (ZNF174) domains are shown schematically. PLZF consists of a single BTB/POZ domain located at the N-terminus of the protein followed by nine C<sub>2</sub>H<sub>2</sub> Kruppel-type zinc fingers. The KRAB domain is a transcriptional repressor. Unlike BTB/POZ domain, the KRAB domain exists only in vertebrates. It can be divided into A and B boxes where the A box plays the major role in transcriptional repression. Like BTB/POZ domain-containing proteins, KRAB domain-containing proteins also contain several zinc fingers near the C-terminus. SCAN domains, unlike BTB/POZ and KRAB domain-containing proteins, can exist as an isolated SCAN domain protein. It is also found linked to KRAB domains to form a KRAB-SCAN-KRAB-(C<sub>2</sub>H<sub>2</sub>)<sub>x</sub> domain alignment, for example, the hypoalphalipoproteinemia susceptibility gene, ZNF202 encodes a protein predicted to contain a SCAN box associated with KRAB domain and eight C<sub>2</sub>H<sub>2</sub> zinc finger motifs.[41]

### 3. Functions of zinc finger proteins

To date, numerous types of zinc finger domains have been identified and the numbers continue to grow. Currently there are at least 14 distinct, well-defined classes known,[42] involved either in nucleic acid binding or protein-protein interactions. Some families, such as the GATA family, possess dual functions including both protein-protein interaction and DNA binding activity. More generally, zinc finger proteins can be grouped into two major groups based on the function: nucleic acid binding and protein-protein interaction.

#### 3.1 Zinc finger involved in nucleic acid binding

Following the discovery of TFIIIA which is able to regulate transcriptional activity, the zinc finger was thought originally to serve as a DNA binding regulator. Of the family members that do bind nucleic acids, C<sub>2</sub>H<sub>2</sub> zinc finger is the most common type and applied widely in drug development. The nuclear receptor DNA-binding domain, as its name suggests mediates the DNA binding activity of nuclear receptors involved in the transcription process. Gal4 protein binds two zinc ions via six cysteines and is able to bind half-site DNA. These diverse family members yield various insights into DNA recognition.

##### 3.1.1 Classical C<sub>2</sub>H<sub>2</sub> zinc finger

The C<sub>2</sub>H<sub>2</sub> zinc finger was first discovered in the *Xenopus laevis* transcriptional factor IIIA and contains 30 amino acids. The consensus sequence of C<sub>2</sub>H<sub>2</sub> zinc finger is X<sub>1-5</sub>-hX<sub>1</sub>-C-X<sub>2-4</sub>-C-X<sub>3</sub>-hX<sub>5</sub>-hX<sub>2-3</sub>-H-X<sub>3-4</sub>-H where X represents any amino acid and h is a hydrophobic residue. These sequences fold to form a ββα domain which is stabilized by a zinc ion. The zinc is chelated between two cysteine residues located between the β-strands and two histidine residues on the α-helix. In addition, a conserved hydrophobic core serves to stabilize the fold. Mutation of zinc binding ligands leads to a loss of both metal coordination and the structural fold.[43]

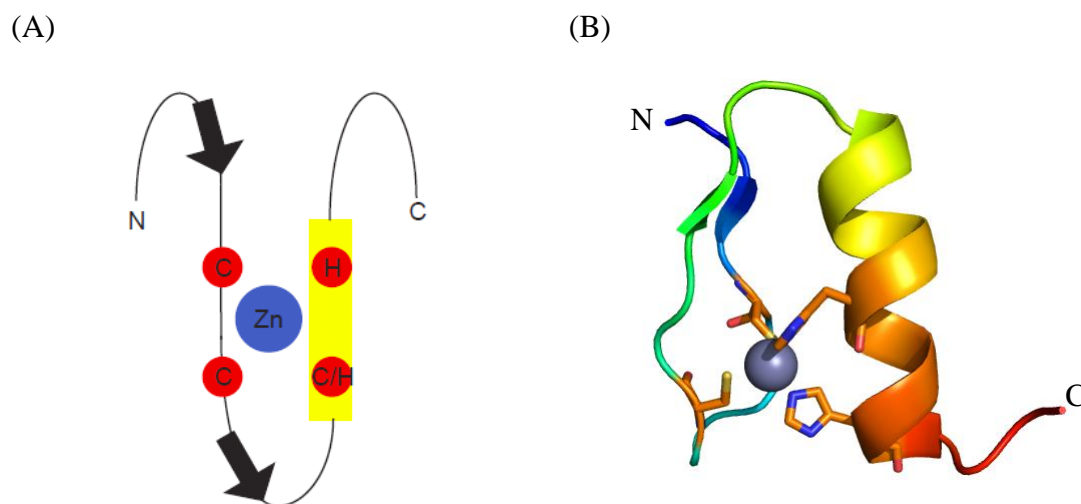


Fig.5 (A) Topology of the classical  $C_2H_2$  zinc finger. The zinc ion is coordinated by two cysteine and two histidine residues. (B) The structure of zinc finger in Zif268 (PDB code: 1aay).[44] It consists of two short  $\beta$ -strands, a short loop and an  $\alpha$ -helix which is nearly parallel to the  $\beta$  hairpin.

Most  $C_2H_2$  zinc fingers are transcriptional factors that are able to recognize specific DNA sequences. The DNA binding site of Zif268 is formed by the  $\alpha$  helix of the zinc finger which binds in the DNA major groove through hydrogen bond interactions from helical position -1, 3 and 6 to one strand of DNA and through a secondary interaction from helical position 2 to the other strand (Fig.6B).[45] Comparison of  $C_2H_2$  zinc finger-DNA interactions involving Zif268 have shown that the vast majority of the base-specific contacts in the zinc finger-DNA complexes are made from positions -1, 2, 3 and 6 of the  $\alpha$ -helix.[46]  $C_2H_2$  zinc fingers are arranged as repeated arrays in order to achieve binding selectivity and specificity that could not be achieved using a single domain. The DNA binding properties of Zif268 provided a strategy for zinc finger design and was used to develop novel drugs.[47]

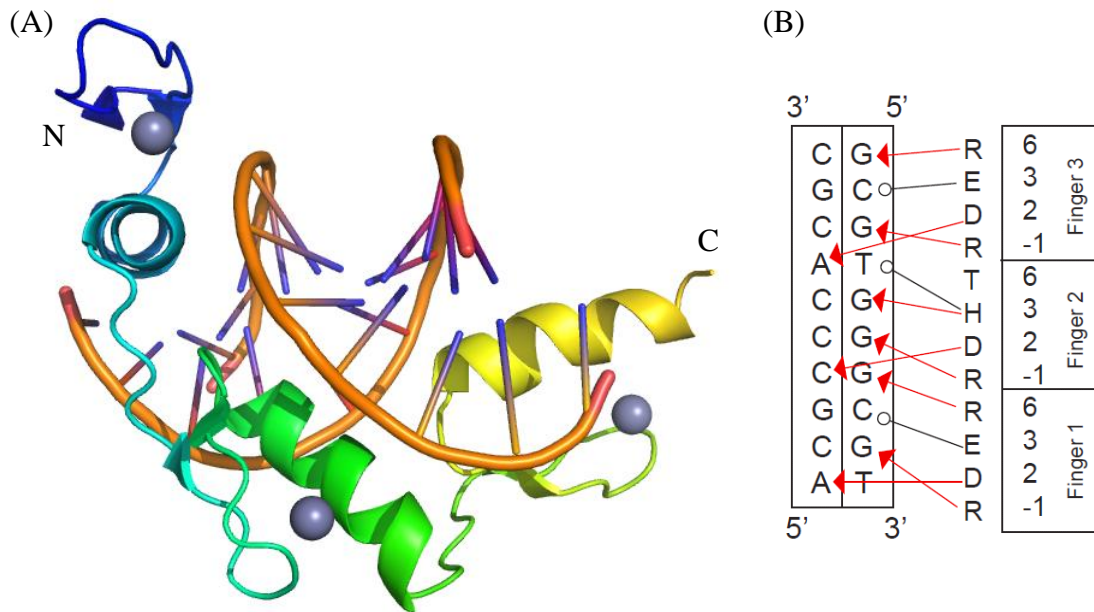


Fig.6 (A) Structure of zinc finger protein Zif268 in a complex with DNA (PDB code: 1aay).[13] Zif268 consists of three zinc fingers each with a  $C_2H_2$  motif. All three Zif268 zinc fingers have identical residues at positions -1 and 2. The  $\alpha$ -helical portion of each finger fits in the major groove of the DNA. Base contacts made from positions -1, 2, 3 and 6 of each  $\alpha$ -helix are indicated in (B). Arrows indicate contacts mediated by hydrogen bonds; open circles indicate hydrophobic interactions.

### 3.1.2 Nuclear receptor DNA binding domain

Nuclear receptors (NRs) are transcriptional factors that regulate the activity of transcription. NRs contain two conserved domains: the ligand-binding domain (LBD) and DNA-binding domain (DBD) separated by a “hinge” domain and flanked by less-conserved domains.[48] NR DBDs contain eight highly conserved cysteine residues that coordinate two zinc ions (fig.7A). Removal of the zinc ions leads to protein unfolding and a loss of DNA binding activity.[49] A subset of NR DBDs binds to hormone response elements as homodimers. The response elements comprise six-nucleotide half sites that are arranged as inverted or direct repeats. However, in solution, the isolated NR DBD exists as monomer. The glucocorticoid receptor (GR) DBD was one of the first zinc finger structures to be determined (Fig.7A).[12] It folds into a globular domain consisting of two zinc-finger structures (Fig.7B). In each zinc-binding motif, the third and fourth cysteine ligands are found at the N terminus of an  $\alpha$ -helix. The two  $\alpha$ -helices are almost perpendicular, held together by a hydrophobic core. However, the roles of the two zinc binding modules differ and cannot be considered as independent units. The first zinc binding module contains two short anti-parallel  $\beta$  strands which partly contact phosphate groups of the DNA and the second pair of zinc coordinating cysteine lie on an  $\alpha$ -helix that makes contact with the major groove of

DNA (Fig.7B). The second zinc binding module provides the dimerization interface.[50] Unlike the  $C_2H_2$  zinc finger in which a single  $\alpha$ -helix of each zinc finger is involved in DNA interactions, NRs interact with DNA through the dimerization of DBD. The arrangement of NRs DBD on the DNA can be either head-to-head or head-to-tail. The homodimer interacting with DNA is symmetric and each DBD contributes the same region to the dimerization interface. The structure of GR DBD bound to GR response element ( $GRE_{4S}$ ) shows that the receptor binds as head to head homodimers (Fig.7C).[51] In contrast, the vitamin D receptor (VDR) DBD bound to DR3 element reveals a head-to-tail protein dimer.[52] Therefore, although the structures of NR DBD are quite similar and the sequences are highly conserved, binding to DNA response elements is selectively achieved by a combination of protein-DNA interactions involving both receptor monomers.





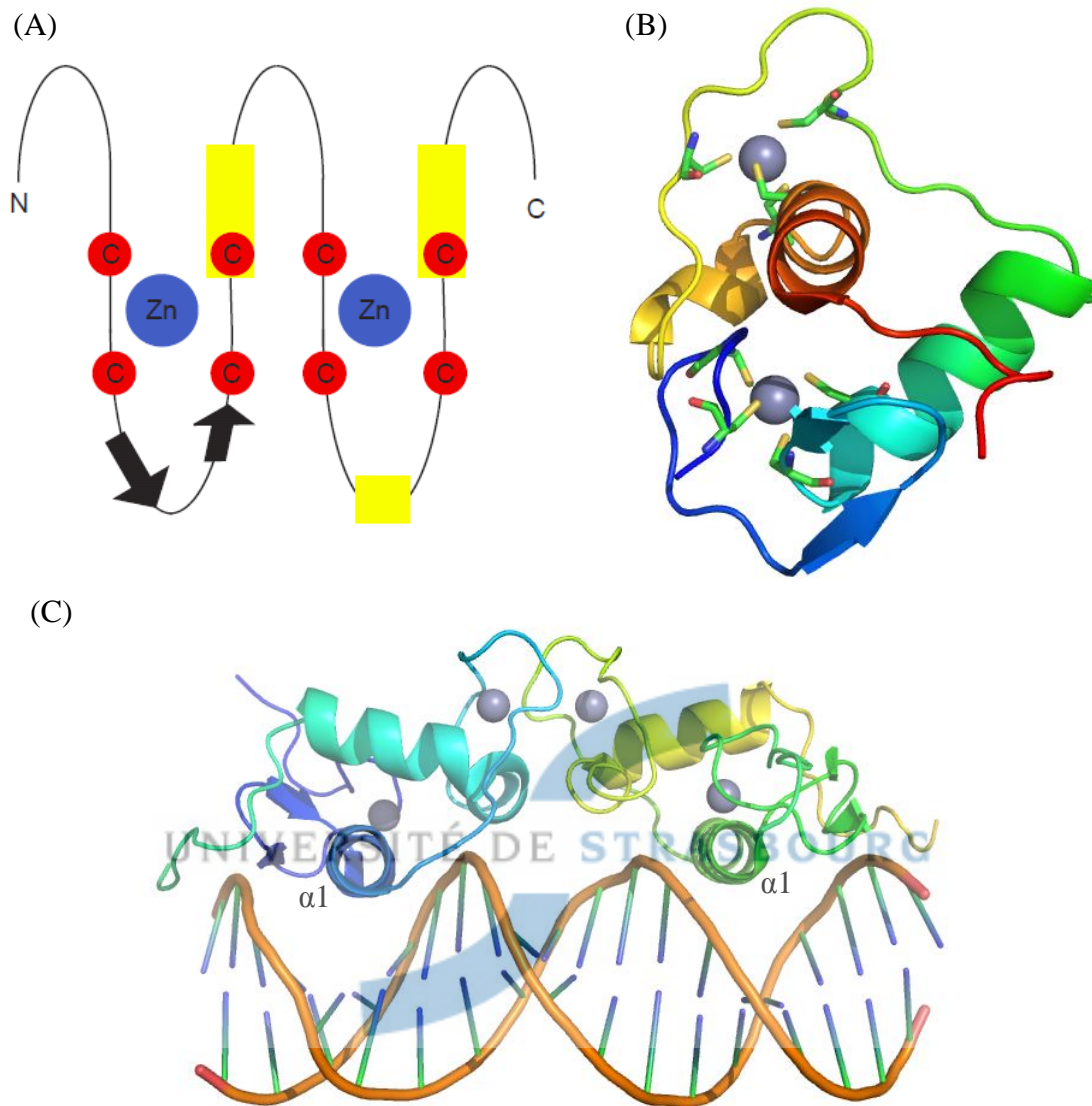


Fig.7 (A) The zinc binding sites of nuclear receptor DNA binding domains are arranged sequentially. In each zinc binding site, the third and fourth cysteine ligands lie on an  $\alpha$ -helix. The zinc binding motifs fold as the treble clef which is assembled around the center zinc ion and consists of a zinc knuckle, loop,  $\alpha$ -helix and  $\beta$ -hairpin.[53] (B) Structure of glucocorticoid receptor DNA binding domain (GR-DBD) (rat) (PDB code: 1gdc).[54] The first zinc module is responsible for DNA recognition and the second zinc binding module is involved in dimerization. (C) Structure of GR-DBD binding to GR response element (IR3) (PDB code: 1r4o).[51] GR-DBD binds to IR3 as a dimer. The dimerization of GR-DBD results in a head-to-head conformation. The  $\alpha$ -helix of the first zinc binding module in each GR-DBD monomer is responsible for DNA binding, contacting the major groove of DNA.

### 3.1.3 Gal4 transcription factor

The Gal4 transcription factor is 881 amino-acid protein which regulates transcription activation in the presence of galactose.[55, 56] The DNA binding activity of Gal4 is located in the N-terminal region and, in particular, a cysteine-rich domain with a conserved C<sub>11</sub>-X<sub>2</sub>-C<sub>14</sub>-X<sub>6</sub>-C<sub>21</sub>-X<sub>6</sub>-C<sub>28</sub>-X<sub>2</sub>-C<sub>31</sub>-X<sub>6</sub>-C<sub>38</sub> sequence. This cysteine-rich sequence suggested that GAL4 DNA binding domain could be a zinc finger motif. <sup>113</sup>Cd NMR showed, however that Gal4 bind to two Cd(II) ions, indicating the formation of a binuclear metal cluster involving the six cysteine residues rather than a typical zinc finger motif.[57] The NMR structure shows that residues 7-40 of free Gal4 form a Zn<sub>2</sub>C<sub>6</sub> binuclear cluster while residues 41-65 are unstructured. The binuclear cluster domain of Gal4 is a monomer in solution but binds to DNA as a homodimer. The crystal structure of Gal4 binding to DNA shows that an ordered extended loop region and a coiled-coil dimerization element contact CGG DNA half-sites.[58] The extended coiled-coil region consists of three  $\alpha$ -helices connected by flexible loops. This helical bundle region not only mediates DNA binding activity but also activates transcription by interaction with other transcription factors such as Gal11P.[59]

The logo of the University of Strasbourg, featuring two blue curved segments that form a stylized 'S' shape.

UNIVERSITÉ DE STRASBOURG



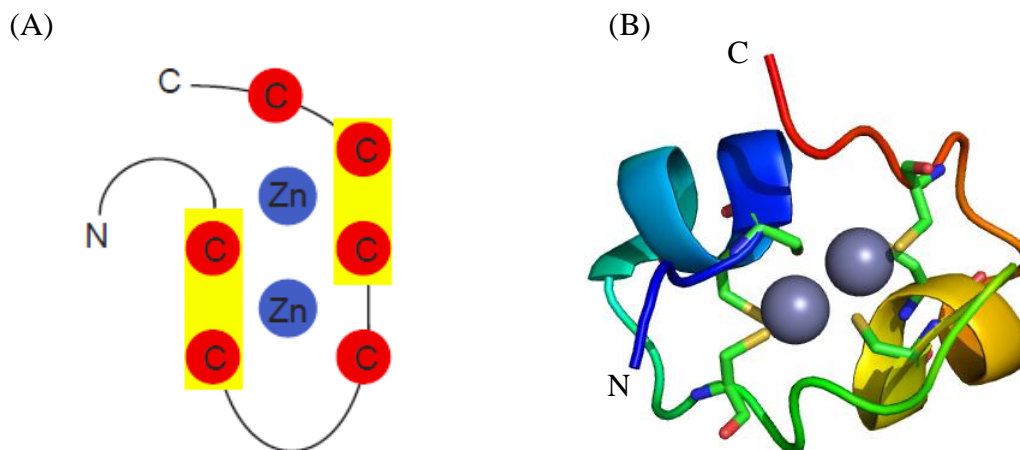


Fig.8 Structure of the zinc-binding DNA recognition module (residue 8-40) of GAL4 (PDB code: 3coq).[59] The core consists of two zinc ions coordinated by six cysteine ligands. Two of the cysteines ligate both zinc ions, creating a “binuclear cluster”. The structure is similar to that of metallothioneins. The folded polypeptide consists of two  $\alpha$ -helices, each followed by an extended strand.

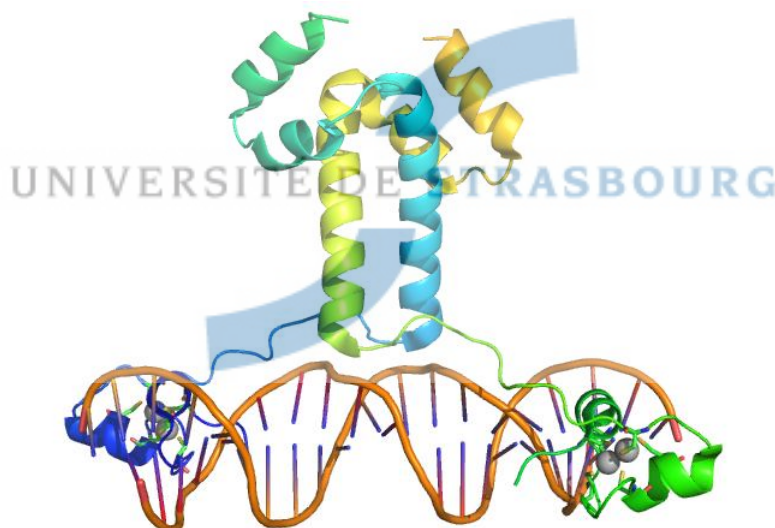


Fig.9 Structure of Gal4/DNA complex (PDB code: 3coq).[59] Gal4 interacts with DNA as a dimer. The dimer interface is formed by packing of the helices, two of which form a parallel coiled-coil. The coiled-coil is positioned over the DNA minor groove and perpendicular to the DNA helix axis. In addition, the two binuclear clusters contact the major groove of the DNA centered over the conserved CCG triplets.

### 3.1.4 Nucleocapsid protein

The nucleocapsid (NC) is a viral capsid protein which is associated with viral nucleic acids. Nucleocapsid protein type7 (NCp7) was isolated from human immunodeficiency virus (HIV) type 1.[60] NCp7 encapsulates and protects viral dimeric unspliced RNA.[61] It contains two successive zinc fingers with Cys-X<sub>2</sub>-Cys-X<sub>4</sub>-His-X<sub>4</sub>-Cys

motifs linked by basic residues.[62] The aromatic residues forming hydrophobic patches on the zinc finger surfaces are essential for RNA recognition.[63] In addition, NCp7 can also recognize DNA sequences. The structure of NCp7 bound to single-stranded DNA shows DNA recognition by the two zinc fingers which behave as flexible independent domains (Fig.10).[63, 64] The linker between the two zinc fingers is important for folding and interaction; deletion of this short sequence leads to a complete loss of protein activity.[65] Although the N- and C-terminal parts of NCp7 are highly flexible, NCp7 exhibits a dramatic reduction in conformational flexibility upon nucleic acid binding.[66]

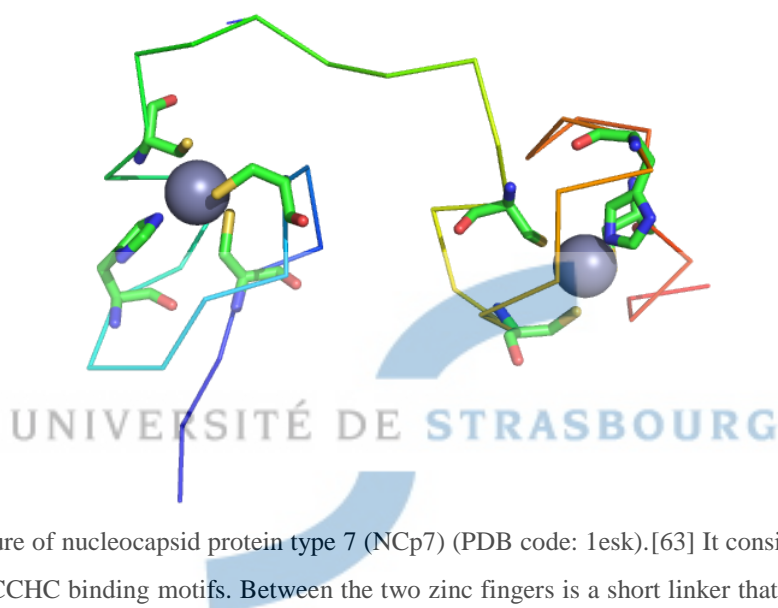


Fig.10 Structure of nucleocapsid protein type 7 (NCp7) (PDB code: 1esk).[63] It consists of two zinc fingers with CCHC binding motifs. Between the two zinc fingers is a short linker that contains basic residues and is important for RNA interaction.

### 3.1.5 Zinc finger engineering

The crystal structure of Zif268 bound to target DNA revealed the molecular basis for the specificity of DNA sequence recognition. The diversity of  $C_2H_2$  zinc fingers in eukaryotes effectively samples large numbers of recognized DNA sequences, creating a “recognition code” (Fig.11). This recognition code can be exploited to design artificial transcription factors to regulate the specific genes for therapeutic purpose. Based on the DNA binding specificity of Zif268, the strategies for the use of the  $C_2H_2$ -type zinc finger as a template were drawn along the following lines:

1. One finger only recognizes 3 to 4 contiguous base pairs of the DNA sequence. The recognition site is positioned specifically at -1, 3 and 6 on the  $\alpha$ -helix and is responsible for DNA contact. The RDER motif in each zinc finger provides the binding contact for DNA interaction (Fig.6B).
2. The  $C_2H_2$ -type zinc finger proteins typically contain several zinc fingers linked by

a short consensus linker, TGE[K/R]P sequence. Only two or three zinc fingers linked in tandem array are sufficient for specific DNA binding with no involvement of other protein partners.

3. Most importantly, tandem zinc finger domains can bind to cognate DNA as a monomer. They can recognize non-palindromic sequences which enhances the specificity for the target in drug design.

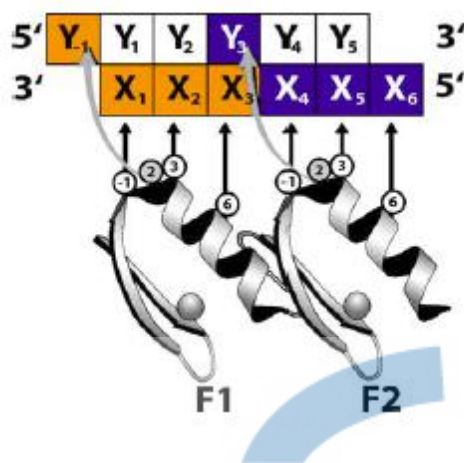


Fig.11 Model of DNA recognition by classical  $C_2H_2$  zinc finger proteins.[67] Each classical zinc finger consists of two  $\beta$ -strands (represented as large arrows) and an  $\alpha$ -helix internally stabilized by chelation of a single zinc ion. The nucleotides ( $X_1\sim X_6$ ) in the primary DNA strand contact amino acids at positions -1, 3 and 6 of  $\alpha$ -helix, while the complementary strand of DNA (positions  $Y_{-1}$  to  $Y_5$ ) is contacted by the amino acid at position 2 (gray circle). This cross-strand interaction is responsible for the overlap between the F1 (orange) and F2 (blue) binding sites.

### 3.2 Zinc fingers involved in protein-protein interaction

An increasing number of zinc fingers are found to be involved in protein-protein interactions. The growing number of folds associated with the zinc finger is followed by the characterization of biological regulatory mechanisms. The mechanisms of interactions are various including tandem array of multiple zinc fingers, cooperative effects between the repeats of single zinc finger domains, interaction directly with target proteins or serving as adaptor proteins to associate other functional motifs. As new functions have been explored, new folds of zinc fingers have been discovered. For example, PHD fingers which are found in nuclear proteins are involved in chromatin regulation.[68] RING fingers regulate E3 ligase activity.[69] LIM domains contain multiple zinc fingers and can serve as adaptors. Cytoskeletal cysteine rich protein (CRP) contains two LIM domains. It interacts with another LIM protein, zyxin and associates with the actin cytoskeleton.[70] FYVE domains are associated with high affinity

phosphoinositide binding, thus regulating lipid signaling and membrane trafficking.[71] The TAZ2 domain, a component of CBP/p300, is able to bind transcription factors and regulate transactivation.[72]

### 3.3.1. RING finger

The RING finger is a cysteine-rich amino acid motif which was found first in the human *ring1* gene (really interesting new gene 1).[14] The basic sequence of the RING is Cys-X<sub>2</sub>-Cys-X<sub>(9-39)</sub>-Cys-X<sub>(1-3)</sub>-His-X<sub>(2-3)</sub>-Cys-X<sub>2</sub>-Cys-X<sub>(4-48)</sub>-Cys-X<sub>2</sub>-Cys. The most significant feature of RING finger is that it binds two zinc ions in a cross-braced arrangement (Fig.12A). This binding topology is distinct from other zinc finger families such as PHD and LIM domains. The conserved spacing between the second and third pairs of CCCH zinc binding residues implies that the distance between two zinc binding sites is also conserved. Interestingly, the ligands that coordinate zinc such as cysteine or histidine can be replaced by another residue, such as Asp in Rbx/Roc1.[73] Early studies of RING fingers focused on the association of RING with other proteins. BRCA1 was shown to interact with protein BARD1 via the RING domains of each protein.[74] Clinical studies have shown that 20% of mutations in BRCA1 occur at the N-terminal region which harbors the RING motif and these mutations are related to breast cancer.[75] RING fingers can also associate with other domains and form larger interaction motifs such as the RING finger B-box  $\alpha$ -helical coiled-coil motif (RBCC).[76] The function of the B-box domains remains unclear.

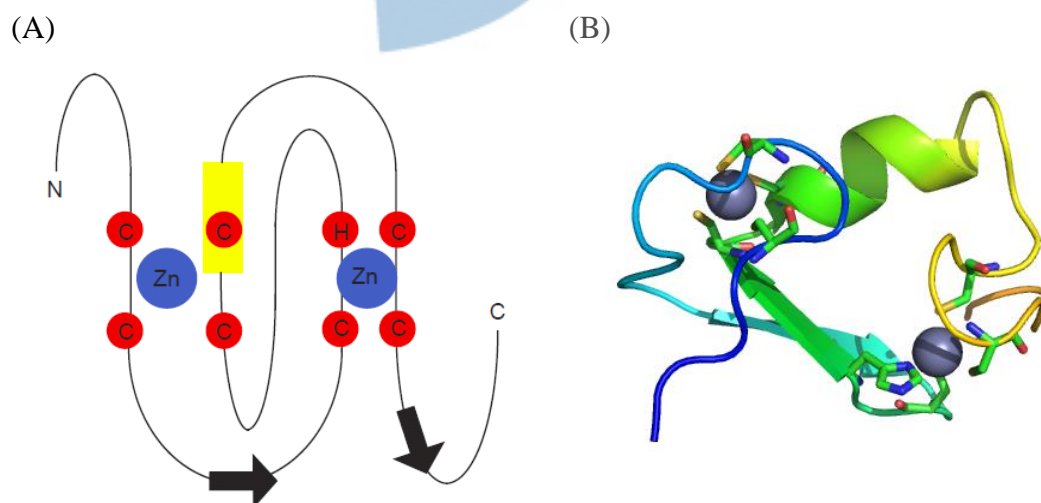


Fig.12 (A) Topology of the RING finger. (B) Structure of the RING finger in equine herpes virus protein (PDB code: 1chc).[15] It consists of three  $\beta$ -strands and an  $\alpha$ -helix with a cross-braced arrangement of zinc-binding residues.

RING domains have been shown to regulate E3 ligase activity. The function of RING in relation to ubiquitination was first identified in 1997 for Rad18, a member of the original cohort of RING proteins, that promotes the ubiquitination of histone.[77] Ubiquitination involves three enzymes known as E1 (ubiquitin-activation), E2 (ubiquitin-conjugation) and E3 (ubiquitin-ligase) enzymes. The first step in ubiquitination is the activation of ubiquitin (Ub) by E1. The activated ubiquitin is transferred from the active-site cysteine of E1 to the active-site cysteine of an E2 through the formation of a thioester linkage. The E2 enzyme then interacts with E3 which causes the transfer of Ub from E2-Ub to a lysine residue of a substrate. The transfer may occur in two different ways: either directly from E2 or via the E3 enzyme, mediated by a HECT domain. The ubiquitin-ligase activity of RING domains is conserved from yeast to human and the fold has been unambiguously associated with proteasome-mediated protein catabolism. Molecular insight into RING domain function came from the crystal structure of the RING domain of Cbl in complex with UbcH7 (Fig.13).[69] The RING domain plays a central role in c-Cbl function as its deletion or disruption abolishes the function of c-Cbl.[78] The crystal structure shows that the RING domain makes a rigid contact with UbcH7 (fig.13). However, upon UbcH7 binding to c-Cbl, no conformational change occurs. The closest distance between the UbcH7 active site cysteine and any RING domain residues is about 15Å which does not suggest the previous study that RING domains in E3s activate E2 enzyme allosterically.[79]

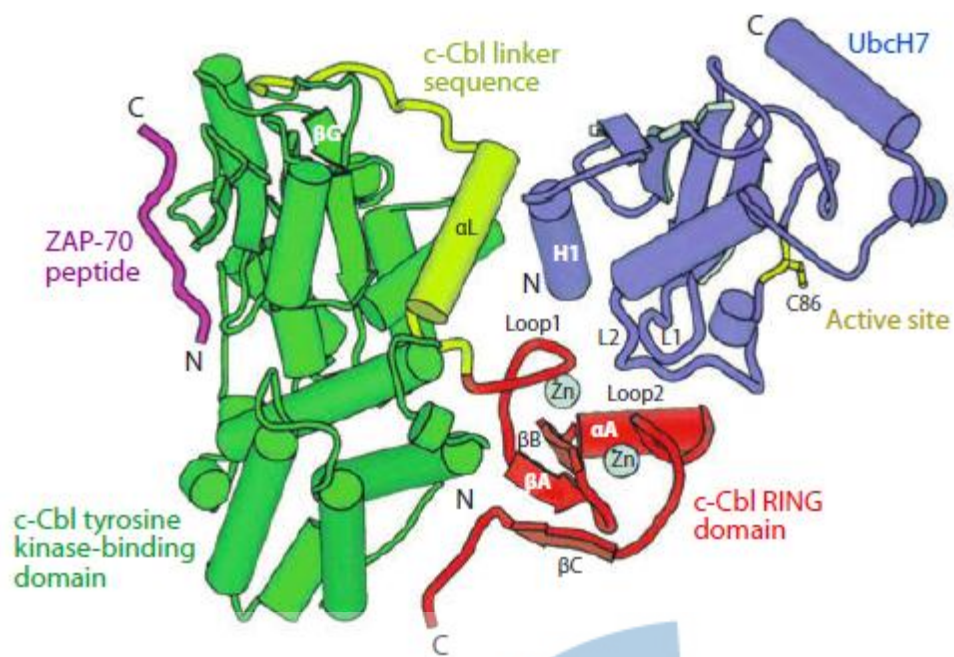


Fig.13 Structure of the c-Cbl- UbcH7 complex (PDB code: 1fbv).[69] The RING domain provides a shallow groove into which the tips of the L1 and L2 loops of UbcH7 bind. The groove is composed of the  $\alpha$ -helix and the two zinc-chelating loops of the RING domain. The contact involves F63 of UbcH7 L1 and P97 and A98 in UbcH7 L2.



Although the crystal structure of c-Cbl bound to UbcH7 provided some insight into the mechanism of substrate recognition mediated by the RING domain, it still left many unanswered questions. One possibility was that RING-E3s are activated by the successive addition of ubiquitin to a substrate. RING-E3s promote the direct transfer of ubiquitin from E2s to substrate (Fig.14) but the transfer requires a large conformational change to bridge the 50Å that separate the substrate docking site from the E2-bound Ub.[69, 73] A further possibility was that direct transfer occurs through dimerization of RING domains in E3s. The dimerization of RING domains has been observed in Brca1-Bard1 and Mdm2-MdmX.[74, 80] However, the former has only structural role and does not interact with E2. In contrast, the latter is able to recruit E2 through the dimerization of RING domains.

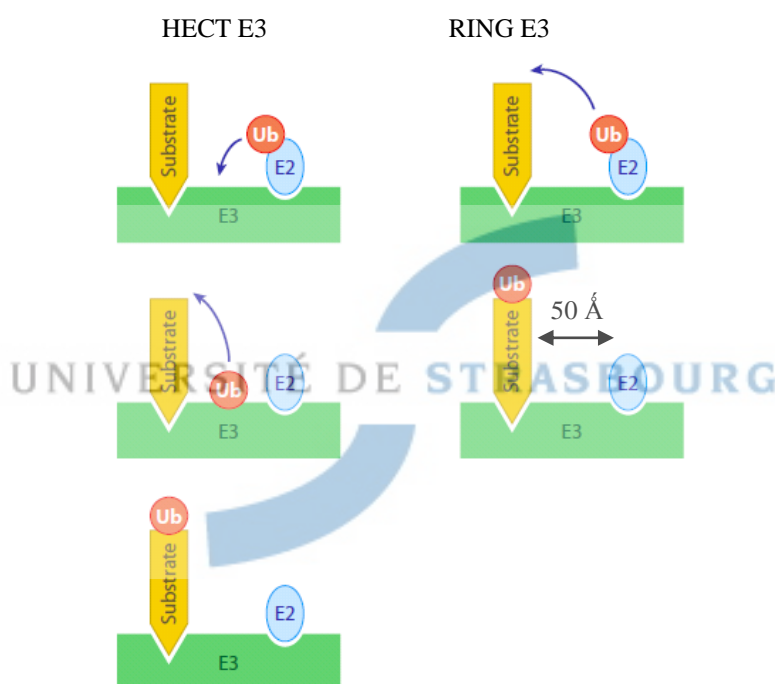


Fig.14 RING-E3 promotes the direct transfer of ubiquitin from E2-Ub to substrate. HECT E3s have a conserved cysteine residue which binds to ubiquitin through a thioester bond. Ubiquitin is then transferred covalently from E3 intermediate to the substrate. In contrast, RING-E3s can directly promote ubiquitin transfer from E2 to substrate. However, the substrate and ubiquitin thioester bond may be separated by a considerable distance.[81]

### Roles of RING domains in post-translational modification

In addition to ubiquitination, RING-E3s can also be involved in post-translational modifications such as sumoylation or phosphorylation. For example, Rfp1 and Rfp2 proteins which were identified in yeast and are essential for cell growth and genome stability. They contain an N-terminal SUMO-interacting motif and a C-terminal RING domain which can recruit Slx8, an active RING-finger ubiquitin ligase, through a

RING–RING interaction, to form a functional E3 complex.[82] Although Rfp1 and Rfp2 lack E3 activity, they are able to ubiquitinate SUMO-containing substrates *in vitro*. The mammalian RING finger protein RNF4, an orthologue of Rfp1 and Rfp2, targets poly-SUMO-modified proteins for degradation mediated by ubiquitin.[83] However, these results do not indicate a direct role of the RING domain in sumoylation. On the other hand, TNF receptor-associated factor 2 (TRAF2) is an adaptor protein which regulates the activation of c-Jun NH(2) terminal kinase (JNK)/c-Jun and IκB kinase (IκK)/nuclear factor-κB (NF-κB) signaling cascades in response to TNFα stimulation. Phosphorylation of the RING domain in TRAF2 was observed to inhibit stress-induced cell death and suppress JNK activation.[84] This finding suggests that phosphorylation of TRAF2 is related to the regulation of NF-κB activity in some cancers. Although several models of phosphorylation and sumoylation regulated by RING-E3s have been proposed,[81] the mechanism remains unclear. The elucidation of modifications of RING domains in E3 ligase proteins will be a considerable challenge for the future.

### 3.3.2 PHD fingers

The PHD finger was identified first in plants as a sequence with spaced cysteines that was conserved between two plant homeodomain proteins.[85] The PHD finger was classified in the same group as RING fingers based on the similarities in sequence and metal-binding residues. PHD fingers were later found in a wide variety of eukaryotic proteins. The most remarkable feature of the PHD finger is an additional conserved aromatic residue near the C-terminus, most often a tryptophan. The first structure of a PHD finger was solved in 2000 and showed that the two zinc binding motifs adopt the cross-braced topology (Fig.15).[86] The structure also revealed some differences between RING and PHD fingers, especially in the surface area of structures. The RING domain surface contains a conserved α-helix in order to dock E2 ligases whereas the equivalent surface is absent in the PHD finger.[86, 87] This significant difference indicates that PHD and RING fingers are two distinct zinc finger families.



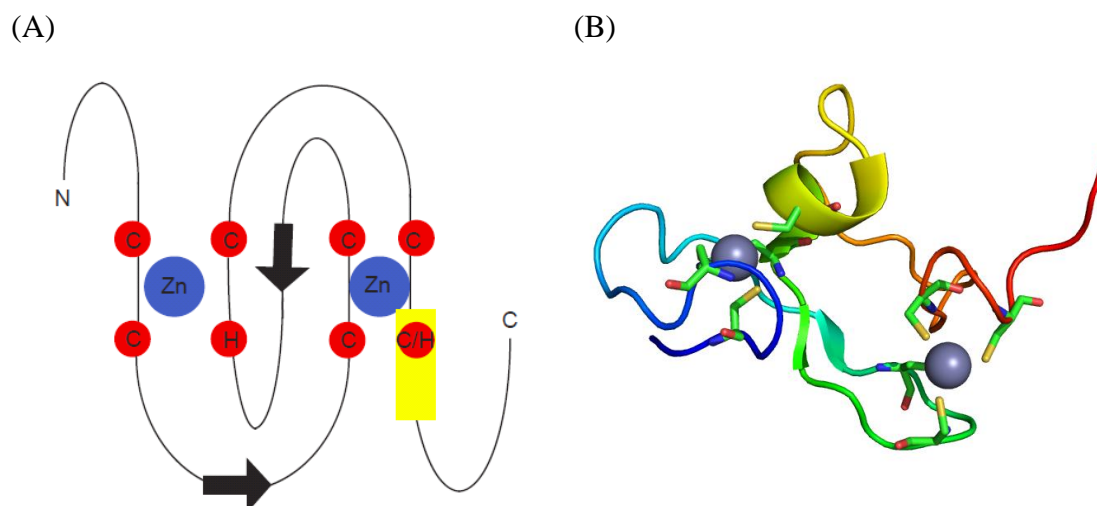


Fig.15 Topology of the PHD finger. It contains  $C_4HC_3$  zinc binding motif with zinc binding sites adopting the cross-braced topology. (B) Structure of PHD finger from human Williams-Beuren syndrome transcription factor (PDB code: 1f62).[86] The structure consists of three loops stabilized by two zinc binding clusters and a small double-stranded antiparallel  $\beta$ -sheet.

Because of the similarity in zinc binding topology between PHD and RING fingers, the PHD finger was originally thought to be involved in ubiquitination. The zinc fingers of MEKK1 and the MIR proteins were shown to act as E3 ubiquitin ligases.[88, 89] The latter reports showed in fact that the zinc fingers of these proteins had been misclassified as PHD fingers.[90, 91] However, the function of PHD fingers remained unclear. An initial hypothesis was that PHD fingers might be involved in gene expression since they are nuclear proteins.[68] A nucleosome-binding activity was observed first for the PHD finger of p300, a transcription coactivator with histone acetyltransferase activity.[92] It cooperates with the adjacent bromodomain to associate with native hyperacetylated nucleosomes. The substitution of the PHD domain of p300 by heterologous canonical PHD fingers causes loss of binding activity.[92] However, the mechanism of interaction between PHD fingers and nucleosomes was not clear. Recent studies have shown that PHD fingers can bind to histone peptides and act as reader modules in chromosomes. Trimethylated lysine 4 on histone 3 (H3K4) was identified as one binding target for the BPTF-PHD finger.[24] Further PHD fingers have since been found to regulate gene expression through interactions with specifically modified H3 tails. Three types of reader modules for histone modification recognition by PHD fingers have been identified including H3K4me3, H3K4me0 and nonhistone binding (Fig.16). The PHD fingers within their complexes interact with the histone H3 tails and play different roles in gene regulation.

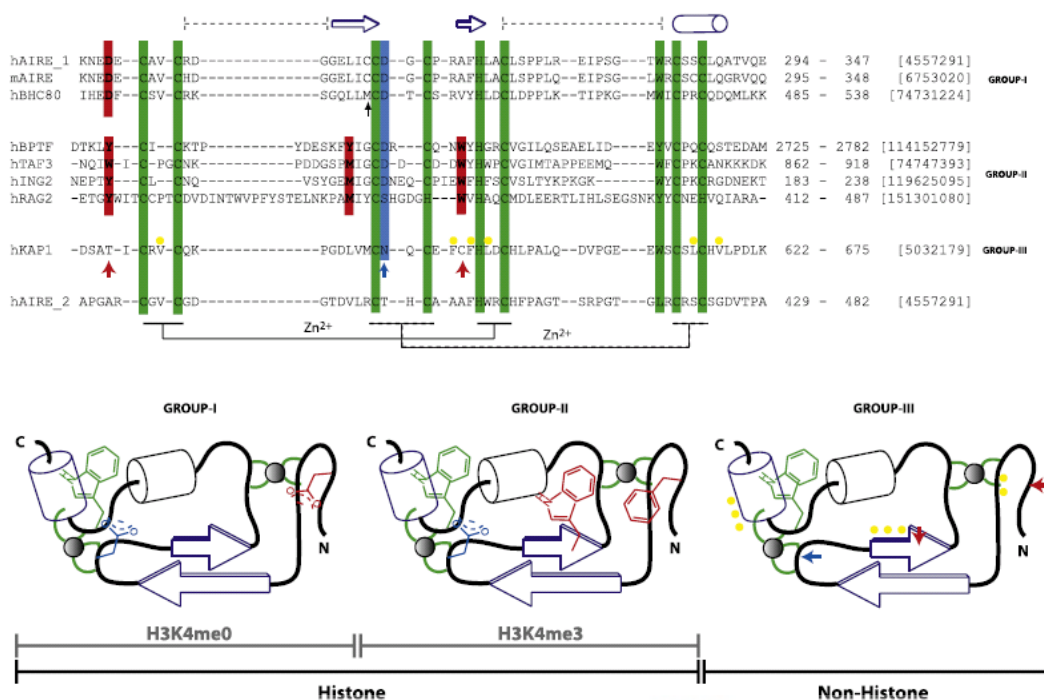


Fig.16 Classification of the PHD finger family. The upper figure shows three groups characterized by distinct sequences and ligand binding specificities. The recognition of H3K4me3 is characterized by aromatic cage residues whereas that of K4me0 involves the formation of an ion pair with a distinct N-terminal Asp residue. The characteristic histone-peptide interacting positions are shown in red, the Zn-chelating residues are connected by regular and dotted lines for the first and second tetrads and conserved C-terminal aromatic residues, characteristic of the entire PHD family are shown in green. The bottom figure indicates the topology of PHD fingers and their interactions with either histone peptides or non-histones.[93]

The PHD fingers in the tumor suppressor ING2 (inhibitor of growth 2) (Fig.17) [94], TAF3 (TATA box-binding protein-associated factor) [95] and RAG2 [96] recognize specifically the N-terminal tail of histone H3 in which Lys4 is trimethylated (H3K4me3). The aromatic residues in these PHD fingers are important in the formation of a H3K4me3 binding pocket and mutation of these residues disrupts the histone binding affinity.[95] However, the binding affinities in this PHD finger family vary widely: the  $K_D$  of histone peptide binding toTAF3-PHD is nearly 1000-fold less than that of ING2-PHD despite the similarities in three dimensional structures of the two domains.[95]

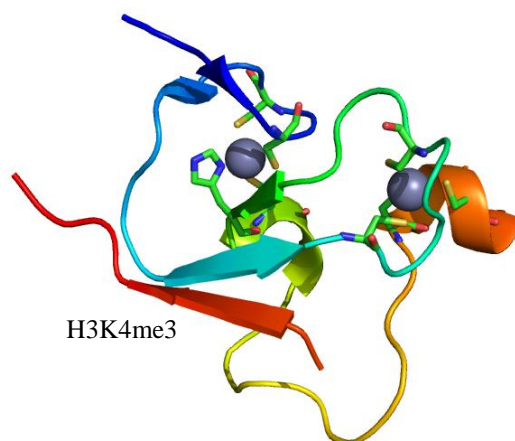


Fig.17 Structure of ING2-PHD complex with histone peptide H3K4me3. The overall fold of ING-PHD in the complex is similar that found in ligand-free PHD (PDB code: 2g6q).[94] The histone H3K4me3 adopts an extended  $\beta$ -strand like conformation and contacts one side of the  $\beta$ -sheet of ING2-PHD through hydrogen bonds.

The second group involves PHD fingers that recognize non-methylated lysine 4 in histone 3 (H3K4me0). Human autoimmune regulator (AIRE) belongs to this group. It consists of two PHD fingers and several domains found in transcription regulators such as the SAND domain. Unlike PHD recognition of H3K4me3 by an aromatic cage, AIRE-PHD1 binds to H3K4me0 through the amino-terminal aspartic acid interacting with Arg2 guanidinium group in H3K4me0. Methylation of Arg2 in H3K4me0 impairs this interaction, and mutations in AIRE-PHD cause autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED).[93, 97]

In addition to interacting with histone tails, PHD fingers can also interact with other proteins. The tandem PHD-finger-bromodomain (PHD-BRD) in human co-repressor KRAB-associated protein (KAP1) regulates gene silencing.[98] Structure analysis has shown that the PHD domain alone or with bromodomain is unable to bind to lysine-methylated peptides in histones such as H3K4, H3K9, and H3K27.[99] The PHD finger cooperates with the bromodomain to regulate lysine SUMOylation. The tandem array of PHD finger-bromodomain in KAP1 facilitates the interaction with E2 conjugating enzyme UBC9.[99]

### 3.3.3 LIM domain

The name LIM came from three proteins containing a conserved domain: LIN11, Isl-1 and MEC3. LIM proteins are involved in a range of biological functions including gene expression, cytoskeleton organization and development.[100] LIM domains contain two sequential zinc binding modules and eight conserved zinc-binding ligands,

usually cysteine or histidine residues (Fig.18), although these can be substituted with aspartate and glutamate residues in some cases. The LIM domains of the protein LIM only domain 4(LMO4) have CCHC and CCCD zinc binding motifs.[100]

LIM domains can be located either near the N- or C- terminus of proteins. Because of the tandem array of zinc fingers, LIM domains can act as multifunctional modules which link to other domains such as homeodomain or catalytic domains. They are also found linked to protein-binding modules such as Src-homology-3 (SH3) or PDZ domains.[101, 102] In cells, LIM proteins are observed in both the nucleus and the cytoplasm. Many LIM domains are identified as cytoskeleton-associated proteins such as zyxin, four-and-a-half LIM (FHL) and cysteine-rich protein (CRP) families. Although LIM domains are found as multiple domains in a protein, each domain may exhibit different dynamic behavior. The cysteine- and glycine-rich protein family (CRP) contains LIM1 (amino-terminal) and LIM2 (carboxyl-terminal) for which dynamic analysis showed independent motional behavior. It suggests that LIM domains in proteins may function as adaptors and interact with two or more other protein to form a macromolecular complex.[103]

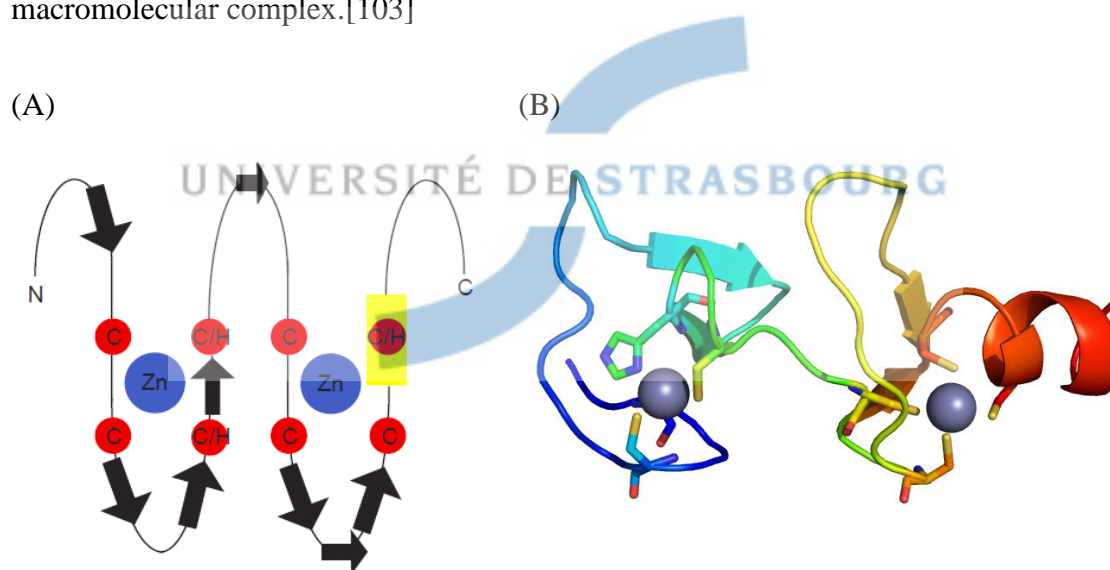


Fig.18 (A) Structural topology of LIM domain. The two zinc binding sites adopt a sequential-braced arrangement. (B) Structure of the N-terminal LIM domain from cysteine- and glycine-rich protein CRP2 (LIM1) (PDB code: 1a7i).[103] The LIM1 domain consists of two zinc binding sites of the CCHC and the CCCD type, respectively. Both zinc binding modules contain two orthogonally arranged antiparallel  $\beta$ -sheets and are packed together with a hydrophobic core composed of residues from the zinc finger loop regions. The CCCD zinc finger is followed by a short  $\alpha$ -helical stretch.

### 3.3.4 MYND domains

The MYND domain is defined by a  $C_4C_2HC$  zinc-chelating motif that is found in a number of nuclear proteins such as Nery, DEAF-1, BS69, PDCD2, and AML/ETO. It

is implicated in transcriptional repression. DEAF-1, for example, regulates the transcription by DNA recognition through its SAND domain. The MYND domain is composed of two zinc binding modules which coordinate with CCCC and CCHC motifs, respectively. The topology of zinc coordination was originally thought to be sequential-braced but this was subsequently corrected and found to be similar to PHD and RING fingers.[104] Functionally, MYND domain is involved in protein-protein interactions. The MYND domain of AML1/ETO recognizes a “PPPLI” motif in the silencing mediator of retinoid and thyroid hormone receptor (SMRT) and nuclear receptor corepressor (N-CoR).

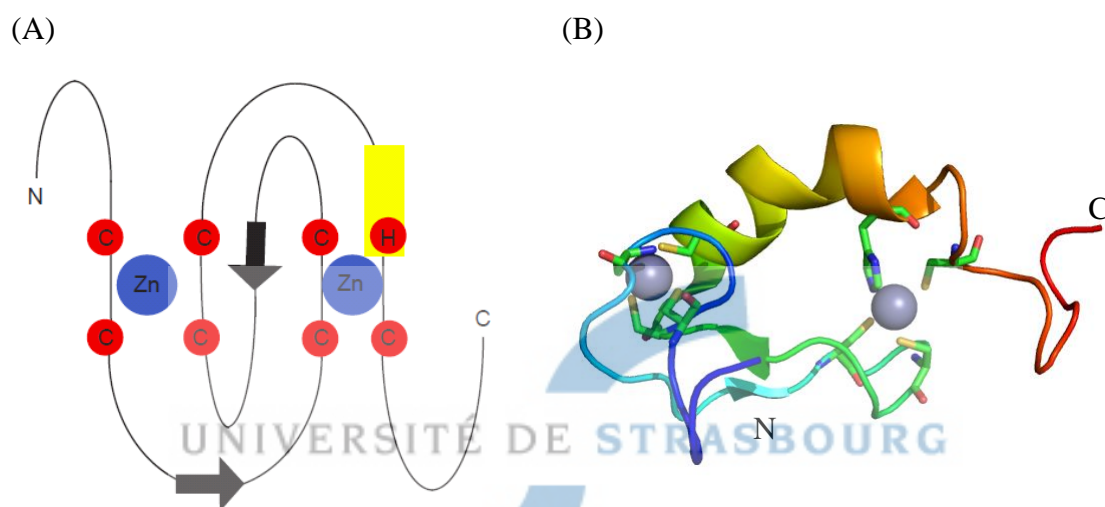


Fig.19 Structural topology of the MYND domain. The zinc binding sites adopt the cross-braced arrangement. (B) Structure of AML1/ETO MYND domain (PDB code: 2od1).[104] It is structurally homologous to the RING and PHD finger in proteins.

### 3.3.5 TAZ and ZZ domains

Transcriptional activation in eukaryotes involves many components such as activators which are bound to DNA and regulate the transcription and coactivators which can mediate interaction between activators and target DNA. Of these coactivators, CREB-binding protein (CBP) with its paralog p300 is a large multidomain protein which can mediate interaction between transcription factors and specific DNA enhancers.[72] CBP contains three regions that bind to zinc ions which were named CH1, CH2 and CH3 (Fig.20A) because of their content of cysteine and histidine residues. CH1 domain was also named the TAZ1 domain. CH2 region contains a PHD motif and CH3 region includes two separate zinc binding motifs, termed ZZ and TAZ2.[105]

Through these TAZ1 and TAZ2 domains, CBP can interact with a number of transcription factors. The TAZ1 and TAZ2 domains are homologous and each contains

three HCCC zinc-binding motifs and four  $\alpha$ -helices (Fig.20B). The TAZ1 domain adopts a well-defined conformation although early studies suggested it to be a molten globule, even in the presence of  $Zn^{2+}$ . [106, 107] The structure of the TAZ2 domain is similar to that of the TAZ1 domain but differs in the orientation of the fourth  $\alpha$ -helix with respect to  $\alpha_1$  and  $\alpha_4$  (Fig.20C and 20D). The structural differences between TAZ1 and TAZ2 domains may result in the recognition of different transcription factors: the TAZ1 domain binds to the activation domains of hypoxia inducible factor (HIF-1 $\alpha$ ) and CBP/p300-interacting transactivator with ED-rich tail (CITED2), while the TAZ2 domain can interact with p53. [108]





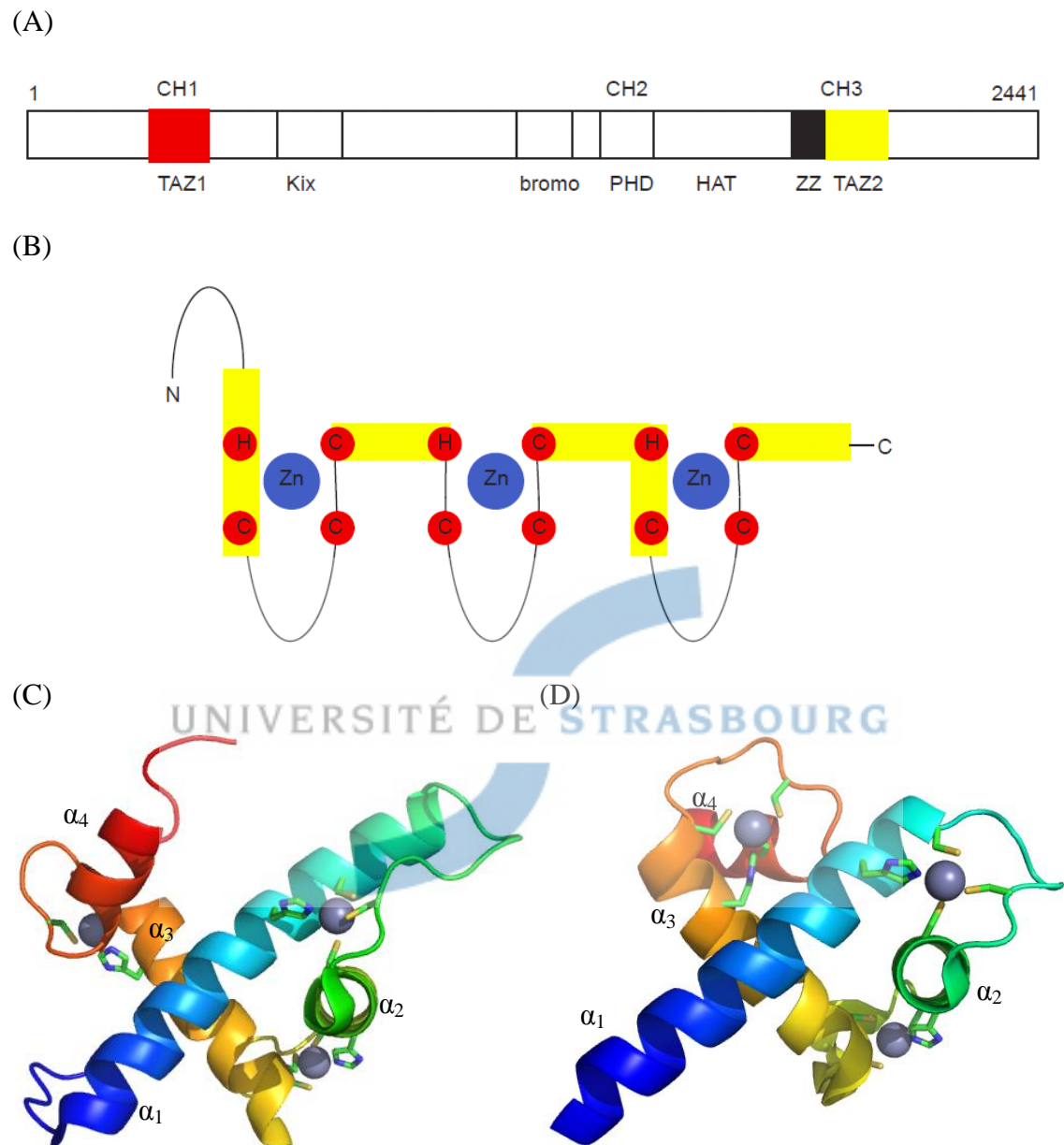


Fig.20 (A) The domain structure of CBP, consisting of CH1, CH2 and CH3 domains. The CH3 domain is composed of the ZZ and TAZ2 domains. Kix, bromo domains and PHD are also components of CBP. (B) The topology the TAZ domain. It consists of three zinc binding sites, each defined by a HCCC motif. The protein contains four  $\alpha$ -helices linked by three short loops. Each loop contains high proportion of non-polar residues. The four helices pack against each other to form the hydrophobic core. Structures of the TAZ1 (C) and TAZ2 (D) domain of CBP (PDB code: 1u2n and 1f81).[19] These domains exhibit a similar fold except in the position of the three zinc binding sites. In the TAZ 1 domain,  $\alpha_1$  and  $\alpha_4$  face the same side of  $\alpha_3$  while in TAZ2 these helices are on opposite sides of  $\alpha_3$ .



Another zinc binding domain in CBP, the ZZ domain, lies close to the N-terminus of the TAZ2 domain (Fig.20B). Given its proximity to the TAZ2 domain, ZZ domain was thought to also interact with p53. However, binding assays showed that the interaction involves only the TAZ2 domain.[108] The ZZ domain adopts the cross-braced zinc-binding topology which was also found in PHD and RING fingers (Fig.21A).[108] In particular, the second zinc binding site contains an unusual HxH motif. Compared with the high degree of conservation in the first zinc binding site, the second site shows considerable variability in the positioning of the two histidine residues, suggesting that molecular recognition involving the ZZ domain may be defined in this region. However, the precise function of ZZ domain remains unclear.

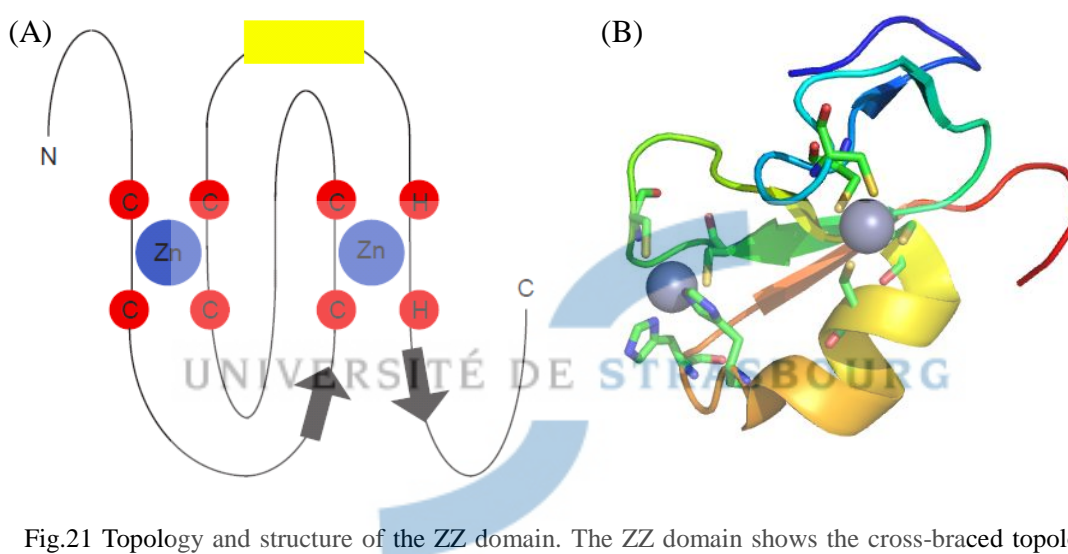


Fig.21 Topology and structure of the ZZ domain. The ZZ domain shows the cross-braced topology. Two zinc binding motifs coordinate with CCCC and CCHH motif, respectively. In particular, the second zinc binding site contains an unusual HxH motif.

### 3.3 Zinc finger proteins involved in membrane trafficking: FYVE domain

The FYVE domain was named after the set of four proteins in which it was first identified: Fab1, YOTB/ZK632.12, Vac1 and EEA1.[71] It was identified in 1996 as a zinc finger and exists in wide range of eukaryotic proteins, involved in the regulation of signal transduction, membrane trafficking, the cytoskeleton and apoptosis.[18] The FYVE domain is a small zinc binding module that recognizes phosphoinositide (PtdIns3P) with high specificity and affinity.[109] Structures of FYVE domains revealed a fold that contains two double-stranded antiparallel  $\beta$ -sheets and a C-terminal  $\alpha$ -helix.[110, 111] FYVE domain consists of two zinc binding modules which are bound by four CXXC motifs in a cross-braced topology. In particular, the unfolded structure that results from the removal of zinc ion can be restored to the folded protein upon reintroduction of zinc ions.[112]

Distinguishing it from other zinc fingers that bind nucleic acids or proteins, the FYVE domain recognizes PtdIns3P specifically. The mechanism of binding to PtdIns3P is in multiple steps involving non-specific insertion of a hydrophobic loop into the lipid bilayer, positioning and activation of the binding pocket. For example, in the complex of EEA1 bound to PtdIns(3)P, three conserved regions in FYVE domains including the N-terminal WxxD, the central basic RR/KHHCR and the C-terminal RVC motifs comprise together a binding pocket.[113] Ligation of PtdIns3P causes a global structural change, drawing the protein termini over the bound PtdIns3P by extension of a hinge.[113] In addition, functions of FYVE domain in the cells may be affected by pH. The FYVE domain of EEA1 exists in the bound form at pH 6.0-6.6 physiologically.[71]

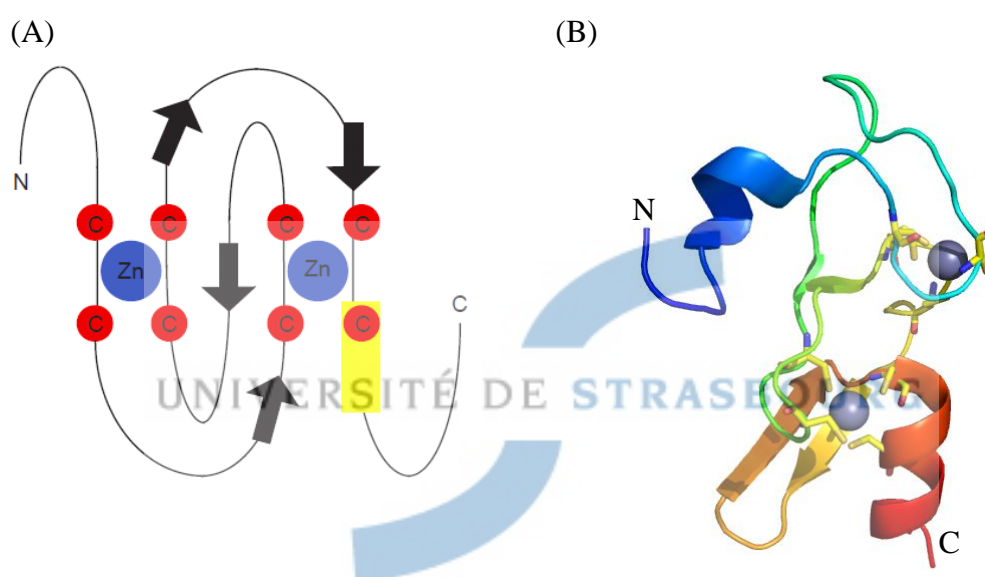


Fig.22 (A) The structural topology of the FYVE domain. The FYVE domain fold is stabilized by two zinc ions which adopt the cross-braced topology. The zinc coordination pattern is similar to that of the RING finger. (B) The structure of the FYVE domain of early endosome antigen1 protein (EEA1) (PDB code: 1hyj).[113] The first zinc binding module consists of two short anti-parallel  $\beta$ -strands which play an important role in PtdIns(3)P binding. Two zinc binding motifs lie beneath and position the PtdIns(3)P binding site.

#### 4. Conclusion

A large number of zinc fingers have now been characterized either structurally or functionally or both. No such extensive family of other metal-binding proteins, such as Cd or Ni containing proteins, exists in nature. Although no strong physiological evidence can support the phenomena, experiments of metal-substitution in zinc fingers may help to explain this puzzle.

In the periodic table, cadmium and zinc belong to the same group and therefore have similar chemical properties. The zinc ion is shown to favor tetrahedral coordination.

However, substitution of zinc by cadmium reflects the higher affinity of cadmium for sulphur linkage ligand and also alters the metal binding conformation.[49] In addition, replacement of zinc by other metals may be related to the causes of physiological disorders. Cadmium exerts complex effects on the growth and survival of normal and cancer cells. It not only disrupts zinc coordination but also causes cellular toxicity and may result in tumor growth.[114] These observations may explain the importance of the zinc ion in zinc finger proteins in eukaryotes which not only maintains the fold but also reflect adaptation of the cellular environment.[115] The diversity of functions also provides more possibilities to develop new drugs and provides ideas for the exploration of unknown scientific fields in the universe.

## References

1. Auld DS: The ins and outs of biological zinc sites. *Biometals* 2009, 22(1):141-148.
2. Vallee BL, Auld DS: Active zinc binding sites of zinc metalloenzymes. *Matrix Suppl* 1992, 1:5-19.
3. Vallee BL, Auld DS: Zinc coordination, function, and structure of zinc enzymes and other proteins. *Biochemistry* 1990, 29(24):5647-5659.
4. Lee YM, Lim C: Physical basis of structural and catalytic Zn-binding sites in proteins. *J Mol Biol* 2008, 379(3):545-553.
5. Miller J, McLachlan AD, Klug A: Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus oocytes*. *EMBO J* 1985, 4(6):1609-1614.
6. Karpel RL, Henderson LE, Oroszlan S: Interactions of retroviral structural proteins with single-stranded nucleic acids. *J Biol Chem* 1987, 262(11):4961-4967.
7. Copeland TD, Morgan MA, Oroszlan S: Complete amino acid sequence of the basic nucleic acid binding protein of feline leukemia virus. *Virology* 1984, 133(1):137-145.
8. Berg JM: Potential metal-binding domains in nucleic acid binding proteins. *Science* 1986, 232(4749):485-487.
9. Parraga G, Horvath SJ, Eisen A, Taylor WE, Hood L, Young ET, Klevit RE: Zinc-dependent structure of a single-finger domain of yeast ADR1. *Science* 1988, 241(4872):1489-1492.
10. Lee MS, Gippert GP, Soman KV, Case DA, Wright PE: Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science* 1989, 245(4918):635-637.
11. Summers MF, South TL, Kim B, Hare DR: High-resolution structure of an HIV zinc fingerlike domain via a new NMR-based distance geometry approach. *Biochemistry* 1990, 29(2):329-340.

12. Hard T, Kellenbach E, Boelens R, Maler BA, Dahlman K, Freedman LP, Carlstedt-Duke J, Yamamoto KR, Gustafsson JA, Kaptein R: Solution structure of the glucocorticoid receptor DNA-binding domain. *Science* 1990, 249(4965):157-160.
13. Pavletich NP, Pabo CO: Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 1991, 252(5007):809-817.
14. Freemont PS, Hanson IM, Trowsdale J: A novel cysteine-rich sequence motif. *Cell* 1991, 64(3):483-484.
15. Barlow PN, Luisi B, Milner A, Elliott M, Everett R: Structure of the C3HC4 domain by 1H-nuclear magnetic resonance spectroscopy. A new structural class of zinc-finger. *J Mol Biol* 1994, 237(2):201-211.
16. Borden KL, Boddy MN, Lally J, O'Reilly NJ, Martin S, Howe K, Solomon E, Freemont PS: The solution structure of the RING finger domain from the acute promyelocytic leukaemia proto-oncoprotein PML. *EMBO J* 1995, 14(7):1532-1541.
17. Andreini C, Banci L, Bertini I, Rosato A: Counting the zinc-proteins encoded in the human genome. *J Proteome Res* 2006, 5(1):196-201.
18. Stenmark H, Aasland R, Toh BH, D'Arrigo A: Endosomal localization of the autoantigen EEA1 is mediated by a zinc-binding FYVE finger. *J Biol Chem* 1996, 271(39):24048-24054.
19. De Guzman RN, Liu HY, Martinez-Yamout M, Dyson HJ, Wright PE: Solution structure of the TAZ2 (CH3) domain of the transcriptional adaptor protein CBP. *J Mol Biol* 2000, 303(2):243-253.
20. Michelsen JW, Schmeichel KL, Beckerle MC, Winge DR: The LIM motif defines a specific zinc-binding protein domain. *Proc Natl Acad Sci U S A* 1993, 90(10):4404-4408.
21. Schmeichel KL, Beckerle MC: The LIM domain is a modular protein-binding interface. *Cell* 1994, 79(2):211-219.
22. Everett RD, Barlow P, Milner A, Luisi B, Orr A, Hope G, Lyon D: A novel arrangement of zinc-binding residues and secondary structure in the C3HC4 motif of an alpha herpes virus protein family. *J Mol Biol* 1993, 234(4):1038-1047.
23. Perez-Alvarado GC, Miles C, Michelsen JW, Louis HA, Winge DR, Beckerle MC, Summers MF: Structure of the carboxy-terminal LIM domain from the cysteine rich protein CRP. *Nat Struct Biol* 1994, 1(6):388-398.
24. Li H, Ilin S, Wang W, Duncan EM, Wysocka J, Allis CD, Patel DJ: Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* 2006, 442(7098):91-95.
25. Yanagisawa S: Dof domain proteins: Plant-specific transcription factors associated with diverse phenomena unique to plants. *Plant and Cell Physiology* 2004, 45(4):386-391.

26. Fukata Y, Iwanaga T, Fukata M: Systematic screening for palmitoyl transferase activity of the DHHC protein family in mammalian cells. *Methods* 2006, 40(2):177-182.
27. Anzellotti AI, Farrell NP: Zinc metalloproteins as medicinal targets. *Chem Soc Rev* 2008, 37(8):1629-1651.
28. Tadepally HD, Burger G, Aubry M: Evolution of C2H2-zinc finger genes and subfamilies in mammals: species-specific duplication and loss of clusters, genes and effector domains. *BMC Evol Biol* 2008, 8:176.
29. Malgieri G, Russo L, Esposito S, Baglivo I, Zaccaro L, Pedone EM, Di Blasio B, Isernia C, Pedone PV, Fattorusso R: The prokaryotic Cys2His2 zinc-finger adopts a novel fold as revealed by the NMR structure of *Agrobacterium tumefaciens* Ros DNA-binding domain. *Proc Natl Acad Sci U S A* 2007, 104(44):17341-17346.
30. Bouhouche N, Syvanen M, Kado CI: The origin of prokaryotic C2H2 zinc finger regulators. *Trends Microbiol* 2000, 8(2):77-81.
31. Gray MW, Burger G, Lang BF: Mitochondrial evolution. *Science* 1999, 283(5407):1476-1481.
32. Urrutia R: KRAB-containing zinc-finger repressor proteins. *Genome Biol* 2003, 4(10):231.
33. Dehal P, Predki P, Olsen AS, Kobayashi A, Folta P, Lucas S, Land M, Terry A, Ecale Zhou CL, Rash S *et al*: Human chromosome 19 and related regions in mouse: conservative and lineage-specific evolution. *Science* 2001, 293(5527):104-111.
34. Emerson RO, Thomas JH: Adaptive evolution in zinc finger transcription factors. *PLoS Genet* 2009, 5(1):e1000325.
35. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W *et al*: Initial sequencing and analysis of the human genome. *Nature* 2001, 409(6822):860-921.
36. Bellefroid EJ, Poncelet DA, Lecocq PJ, Revelant O, Martial JA: The evolutionarily conserved Kruppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. *Proc Natl Acad Sci U S A* 1991, 88(9):3608-3612.
37. Sander TL, Stringer KF, Maki JL, Szauter P, Stone JR, Collins T: The SCAN domain defines a large family of zinc finger transcription factors. *Gene* 2003, 310:29-38.
38. Chung HR, Schafer U, Jackle H, Bohm S: Genomic expansion and clustering of ZAD-containing C2H2 zinc-finger genes in *Drosophila*. *EMBO Rep* 2002, 3(12):1158-1162.
39. Chung HR, Lohr U, Jackle H: Lineage-specific expansion of the zinc finger associated domain ZAD. *Mol Biol Evol* 2007, 24(9):1934-1943.



40. Collins T, Stone JR, Williams AJ: All in the family: the BTB/POZ, KRAB, and SCAN domains. *Mol Cell Biol* 2001, 21(11):3609-3615.
41. Wagner S, Hess MA, Ormonde-Hanson P, Malandro J, Hu H, Chen M, Kehrer R, Frodsham M, Schumacher C, Beluch M *et al*: A broad role for the zinc finger protein ZNF202 in human lipid metabolism. *J Biol Chem* 2000, 275(21):15685-15690.
42. Matthews JM, Sunde M: Zinc fingers--folds for many occasions. *IUBMB Life* 2002, 54(6):351-355.
43. Cook WJ, Mosley SP, Audino DC, Mullaney DL, Rovelli A, Stewart G, Denis CL: Mutations in the Zinc-Finger Region of the Yeast Regulatory Protein Adr1 Affect Both DNA-Binding and Transcriptional Activation. *Journal of Biological Chemistry* 1994, 269(12):9374-9379.
44. Elrod-Erickson M, Rould MA, Nekludova L, Pabo CO: Zif268 protein-DNA complex refined at 1.6 Å: a model system for understanding zinc finger-DNA interactions. *Structure* 1996, 4(10):1171-1180.
45. Wolfe SA, Greisman HA, Ramm EI, Pabo CO: Analysis of zinc fingers optimized via phage display: Evaluating the utility of a recognition code. *Journal of Molecular Biology* 1999, 285(5):1917-1934.
46. Pabo CO, Nekludova L: Geometric analysis and comparison of protein-DNA interfaces: why is there no simple code for recognition? *J Mol Biol* 2000, 301(3):597-624.
47. Rebar EJ, Pabo CO: Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* 1994, 263(5147):671-673.
48. Birnbaumer M, Schrader WT, O'Malley BW: Assessment of structural similarities in chick oviduct progesterone receptor subunits by partial proteolysis of photoaffinity-labeled proteins. *J Biol Chem* 1983, 258(12):7331-7337.
49. Freedman LP, Luisi BF, Korszun ZR, Basavappa R, Sigler PB, Yamamoto KR: The function and structure of the metal coordination sites within the glucocorticoid receptor DNA binding domain. *Nature* 1988, 334(6182):543-546.
50. Claessens F, Verrijdt G, Schoenmakers E, Haelens A, Peeters B, Verhoeven G, Rombauts W: Selective DNA binding by the androgen receptor as a mechanism for hormone-specific gene regulation. *J Steroid Biochem Mol Biol* 2001, 76(1-5):23-30.
51. Luisi BF, Xu WX, Otwinowski Z, Freedman LP, Yamamoto KR, Sigler PB: Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* 1991, 352(6335):497-505.
52. Shaffer PL, Gewirth DT: Structural basis of VDR-DNA interactions on direct repeat response elements. *EMBO J* 2002, 21(9):2242-2252.
53. Grishin NV: Treble clef finger--a functionally diverse zinc-binding structural motif. *Nucleic Acids Res* 2001, 29(8):1703-1714.

54. Baumann H, Paulsen K, Kovacs H, Berglund H, Wright AP, Gustafsson JA, Hard T: Refined solution structure of the glucocorticoid receptor DNA-binding domain. *Biochemistry* 1993, 32(49):13463-13471.
55. Johnston M: A model fungal gene regulatory mechanism: the GAL genes of *Saccharomyces cerevisiae*. *Microbiol Rev* 1987, 51(4):458-476.
56. Ma J, Ptashne M: A new class of yeast transcriptional activators. *Cell* 1987, 51(1):113-119.
57. Pan T, Coleman JE: GAL4 transcription factor is not a "zinc finger" but forms a Zn(II)<sub>2</sub>Cys<sub>6</sub> binuclear cluster. *Proc Natl Acad Sci U S A* 1990, 87(6):2077-2081.
58. Marmorstein R, Carey M, Ptashne M, Harrison SC: DNA recognition by GAL4: structure of a protein-DNA complex. *Nature* 1992, 356(6368):408-414.
59. Hong M, Fitzgerald MX, Harper S, Luo C, Speicher DW, Marmorstein R: Structural basis for dimerization in DNA recognition by Gal4. *Structure* 2008, 16(7):1019-1026.
60. Veronese FD, Rahman R, Copeland TD, Oroszlan S, Gallo RC, Sarngadharan MG: Immunological and chemical analysis of P6, the carboxyl-terminal fragment of HIV P15. *AIDS Res Hum Retroviruses* 1987, 3(3):253-264.
61. Darlix JL, Gabus C, Nugeyre MT, Clavel F, Barre-Sinoussi F: Cis elements and trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. *J Mol Biol* 1990, 216(3):689-699.
62. Lever A, Gottlinger H, Haseltine W, Sodroski J: Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. *J Virol* 1989, 63(9):4085-4087.
63. Morellet N, Jullian N, De Rocquigny H, Maigret B, Darlix JL, Roques BP: Determination of the structure of the nucleocapsid protein NCp7 from the human immunodeficiency virus type 1 by 1H NMR. *EMBO J* 1992, 11(8):3059-3065.
64. Morellet N, de Rocquigny H, Mely Y, Jullian N, Demene H, Ottmann M, Gerard D, Darlix JL, Fournie-Zaluski MC, Roques BP: Conformational behaviour of the active and inactive forms of the nucleocapsid NCp7 of HIV-1 studied by 1H NMR. *J Mol Biol* 1994, 235(1):287-301.
65. De Rocquigny H, Gabus C, Vincent A, Fournie-Zaluski MC, Roques B, Darlix JL: Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers. *Proc Natl Acad Sci U S A* 1992, 89(14):6472-6476.
66. Ramboarina S, Srividya N, Atkinson RA, Morellet N, Roques BP, Lefevre JF, Mely Y, Kieffer B: Effects of temperature on the dynamic behaviour of the HIV-1 nucleocapsid NCp7 and its DNA complex. *J Mol Biol* 2002, 316(3):611-627.



67. Papworth M, Kolasinska P, Minczuk M: Designer zinc-finger proteins and their applications. *Gene* 2006, 366(1):27-38.
68. Aasland R, Gibson TJ, Stewart AF: The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends Biochem Sci* 1995, 20(2):56-59.
69. Zheng N, Wang P, Jeffrey PD, Pavletich NP: Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell* 2000, 102(4):533-539.
70. Crawford AW, Pino JD, Beckerle MC: Biochemical and molecular characterization of the chicken cysteine-rich protein, a developmentally regulated LIM-domain protein that is associated with the actin cytoskeleton. *J Cell Biol* 1994, 124(1-2):117-127.
71. Kutateladze TG: Phosphatidylinositol 3-phosphate recognition and membrane docking by the FYVE domain. *Biochim Biophys Acta* 2006, 1761(8):868-877.
72. Goodman RH, Smolik S: CBP/p300 in cell growth, transformation, and development. *Genes Dev* 2000, 14(13):1553-1577.
73. Zheng N, Schulman BA, Song L, Miller JJ, Jeffrey PD, Wang P, Chu C, Koepf DM, Elledge SJ, Pagano M *et al*: Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. *Nature* 2002, 416(6882):703-709.
74. Brzovic PS, Rajagopal P, Hoyt DW, King MC, Klevit RE: Structure of a BRCA1-BARD1 heterodimeric RING-RING complex. *Nat Struct Biol* 2001, 8(10):833-837.
75. Welch PL, King MC: BRCA1 and BRCA2 and the genetics of breast and ovarian cancer. *Hum Mol Genet* 2001, 10(7):705-713.
76. Massiah MA, Simmons BN, Short KM, Cox TC: Solution structure of the RBCC/TRIM B-box1 domain of human MID1: B-box with a RING. *J Mol Biol* 2006, 358(2):532-545.
77. Bailly V, Lauder S, Prakash S, Prakash L: Yeast DNA repair proteins Rad6 and Rad18 form a heterodimer that has ubiquitin conjugating, DNA binding, and ATP hydrolytic activities. *J Biol Chem* 1997, 272(37):23360-23365.
78. Andoniou CE, Thien CB, Langdon WY: Tumour induction by activated abl involves tyrosine phosphorylation of the product of the cbl oncogene. *EMBO J* 1994, 13(19):4515-4523.
79. Joazeiro CA, Wing SS, Huang H, Levenson JD, Hunter T, Liu YC: The tyrosine kinase negative regulator c-Cbl as a RING-type, E2-dependent ubiquitin-protein ligase. *Science* 1999, 286(5438):309-312.
80. Linares LK, Hengstermann A, Ciechanover A, Muller S, Scheffner M: HdmX stimulates Hdm2-mediated ubiquitination and degradation of p53. *Proc Natl Acad Sci U S A* 2003, 100(21):12009-12014.

81. Deshaies RJ, Joazeiro CA: RING domain E3 ubiquitin ligases. *Annu Rev Biochem* 2009, 78:399-434.
82. Sun H, Levenson JD, Hunter T: Conserved function of RNF4 family proteins in eukaryotes: targeting a ubiquitin ligase to SUMOylated proteins. *EMBO J* 2007, 26(18):4102-4112.
83. Tatham MH, Geoffroy MC, Shen L, Plechanovova A, Hattersley N, Jaffray EG, Palvimo JJ, Hay RT: RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. *Nat Cell Biol* 2008, 10(5):538-546.
84. Thomas GS, Zhang L, Blackwell K, Habelhah H: Phosphorylation of TRAF2 within its RING domain inhibits stress-induced cell death by promoting IKK and suppressing JNK activation. *Cancer Res* 2009, 69(8):3665-3672.
85. Schindler U, Beckmann H, Cashmore AR: HAT3.1, a novel Arabidopsis homeodomain protein containing a conserved cysteine-rich region. *Plant J* 1993, 4(1):137-150.
86. Pascual J, Martinez-Yamout M, Dyson HJ, Wright PE: Structure of the PHD zinc finger from human Williams-Beuren syndrome transcription factor. *J Mol Biol* 2000, 304(5):723-729.
87. Dodd RB, Allen MD, Brown SE, Sanderson CM, Duncan LM, Lehner PJ, Bycroft M, Read RJ: Solution structure of the Kaposi's sarcoma-associated herpesvirus K3 N-terminal domain reveals a Novel E2-binding C4HC3-type RING domain. *J Biol Chem* 2004, 279(51):53840-53847.
88. Lu Z, Xu S, Joazeiro C, Cobb MH, Hunter T: The PHD domain of MEKK1 acts as an E3 ubiquitin ligase and mediates ubiquitination and degradation of ERK1/2. *Mol Cell* 2002, 9(5):945-956.
89. Coscoy L, Sanchez DJ, Ganem D: A novel class of herpesvirus-encoded membrane-bound E3 ubiquitin ligases regulates endocytosis of proteins involved in immune recognition. *J Cell Biol* 2001, 155(7):1265-1273.
90. Scheel H, Hofmann K: No evidence for PHD fingers as ubiquitin ligases. *Trends Cell Biol* 2003, 13(6):285-287; author reply 287-288.
91. Aravind L, Iyer LM, Koonin EV: Scores of RINGS but no PHDs in ubiquitin signaling. *Cell Cycle* 2003, 2(2):123-126.
92. Ragvin A, Valvatne H, Erdal S, Arskog V, Tufteland KR, Breen K, AM OY, Eberharter A, Gibson TJ, Becker PB *et al*: Nucleosome binding by the bromodomain and PHD finger of the transcriptional cofactor p300. *J Mol Biol* 2004, 337(4):773-788.
93. Chakravarty S, Zeng L, Zhou MM: Structure and site-specific recognition of histone H3 by the PHD finger of human autoimmune regulator. *Structure* 2009, 17(5):670-679.

94. Pena PV, Davrazou F, Shi X, Walter KL, Verkhusha VV, Gozani O, Zhao R, Kutateladze TG: Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* 2006, 442(7098):100-103.
95. van Ingen H, van Schaik FM, Wienk H, Ballering J, Rehmann H, Dechesne AC, Kruijzer JA, Liskamp RM, Timmers HT, Boelens R: Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. *Structure* 2008, 16(8):1245-1256.
96. Matthews AG, Kuo AJ, Ramon-Maiques S, Han S, Champagne KS, Ivanov D, Gallardo M, Carney D, Cheung P, Ciccone DN *et al*: RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature* 2007, 450(7172):1106-1110.
97. Org T, Chignola F, Hetenyi C, Gaetani M, Rebane A, Liiv I, Maran U, Mollica L, Bottomley MJ, Musco G *et al*: The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression. *EMBO Rep* 2008, 9(4):370-376.
98. Sripathy SP, Stevens J, Schultz DC: The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression. *Mol Cell Biol* 2006, 26(22):8623-8638.
99. Zeng L, Yap KL, Ivanov AV, Wang X, Mujtaba S, Plotnikova O, Rauscher FJ, 3rd, Zhou MM: Structural insights into human KAP1 PHD finger-bromodomain and its role in gene silencing. *Nat Struct Mol Biol* 2008, 15(6):626-633.
100. Kadrmas JL, Beckerle MC: The LIM domain: from the cytoskeleton to the nucleus. *Nat Rev Mol Cell Biol* 2004, 5(11):920-931.
101. Pawson T, Nash P: Assembly of cell regulatory systems through protein interaction domains. *Science* 2003, 300(5618):445-452.
102. Xia H, Winokur ST, Kuo WL, Altherr MR, Bredt DS: Actinin-associated LIM protein: identification of a domain interaction between PDZ and spectrin-like repeat motifs. *J Cell Biol* 1997, 139(2):507-515.
103. Konrat R, Krautler B, Weiskirchen R, Bister K: Structure of cysteine- and glycine-rich protein CRP2. Backbone dynamics reveal motional freedom and independent spatial orientation of the lim domains. *J Biol Chem* 1998, 273(36):23233-23240.
104. Liu Y, Chen W, Gaudet J, Cheney MD, Roudaia L, Cierpicki T, Klet RC, Hartman K, Laue TM, Speck NA *et al*: Structural basis for recognition of SMRT/N-CoR by the MYND domain and its contribution to AML1/ETO's activity. *Cancer Cell* 2007, 11(6):483-497.

105. Ponting CP, Blake DJ, Davies KE, Kendrick-Jones J, Winder SJ: ZZ and TAZ: new putative zinc fingers in dystrophin and other proteins. *Trends Biochem Sci* 1996, 21(1):11-13.
106. De Guzman RN, Wojciak JM, Martinez-Yamout MA, Dyson HJ, Wright PE: CBP/p300 TAZ1 domain forms a structured scaffold for ligand binding. *Biochemistry* 2005, 44(2):490-497.
107. Dial R, Sun ZY, Freedman SJ: Three conformational states of the p300 CH1 domain define its functional properties. *Biochemistry* 2003, 42(33):9937-9945.
108. Legge GB, Martinez-Yamout MA, Hambly DM, Trinh T, Lee BM, Dyson HJ, Wright PE: ZZ domain of CBP: an unusual zinc finger fold in a protein interaction module. *J Mol Biol* 2004, 343(4):1081-1093.
109. Gaullier JM, Simonsen A, D'Arrigo A, Bremnes B, Stenmark H, Aasland R: FYVE fingers bind PtdIns(3)P. *Nature* 1998, 394(6692):432-433.
110. Misra S, Hurley JH: Crystal structure of a phosphatidylinositol 3-phosphate-specific membrane-targeting motif, the FYVE domain of Vps27p. *Cell* 1999, 97(5):657-666.
111. Mao Y, Nickitenko A, Duan X, Lloyd TE, Wu MN, Bellen H, Quioco FA: Crystal structure of the VHS and FYVE tandem domains of Hrs, a protein involved in membrane trafficking and signal transduction. *Cell* 2000, 100(4):447-456.
112. Kutateladze TG, Ogburn KD, Watson WT, de Beer T, Emr SD, Burd CG, Overduin M: Phosphatidylinositol 3-phosphate recognition by the FYVE domain. *Mol Cell* 1999, 3(6):805-811.
113. Kutateladze T, Overduin M: Structural mechanism of endosome docking by the FYVE domain. *Science* 2001, 291(5509):1793-1796.
114. Meplan C, Mann K, Hainaut P: Cadmium induces conformational modifications of wild-type p53 and suppresses p53 response to DNA damage in cultured cells. *J Biol Chem* 1999, 274(44):31663-31670.
115. Schmidt D, Durrett R: Adaptive evolution drives the diversification of zinc-finger binding domains. *Mol Biol Evol* 2004, 21(12):2326-2339.

## Chapter 2: Production of zinc finger proteins

### 1. Strategy for the production of zinc finger proteins

The study of protein structure using NMR requires the production of samples in high yield and good purity. In the past, approaches such as peptide synthesis or purification from natural sources were used widely in protein studies.[1] With the development of biotechnology, more and more proteins could be produced readily using recombinant techniques.[2] For NMR studies, this allows target proteins to be labeled with stable isotopes such as  $^{15}\text{N}$  and/or  $^{13}\text{C}$  by incorporation of these nuclei into recombinant protein segments. In addition, this technology also permits the production of proteins on a large scale, corresponding to the need for milligram quantities of proteins for NMR studies. Fig. 1 shows a typical procedure for protein production using recombinant methods. The process of protein expression involves the growth of cells in culture media and induction by IPTG, then harvesting by centrifugation. Purification of proteins is initiated by the disruption of cells to release soluble protein. Standard procedures for protein purification are well established, but, in many cases, modification of these procedures is required, depending on the nature of the desired proteins. For zinc fingers, in particular, several strategies may be followed including selection of affinity tags, optimization of expression and buffer conditions and efficient removal of affinity tags.

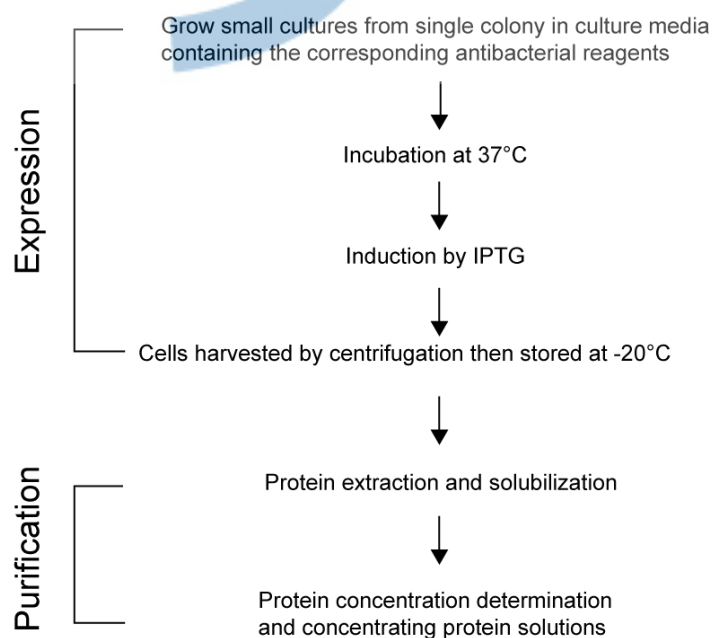


Fig.1 Representative scheme of the procedure of protein production using *E.coli* as hosts.

## 1.1 Expression of zinc finger proteins

Of the many available systems applied for the production of recombinant proteins, the Gram-negative bacterium *Escherichia coli* (*E.coli*) is the most common host that has been developed to express heterologous proteins.[3] Characteristics of *E.coli* such as high rate growth, well-defined genetics as well as low cost result in a high efficiency of protein expression.[4]

### 1.1.1 Selection of affinity tags for zinc finger proteins

For protein expression and to study protein-protein interactions a target protein is often fused to an affinity tag to allow its separation from the cell lysate using affinity chromatography.[5] An affinity tag is an exogenous amino acid sequence which has a high affinity for a specific ligand. The proteins carrying such an affinity tag exhibit a high selectivity for affinity resins. Table 1 lists the common affinity tags for recombinant proteins. A major class of affinity tags consists of peptide or protein sequences that bind a small ligand linked on a solid support. This class includes hexa-histidine (His-tag), glutathione-S-transferase (GST)[6], maltose-binding protein (MBP)[7] and thioredoxin (TRX) [8]. Other types include tags such as antibodies that bind to immobilized proteins. Target proteins fused with specific affinity tags can improve the protein yield [9], prevent proteolysis [10], facilitate protein refolding and increase solubility[11]. However, in some cases, the addition of tags was also shown to have negative effects on target proteins such as inducing a change in protein conformation [12], reducing protein yields [13], altering biological activity [14], inhibiting the enzymatic activity [15], creating unexpected flexibility that is detrimental for structural studies [16] and inducing toxicity [17].

Affinity and solubility tags for recombinant proteins

Tag	Size (aa)	Comments
His-tag	5–15	Purification under native or denaturing conditions
FLAG	8	Calcium-dependent, mAb-based purification
Streptag II	8	Modified streptavidin, elution with biotin analog
HA-tag	9	Influenza virus hemagglutinin tag, Ab-based purification
Softag1, Softag 3	13, 8	Recognized by polyol-responsive mAb
c-myc	10	mAb-based purification
T7-tag	11–16	mAb-based purification
S-tag	15	S-protein resin affinity purification
Elastin-like peptides	18–320	Protein aggregation by temperature shift, intein used to remove tag
Chitin-binding domain	52	Binds only insoluble chitin (see intein, Table 3)
Thioredoxin	109	Affinity purification with modified resin
Xylanase 10A	163	Cellulose based capture, elution with glucose
Glutathione S-transferase	201	Glutathione or GST-Ab affinity
Maltose binding protein	396	Amylose affinity purification
NusA	495	Increased solubility in <i>E. coli</i> . Affinity tag needed for purification

Table 1 Lists of affinity tags for recombinant proteins.[18]



Of these affinity tags, His-tags are used widely with chelated metal ions as affinity ligands. Histidine residues exhibit a strong interaction with immobilized metal ion matrices, as electron donor groups on the histidine imidazole ring form coordination bonds with the immobilized transition metal. Peptides containing sequences of consecutive histidine residues are retained on affinity chromatography. However, the strong interaction of histidines with chelated metal ions may affect the fold or the solubility of zinc finger proteins. For example, the expression of single zinc finger peptide ZNF191 using a His-tag fusion protein was unsuccessful, resulting in changes to its biological activity [19], although the reasons for this were not clear. In another study, attachment of a His-tag to a zinc finger protein also led to a conformational change at a DNA binding site.[12] Despite these drawbacks of the His-tag, several other zinc finger proteins have been purified successfully from His-tag recombinant proteins and revealed high yield and purity. [20, 21]

A number of zinc finger proteins have been produced with GST tags. GST is a 26-kDa protein which was cloned from *E.Coli*. It was first used as an affinity tag in a single-step purification in 1988.[6] As for the His-tag, the GST tag can also be linked to either the N- or C-terminus of target proteins and can be used in various microorganisms such as bacteria, yeast or mammalian cells.[6, 22] GST can help to protect the recombinant protein against intracellular protease cleavage. In most cases, GST fusion proteins are soluble or partly soluble; however, the insolubility of GST fusion proteins may result from the presence of hydrophobic regions.[23] Although GST is not expected to affect the activity of target proteins, removal of the tag is essential because of its large size. To date, more and more zinc finger proteins were purified from GST fusion proteins and showed high solubility during expression.[21, 24, 25]

Zinc finger proteins can also be attached to other kinds of affinity tags such as MBP [25] or ketosteroid isomerase (KSI) [19]. However, successful protein purification depends not only on choosing an appropriate affinity tag but also on other factors such as optimal expression conditions.

### 1.1.2 Optimal expression condition

In the process of expression, the zinc ion not only stabilizes the overall fold of zinc finger proteins but is also important for promoting culture growth. As a result, the investigation of optimal conditions for expression becomes a critical issue. A shortage of zinc during over-expression may produce misfolded proteins, resulting in protein aggregation during purification [26] whereas an excess of zinc may inhibit cell growth and result in cell toxicity.[25, 27] Biochemical studies have shown that a maximal zinc concentration for the growth of *E.Coli* is about 30 $\mu$ M.[27] As a result, to avoid protein



aggregation, addition of zinc in culture media before starting protein expression is necessary. The optimal dosage of zinc ranges from 6 to 25  $\mu\text{M}$ , depending on the solubility of the protein upon expression at various temperatures.[25] In addition, induction at lower temperatures can also minimize protein aggregation.[28]

## **1.2 Purification of zinc finger proteins**

Several factors should be considered before starting the purification of zinc finger proteins, including the selection of an appropriate buffer, affinity tags for recombinant proteins and the removal of tags properly by proteases. Since the zinc ion in zinc containing proteins plays a critical role in the fold, all the purification procedures should be performed carefully.

### **1.2.1 Preparation of an optimal buffer**

Several buffer parameters should be considered when selecting a buffer including pH value, its variation with temperature, ionic strength, and interaction of the protein with other buffer components such as detergents and solubility. Tris and phosphate are the most common components used in buffer preparations. It should be noted that the pH of Tris changes by 1 pH unit when the temperature is raised from 4 to 37°C.[29] An optimal buffer is able to enhance the solubility and stability of the protein during purification. Protein stability is enhanced as the NaCl concentration is increased from 0 to 1M; concentrations higher than 1M have no noticeable effect.[30] The salt concentration may, however, have detrimental effects on the protein activity. For instance, C<sub>2</sub>H<sub>2</sub> zinc finger proteins that possess DNA binding activities are highly sensitive to salt and pH conditions.[30, 31] GR-DBD only binds to DNA at low salt concentrations and a shift of pKa of the histidine residues also affects its DNA binding properties.[32] The choice of the optimal pH for a structural study depends on the pI of the protein, the pKa of the histidines if they are present, and rates of amide proton exchange. This latter condition sets the maximum value of pH for most NMR studies to 7.5. Media used to stabilize zinc-containing proteins should avoid oxidation of cysteines. Most commonly, reducing agents such as  $\beta$ -mercaptoethanol, dithiothreitol (DTT) and TECP are used for this purpose. Special care is also taken to remove dissolved oxygen by degassing all solutions.[30] In extreme cases, these methods are not sufficient to prevent protein oxidation and protein manipulation must then be performed under a controlled atmosphere.

### **1.2.2 Removal of tags by proteases**

Although affinity tags can help to stabilize the folds of target proteins during purification, they may also affect the characteristics or functions of proteins and

produce artifactual effects in a protein study. For these reasons, the removal of the tag from the protein of interest is necessary and can be achieved using a site-specific protease. Most importantly, the cleavage should not affect protein activity.

Cleavage of the tag can be performed using different systems. Table 1 lists a number of proteases used for the removal of tags and Fig.2 illustrates the strategies for the removal of tags. In addition to the use of a site-specific protease, it is also possible to adopt “self-cleavage” strategies in which the tag cleavage is removed by using a self-splicing intein.[33] In practice, cleavage of the tag is commonly achieved by proteases such as thrombin [34], factor Xa [35] , and tobacco etch virus (TEV) [36]. Thrombin is the most widely used enzyme to cleave tags. Cleavage can be achieved at temperature between 20 and 37°C for 0.3-16 h. Factor Xa cleaves at the carboxyl side of the four-amino-acid peptide I-E[D]-G-RX1.[37] Cleavage can be carried out at temperatures ranging from 4 to 25°C.

TEV protease recognizes the sequence E-X-X-Y-X-Q-S (where X can be any amino acid, although not all are tolerated) and cleavage occurs between the glutamine and serine.[38] Although TEV protease was shown to have better activity at 34°C, it also can be used at temperatures as low as 4°C and remains active over a wide range of conditions and in the presence of various protease inhibitors.[39, 40] A recent study showed that a fusion protein incubated with TEV protease can exist stably for two weeks at 4°C or room temperature.[41] In spite of the durability of TEV protease, its efficiency of cleavage must be considered since proteins may degrade or precipitate during long reaction times. The ratio of enzyme to proteins and the reaction time are therefore important.[42] To avoid protein precipitation or degradation that may occur during reactions and to increase the efficiency of cleavage, reduction of the reaction time is necessary and investigation of optimal conditions by screening various enzyme/substrate ratios is also critical in purification.

Enzyme	Cleavage sites	Comments	References
Enterolisin	DDDDK*	Secondary sites at other basic amino acids.	[43]
Factor Xa	IDGR*	Secondary sites at GR.	[35]
Thrombin	LVPR* <b>GS</b>	Secondary sites. Biotin labeled for removal of the protease.	[44]
PreScission	LEVLFQ* <b>GP</b>	GST tag for removal of the protease.	[45]
TEV protease	EQLYFQ* <b>G</b>	His-tag for removal of the protease.	[46]
3C protease	ETLFQ*GP	GST tag for removal of the protease.	[47, 48]
Sortase A	LPET* <b>G</b>	Ca <sup>2+</sup> -induction of cleavage requires an additional affinity tag (e.g., His-tag) for on column tag removal.	[49]
Granzyme B	D*X, N*X, M*N, S*X	Serine protease. Risk for unspecific cleavage	[50]
Intein	Self-cleavable	Artificial amino acid left after cleavage in some applications. On column (chitin-beads) cleavage.	[51, 52]
SUMO	Conformation	No affinity purification per se (requires His-tag).	[53]
DAPase (TAGZyme)	Exo(di)peptidase	Cleaves N-terminal. His-tag (C-terminal) for purification and removal.	[54]
<i>Areomonas aminopeptidase</i>	Exopeptidase	Cleaves N-terminal, effective on M, L. Requires Zn.	[55]
Aminopeptidase M	Exopeptidase	Cleaves N-terminal, does not cleave X-P.	[56]
Carboxypeptidase A	Exopeptidase	Cleaves N-terminal. No cleavage at X-R, P.	[56]
Carboxypeptidase B	Exopeptidase	Cleaves N-terminal R, K.	[56]

Table 2 Enzymatic methods for tag removal. The position of endoprotease cleavage is indicated with an asterisk (\*). Residues in bold remain in the protein after endoprotease cleavage. (Adapted from [18])

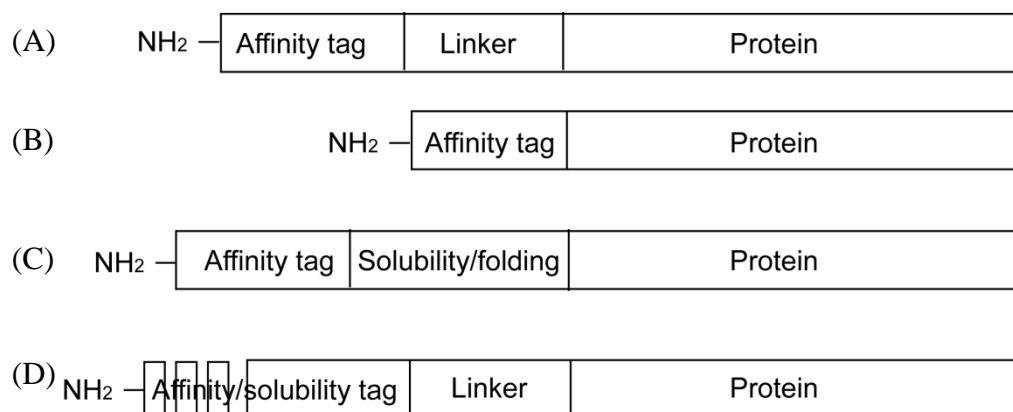


Fig.2 Strategies for using protein tags. N-terminal fusions are shown as examples. (A) A fusion protein contains an affinity tag, a linker region including a specific sequence for endoprotease cleavage and the native protein sequence. (B) a fusion protein designed for exopeptidase removal of the tag (e.g. TAGZyme for N-terminal cleavage) (C) is similar to (A) but the affinity tag (e.g. MBP, GST) increases the solubility of the protein. (D) a fusion protein where a solubility and folding partner is fused N-terminal to the target protein (e.g. SUMO, sortase A). (Adapted from [18])

## 2. Materials and methods

### 2.1. Culture media

Luria Broth (LB) medium

LB medium is a bacterial growth medium for the expression of unlabelled proteins. 1L of LB contains 10.0g of peptone, 5.0g of NaCl and 5.0g of yeast extract.

M9 minimal medium

This medium is used to express labeled proteins. 1L M9 minimal medium contains 17 g of Na<sub>2</sub>HPO<sub>4</sub> 12(H<sub>2</sub>O), 3.0g of KH<sub>2</sub>PO<sub>4</sub>, and 0.5g of NaCl.

The following is the composition of medium for the expression of <sup>15</sup>N labeled protein

Culture volume	100 ml	1L
10X M9 medium	10 ml	100 ml
20% Glucose	2 ml (0.4g)	20 ml (4.0 g)
25% <sup>15</sup> NH <sub>4</sub> Cl	400 µl (0.1g)	4 ml (1.0 g)
1M MgSO <sub>4</sub>	200 µl	2 ml
1M CaCl <sub>2</sub>	10 µl	100 µl
3.5mM FeSO <sub>4</sub>	10 µl	100 µl
Thiamine (1mg/ml)	100 µl	1 ml
Vitamine-mineral stock*	100 µl	1 ml

\*Commercial product: Thermophilus Vitamin-Mineral Stock 100X (TYE Media,

Castenholz Media) (Powder)

## 2.2 Plasmid transformation

Protocols for the transformation of plasmids containing ARDBD and ATXN7:

Protein	ARDBD wild type/T575A mutant	ATXN7 and ATXN7 L3
Vector	pETM30 containing GST tag	pGEX-4T1 containing GST tag
Competent cells	BL21 (DE3) star	BL21 (DE3) pRARE
DNA transformation	Heat shock process	Heat shock process
Length of protein sequence	555-627 (GAMG were added to the N-terminus)	ATXN7: 330-401 (GS were added to the N-terminus) ATXN7 L3: 197-276 (GSHM were added to the N-terminus)

## 2.3 Culture growth and protein expression

Conditions for the expression of ARDBD, ATXN7 and ATXN7 L3 list as follows:

Protein	ARDBD wild type/T575A mutant	ATXN7 and ATXN7 L3
Isotope labeling	<sup>13</sup> C, <sup>15</sup> N	<sup>13</sup> C, <sup>15</sup> N
Media for preculture	LB	LB
Time of incubation for preculture	overnight	overnight
Antibacterial reagent	kanamycin (1000 times dilution)	ampicilin (1000 times dilution)
Culture media	M9	M9
Incubation temperature	37°C	37°C
Incubation period	2~4 hr until the OD <sub>600</sub> reached to 0.6.	2~4 hr depends on the rate of cell growth
Induction conditions	1mM IPTG 18°C, overnight	1mM IPTG 18°C, overnight
Culture harvest	Centrifugation 4000 rpm, 20min	Centrifugation 4000 rpm, 20min
Storage	-20°C	-20°C

## 2.4 Protein purification

### 2.4.1 Buffers used for purification

Buffers used in purification of ARDBD, ATXN7 and ATXN7 L3 list as follows:

Buffer	ARDBD wild type/T575A mutant	ATXN7 and ATXN7 L3
Binding buffer	50mM Tris-HCl pH 7.5/150mM NaCl/10% of glycerol/1mM DTT	20mM Tris pH 7.5/250mM NaCl/2mM DTT
Washing buffer	50mM Tris-HCl pH 7.5/150mM NaCl/1mM DTT	20mM Tris pH 7.5/250mM NaCl/2mM DTT

Protease using for tag cleavage	<b>TEV protease</b>	<b>Thrombin</b>
Buffer using in gel filtration	50mM phosphate pH 6.5/150mM NaCl/1mM DTT	50mM phosphate pH 7.0/200mM NaCl/2mM DTT
Sample buffer for NMR study	50mM phosphate pH 6.5/1mM DTT	50mM phosphate pH 7.0/200mM NaCl/2mM DTT

### 2.4.2 Steps for protein purification

All purifications were performed at 4°C and samples were stored on ice between all steps. Procedures for the purification of ARDBD, ATXN7 and ATXN7 L3 were as follows:

Protein	ARDBD wild type/T575A mutant	ATXN7 and ATXN7 L3
1. Cell lysis	Cells were resuspended in 40mL of binding buffer (for 1L culture) and lysed using sonicator. Parameters of sonicator: Pulse on: 2sec Pulse off: 2sec Time duration: 5min Power amplitude: 30%	As for ARDBD
2. Equilibration of the affinity resin by binding buffer	1.5ml of glutathione sepharose 4B resin was equilibrated by binding buffer.	As for ARDBD
3. Removal of lysed cells	Ultracentrifugation at 40000 rpm, 1hr.	As for ARDBD
4. Incubation with affinity resin	Supernatant was incubated with 1.5 ml of glutathione sepharose 4B for 2 hr.	As for ARDBD
5. Washing the resin	The fusion protein was washed three times with the washing buffer. Each time, the fusion protein was mixed with washing buffer and centrifuged at 800 rpm for 5 min.	As for ARDBD
6. Remove tag by protease	The fusion protein was incubated with <b>TEV protease</b> for <b>2 hr</b> at <b>room temperature</b> . Screening of an optimal conditions for TEV cleavage was performed as in section 2.5.	The fusion protein was incubated with <b>thrombin overnight</b> at <b>4°C</b> .
7. Elution	The protein was eluted from the affinity resin. Quality of eluted proteins was checked using Bradford assay. Proteins were eluted completely using a series of washes with	As for ARDBD

	washing buffer.	
8. Gel filtration: proteins were performed by Superdex G75 (HiLoad 16/60) gel filtration column	Protein was injected in a 5ml-loop attached to the AKTA machine. The flow rate was 1ml/min. The protein was eluted between 80~85ml. Note: for 5ml-loop, the injection volume of proteins is limited to 5ml.	As for ARDBD
9. Sample concentration	The eluted protein was concentrated using 3K amicon by centrifugation. The speed for centrifugation was 3000 rpm and the final volume is about 500µl.	As for ARDBD
10. Measurements of protein concentrations	Protein concentrations were measured either by Bradford assays or nanodrop UV spectroscopy.	The measurement of ATXN7 L3 concentration was performed as for ARDBD whereas the concentration of ATXN7 could only be measured by Bradford assay.

## 2.5 Expression and purification of recombinant TEV protease

1. Plasmid transformation	
Vector	pET vector containing His-tag
Competent cells	BL21 (DE3) <i>pLysS</i>
DNA transformation	Heat shock process
2. Expression	
Medium for preculture	LB medium with 15µg/ml kanamycin and 34µg/ml chloramphenicol
Time of incubation for preculture	2 or 3 clones were selected and grown in 5ml LB at 37°C overnight
Culture media	LB medium with 15µg/ml kanamycin and 34µg/ml chloramphenicol
Incubation temperature	37°C
Incubation duration	2~4 hr until the OD <sub>600</sub> reached to 0.6.
Induction conditions	1mM IPTG, 25°C overnight
Culture harvest	Centrifugation 4000 rpm, 20min
Storage	-20°C
3. Buffer preparation for purification	
Buffer A	50mM Tris-HCl pH 6.8/150mM NaCl/2mM DTT
Buffer B	50mM Tris-HCl pH 6.8/150mM NaCl/2mM DTT/10mM imidazole
Buffer C	50mM Tris-HCl pH 6.8/1M NaCl/2mM DTT/10mM imidazole
Buffer D	50mM Tris-HCl pH 6.8/150mM NaCl/2mM DTT/40mM imidazole



Buffer E	50mM Tris-HCl pH 6.8/150mM NaCl/2mM DTT/300mM imidazole
4. Purification: All procedures were performed at 4°C and samples were stored on ice between all steps.	
1. Cell lysis	Cells were resuspended in 40mL of binding buffer (for 1L culture) and lysed using sonicator. Parameters of sonicator: Pulse on: 2sec Pulse off: 2sec Time duration: 5min Power amplitude: 30%
2. Equilibration of the affinity resin by binding buffer	4ml of Ni-NTA superflow resin was equilibrated by buffer B.
3. Removal of lysed cells	Ultracentrifugation at 36000 rpm, 45 min.
4. Incubation with affinity resin	Supernatant was loaded on a Ni-NTA superflow resin.
5. Washing the resin	The resin was washed first using buffer B and then by buffer C (2~3 times). The final wash was used buffer D.
6. Elution	Elution was performed using buffer E.
7. Desalting	The column was equilibrated using buffer A before desalting. The eluted protein was loaded on PD 10 desalting column and the flow through was eluted immediately from the column. TEV protease was eluted after 3.5ml of buffer A was added to the column.
8. Storage	The protein concentration was determined by Bradford assay and the samples were stored in buffer A with 20% glycerol at -80°C.

## 2.6 Methods for the removal of a tag by TEV protease

### 2.6.1 Selection of a proper approach for TEV protease cleavage

Two protocols can be used to remove the GST tag from AR-DBD fusion proteins. One is to add the protease directly to the eluted protein. Another is to add the protease to the fusion protein bound to an affinity resin and to elute the soluble fraction. The efficiency of elution was assessed by 15% SDS-PAGE. The reactions were carried out at 4°C overnight.

### 2.6.2 Screening of optimal conditions for TEV protease cleavage

A series of tests were performed by adding various ratios of TEV protease to AR-DBD wild type at room temperature. Three ratios of TEV to protein were used in the screen: 1:500, 1:100 and 1:10 and the reaction times were set to 1 and 2 hour. The reaction was stopped by the addition of 2x SDS sample buffer, and the samples were

then analyzed by 15% SDS-PAGE.

### **3. Results and discussion**

In this section, two zinc finger domains, AR-DBD and ATXN7 are discussed individually in terms of their behavior in experiments using biochemical approaches.

#### **3.1 Expression and purification of AR-DBD wild type and T575A mutant**

Given the pI of 9.14 for both proteins, they are expected to be soluble and stable in a buffer at pH 7.0~8.0. As a result, a buffer containing 50mM of Tris-HCl, pH 7.5 was chosen for purification. To prevent the oxidation of cysteine residues, the addition of 1mM of DTT was essential. Moreover, the addition of 10% glycerol was used to enhance protein stability. Fig.3 shows the gel filtration profiles and 15% SDS-PAGE of AR-DBD WT. AR-DBD WT was eluted in the volume between 85 and 90 ml in gel filtration corresponding to a maximum absorption peak. The quality of each eluting fraction in gel filtration was checked by 15% SDS-PAGE and showed a high yield and purity. The yield was about 8.6mg/ml as measured by Bradford assay.



UNIVERSITÉ DE STRASBOURG

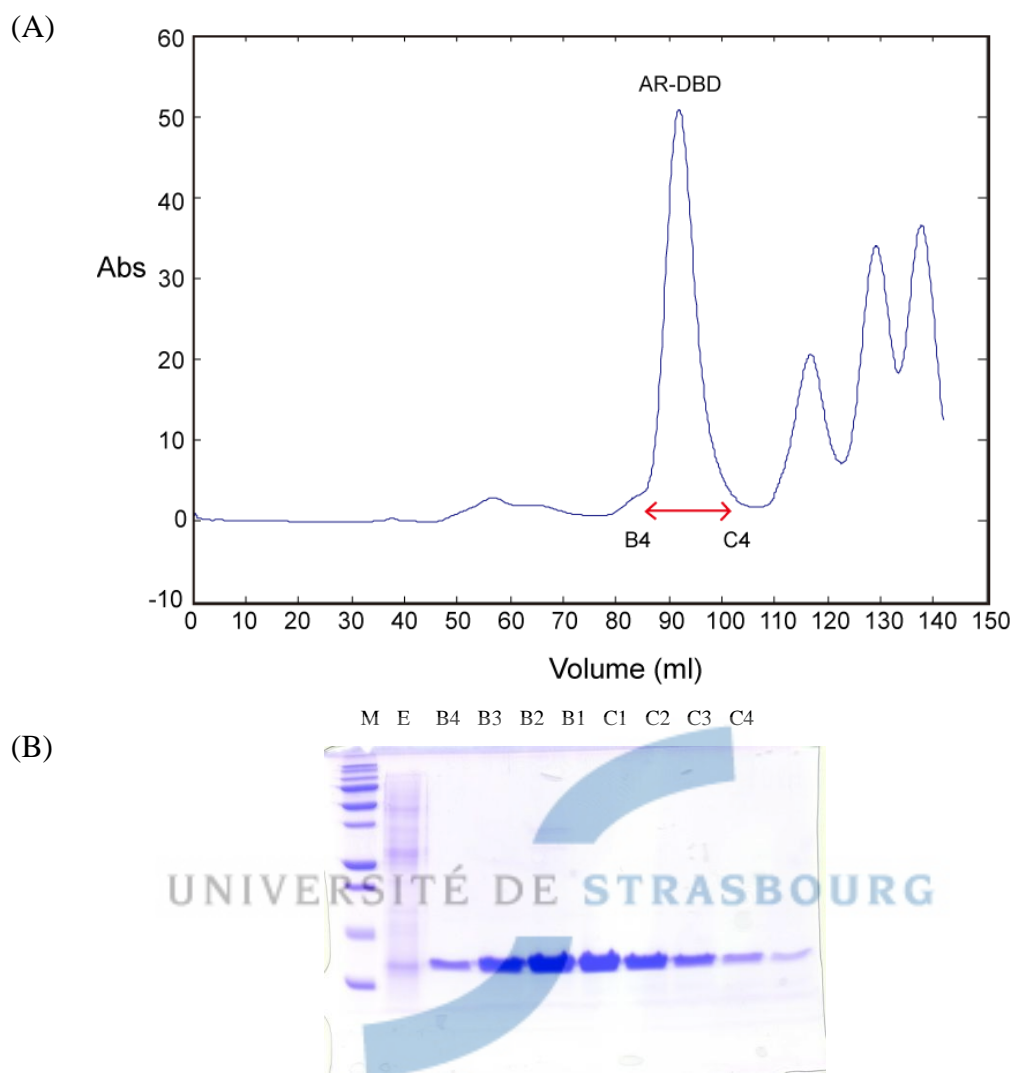


Fig.3 (A) Profile of gel filtration of AR-DBD WT expressed from 1L culture in LB medium. The protein was eluted between 80~90 ml during gel filtration. (B) 15% SDS-PAGE of AR-DBD wild type. Fractions B4 to C4 were collected and run on an electrophoresis gel. Bands of each fraction exhibit high purity. (M: protein marker; E: elute from the beads; B4~C4: elute position marked in the collector)

### 3.2 Comparison of sample quality between ATXN7 fused with His-tag and GST tag

Initial constructs of ATXN7 were designed using a His-tag which gave a low yield of a protein (final sample concentration of protein about only 50 $\mu$ M could be obtained). Despite the low yield, it was possible to record a  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum (Fig.4A). Dispersed peaks appeared in the spectrum, an indication of a folded protein. However, some weak peaks adjacent to the major peaks were also observed in the spectrum. The appearance of additional peaks in spectra may suggest multiple conformations of the protein which is highly detrimental for structure determination. In addition, many peaks are observed in the center of the spectrum (around 8.0ppm in  $^1\text{H}$  chemical shift),

suggesting partial unfolding of the protein. Fig.4B shows the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of ATXN7 purified from GST-tagged fusion protein. In contrast to Fig.4A, each observed peak in Fig.4B is unique and distinct, in particular those in the central region of the spectrum. Most importantly, no additional peaks were detected in the spectrum, indicating a single conformation.

These results revealed that the metal binding properties of His-tags can affect the protein quality of zinc finger domains ATXN7. The presence of cysteine residues or naturally occurring histidine-rich regions may lead to unwanted protein binding during purification.[57] ATXN7 not only binds a zinc ion but also contains four histidines, including H367 which is involved in zinc coordination. The metal-chelating properties of ATXN7 may lead to protein precipitation or aggregation when the fusion protein is bound to the immobilized affinity column. Evidence for this is provided by the appearance of overlapping peaks in the central region of HSQC spectrum in Fig.4A. In addition, the overall conformation of ATXN7 may also be perturbed by metal chelation by the His-tag and result in the occurrence of multiple conformations. The effect may give rise to the minor peaks in Fig.4A.

To summarize, in the case of ATXN7 purification, the protein fused with a GST tag gave high yield and good quality in contrast to the His-tag fusion protein. Since the His-tag could be responsible for this phenomenon, by perturbation of the zinc coordination sphere, another tag was chosen.

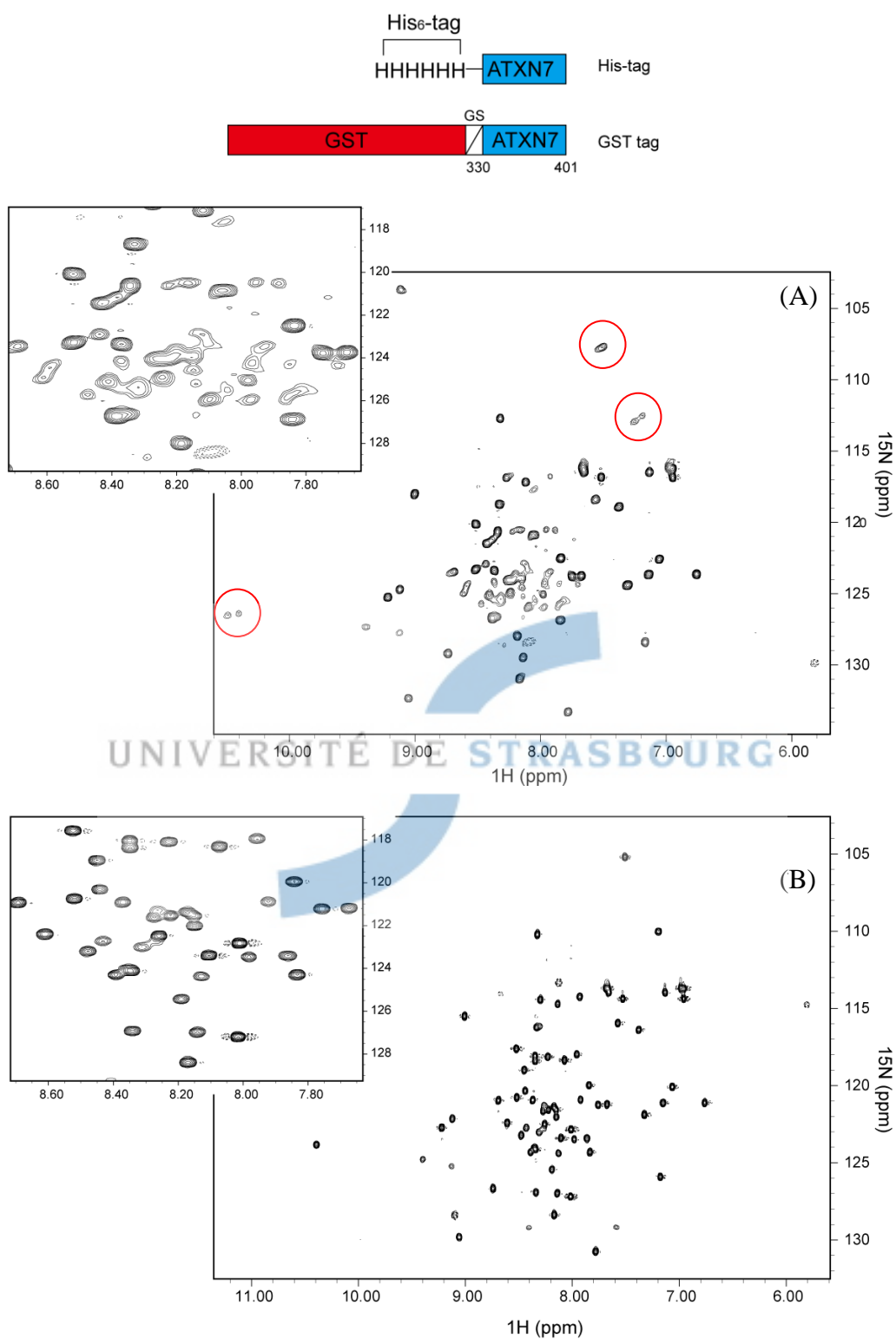


Fig.4  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of ATXN7 purified from fusion proteins with (A) His-tag and (B) GST tag. The constructs are shown schematically on the top of the figure. Several overlapping peaks near 8.0 ppm ( $^1\text{H}$  chemical shift) are shown in (A), revealing the partial unfolding of the protein. In addition, minor peaks (present in red circles) are observed in the spectrum, suggesting that multiple conformations are present.

### 3.3 Removal of affinity tags from recombinant AR-DBD proteins by TEV protease

Fig.5 shows the results of removal of tags from AR-DBD wild-type fused with GST tag. The cleavage reaction was tested at 4°C overnight because the protein is stable at low temperature. The comparison of lanes 2 and 4 clearly shows that cleavage on the resin exhibits better performance than that of the eluted protein.

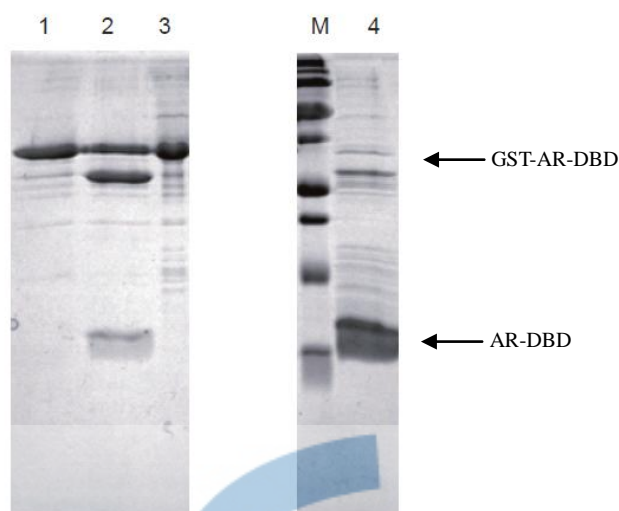


Fig.5 Two different methods for TEV protease cleavage after protein elution (Lane 1~3) and on the resin directly (Lane 4). Lane 1: the protein eluted from the resin. Lane 2: the eluted protein after cleavage by TEV protease at 4°C overnight. Lane 3: the resin after treatment with TEV protease. Some fusion protein was retained. Lane 4: the elution fraction after treatment with TEV protease directly on the resin. The reaction conditions were the same as in lanes 1~3. Almost all the fusion protein was cleaved by TEV protease.

However, the overnight reaction was too long and may lead to protein degradation or precipitation. Furthermore the cleavage conditions were not reproducible. Fig.6A showed that, in separate experiment protein degradation occurred at 4°C following TEV protease cleavage overnight. Lanes E1, E2 and E3 contain not only bands of AR-DBD but also non-specific bands with molecular weight in the range from 10kDa to 15kDa. These impurities were also observed in the gel filtration profile (Fig.6C). To ensure that the non-specific bands were not derived from DNA contaminants, an agarose gel was used to detect DNA. Fig.6D shows that no DNA contaminants are present in the sample. These results indicate that improper conditions of TEV protease cleavage lead to protein mis-folding and thus contribute to the production of an unstable protein. Consequently, it was crucial to investigate the optimal conditions for tag cleavage by TEV protease.

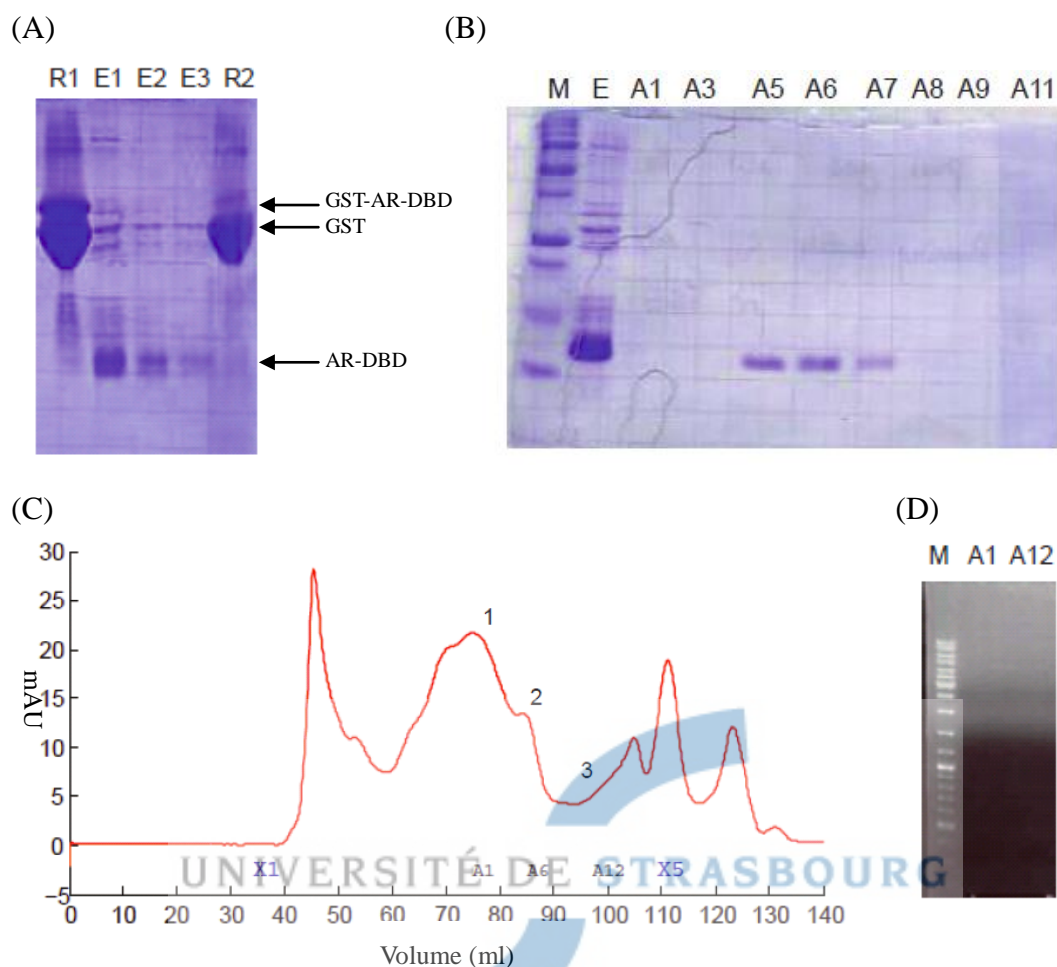


Fig.6 The quality of protein after treatment with TEV protease of AR-DBD wild type fusion protein bound to glutathione sepharose 4B resin at 4°C overnight was checked by (A)(B)15% SDS-PAGE , (C) gel filtration and (D) DNA gel. The arrows in (A) indicate the band of AR-DBD which corresponds to its molecular weight. The lanes in B correspond to the absorption peaks 2 and 3 in C. Lane M: molecular weight marker; Lane R1: affinity resin containing fusion protein before treatment of TEV; Lane E1~E3: Eluting fractions of AR-DBD; Lane R2: the resin after treatment of TEV protease; Lane A1~A11: each fraction after performance of gel filtration.

The investigation of optimal conditions by screening various reaction times and temperature was necessary. Previous reports have shown that TEV protease is maximally active at 34°C and the reaction is only three-fold less active at 4°C than at 20°C. [46] The reaction time and reaction temperature should be controlled simultaneously because while the TEV protease activity may decrease while the temperature increased and may be destroyed under the long reacting time. To compensate the dual effects on temperature and time as well as to enhance the cutting efficiency, the cleavage was performed at room temperature. In addition, the reaction



time was reduced to 2 hours to prevent protein degradation. Fig.7 shows the results of TEV protease cleavage at three different ratios of TEV protease to proteins. With a ratio of TEV protease to proteins was 1:500, the reaction occurred in 1 hour but was not complete after 2 hours because some fusion proteins were still retained on the resin (lane R2 of 1:500). To increase the efficiency of TEV protease cleavage, a ratio of 1:100 TEV proteases was tested. The reaction was almost complete within 1 hour; however, the activity of the TEV protease appeared to decrease in the second hour because no further amounts of proteins could be eluted (Lane E2 of 1:100). In addition, a portion of the fusion proteins were retained on resin with the GST tag not cut by TEV protease (Lane R2 of 1:100). The degree of cleavage reached a maximum at a ratio of 1:10 with no fusion proteins being retained on the resin (Lane R2 of 1:10). In addition, no non-specific bands appeared in gel, suggesting a greater stability during the cleavage.

For the removal of the GST affinity tags from AR-DBD fusion proteins by TEV protease, the optimal conditions for cleavage were found to be incubation of proteins with protease at room temperature for 2 hours. The ratio of TEV to proteins at 1:10 gave the best performance for cleavage.

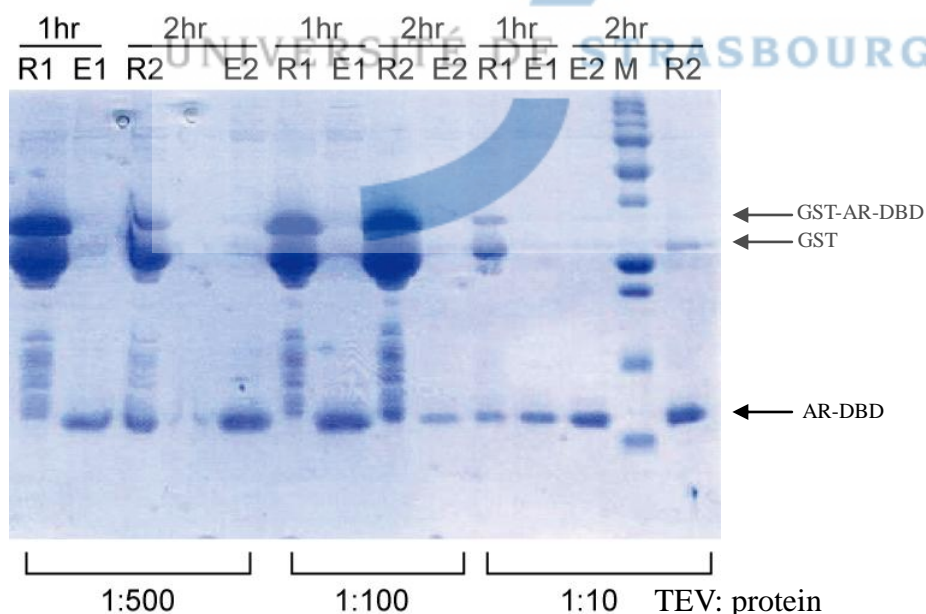


Fig.7 Performance of removal of GST from AR-DBD WT fusion protein by TEV protease with different conditions. The fusion protein was incubated with TEV for 1 and 2 hour at room temperature. Conditions for the removals of GST by TEV protease were tested at ratios of TEV protease to protein of 1: 500, 1:100 and 1:10. The highest efficiency of TEV cleavage was given by a ratio of 1:10 with cleavage being complete within 2 hours. Lane R2 shows that no GST-tagged proteins remained on the resin. Lane R1 and R2: fusion protein on resin after treatment with TEV protease 1 hour and 2 hour, respectively; Lane E1 and E2: eluted fractions for the 1<sup>st</sup> and 2<sup>nd</sup> hour, respectively.

## 4. Conclusions

This chapter has provided some insight into the purification of zinc finger proteins including methods for the maintenance of stability of metal ions during expression, the selection of suitable affinity tags and buffers for purification and proper conditions for tag cleavage by proteases. The goal in all steps is to obtain a sample in high yield and of good quality which is important for studies of protein structure using NMR.

For the preparation of zinc finger proteins, a supplement of zinc during expression enhances protein stability. Careful design of recombinant proteins with affinity tags not only ensures folding ability of expressed proteins but also enhances their stability in culture growth. During purification, the selection of buffer should take into account the effect of pH on histidine protonation. For the purification of AR-DBD, ATXN7 and ATXN7 L3, Tris-HCl buffer is therefore more suitable than phosphate buffer and the use of anti-oxidant reagents such as  $\beta$ -mercaptoethanol or DTT prevented the formation of disulfide bond between cysteine residues, especially for cysteine residues involving in zinc coordination.

For the purification of ATXN7, recombinant proteins with GST tags gave better yields and higher quality of proteins than those produced with His-tags, indicating the detrimental effects of the metal chelating properties of His-tags on the stability of metal binding sites in zinc fingers. In addition, sample of ATXN7 produced using GST tags revealed a single conformation and gave better  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra with a set of cross-peaks that are distinct and entirely resolved.

The removal of tags from recombinant proteins is also important in protein purification. The optimal conditions for tag cleavage from recombinant AR-DBD proteins were found to be reaction at room temperature for 2 hours and an addition of TEV protease at a ratio of protease to proteins of 1:10. For both ATXN7 and ATXN7 L3, tags were cleaved using thrombin at 4°C overnight. In both cases, better performance was achieved when the tags were cut from the recombinant proteins on the resins.

## References

1. Fridkin M, Patchornik A: **Peptide synthesis**. *Annu Rev Biochem* 1974, **43**(0):419-443.
2. Schein CH: **Production of Soluble Recombinant Proteins in Bacteria**. *Bio-Technology* 1989, **7**(11):1141-1147.
3. Sagermann M, Baase WA, Mooers BH, Gay L, Matthews BW: **Relocation or duplication of the helix A sequence of T4 lysozyme causes only modest changes in structure but can increase or decrease the rate of folding**. *Biochemistry* 2004, **43**(5):1296-1301.

4. Baneyx F: **Recombinant protein expression in Escherichia coli.** *Curr Opin Biotechnol* 1999, **10**(5):411-421.
5. Narayanan SR, Crane LJ: **Affinity chromatography supports: a look at performance requirements.** *Trends Biotechnol* 1990, **8**(1):12-16.
6. Smith DB, Johnson KS: **Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase.** *Gene* 1988, **67**(1):31-40.
7. di Guan C, Li P, Riggs PD, Inouye H: **Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein.** *Gene* 1988, **67**(1):21-30.
8. Yasukawa T, Kanei-Ishii C, Maekawa T, Fujimoto J, Yamamoto T, Ishii S: **Increase of solubility of foreign proteins in Escherichia coli by coproduction of the bacterial thioredoxin.** *J Biol Chem* 1995, **270**(43):25328-25331.
9. Rajan SS, Lackland H, Stein S, Denhardt DT: **Presence of an N-terminal polyhistidine tag facilitates stable expression of an otherwise unstable N-terminal domain of mouse tissue inhibitor of metalloproteinase-1 in Escherichia coli.** *Protein Expr Purif* 1998, **13**(1):67-72.
10. Tang W, Sun ZY, Pannell R, Gurewich V, Liu JN: **An efficient system for production of recombinant urokinase-type plasminogen activator.** *Protein Expr Purif* 1997, **11**(3):279-283.
11. Chen H, Xu Z, Xu N, Cen P: **Efficient production of a soluble fusion protein containing human beta-defensin-2 in E. coli cell-free system.** *J Biotechnol* 2005, **115**(3):307-315.
12. Chant A, Kraemer-Pecore CM, Watkin R, Kneale GG: **Attachment of a histidine tag to the minimal zinc finger protein of the Aspergillus nidulans gene regulatory protein AreA causes a conformational change at the DNA-binding site.** *Protein Expr Purif* 2005, **39**(2):152-159.
13. Goel A, Colcher D, Koo JS, Booth BJ, Pavlinkova G, Batra SK: **Relative position of the hexahistidine tag effects binding properties of a tumor-associated single-chain Fv construct.** *Biochim Biophys Acta* 2000, **1523**(1):13-20.
14. Fonda I, Kenig M, Gaberc-Porekar V, Pristovaek P, Menart V: **Attachment of histidine tags to recombinant tumor necrosis factor-alpha drastically changes its properties.** *ScientificWorldJournal* 2002, **2**:1312-1325.
15. Cadel S, Gouzy-Darmon C, Petres S, Piesse C, Pham VL, Beinfeld MC, Cohen P, Foulon T: **Expression and purification of rat recombinant aminopeptidase B secreted from baculovirus-infected insect cells.** *Protein Expr Purif* 2004, **36**(1):19-30.

16. Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B: **Crystal structures of fusion proteins with large-affinity tags.** *Protein Sci* 2003, **12**(7):1313-1322.
17. de Vries EG, de Hooge MN, Gietema JA, de Jong S: **Correspondence re: C. G. Ferreira et al., Apoptosis: target of cancer therapy.** *Clin. Cancer Res.*, **8**: 2024-2034, 2002. *Clin Cancer Res* 2003, **9**(2):912; author reply 913.
18. Arnau J, Lauritzen C, Petersen GE, Pedersen J: **Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins.** *Protein Expr Purif* 2006, **48**(1):1-13.
19. Zhao DX, Ding ZC, Liu YQ, Huang ZX: **Overexpression and purification of single zinc finger peptides of human zinc finger protein ZNF191.** *Protein Expr Purif* 2007, **53**(1):232-237.
20. Doubeikovskaia Z, Aries A, Jeannesson P, Morle F, Doubeikovski A: **Purification of human recombinant GATA-1 from bacteria: implication for protein-protein interaction studies.** *Protein Expr Purif* 2001, **23**(3):426-431.
21. Jiang L, Zhang B, Wang G, Wang K, Xiao X: **Expression, purification and characterization of rat zinc finger protein Mipu1 in Escherichia coli.** *Mol Cell Biochem* 2009, **328**(1-2):137-144.
22. Rudert F, Visser E, Gradl G, Grandison P, Shemshedini L, Wang Y, Grierson A, Watson J: **pLEF, a novel vector for expression of glutathione S-transferase fusion proteins in mammalian cells.** *Gene* 1996, **169**(2):281-282.
23. Terpe K: **Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems.** *Appl Microbiol Biotechnol* 2003, **60**(5):523-533.
24. Gervais V, Busso D, Wasielewski E, Poterszman A, Egly JM, Thierry JC, Kieffer B: **Solution structure of the N-terminal domain of the human TFIIH MAT1 subunit: new insights into the RING finger family.** *J Biol Chem* 2001, **276**(10):7457-7464.
25. Nomine Y, Charbonnier S, Miguet L, Potier N, Van Dorsselaer A, Atkinson RA, Trave G, Kieffer B: **<sup>1</sup>H and <sup>15</sup>N resonance assignment, secondary structure and dynamic behaviour of the C-terminal domain of human papillomavirus oncoprotein E6.** *J Biomol NMR* 2005, **31**(2):129-141.
26. Nomine Y, Ristriani T, Laurent C, Lefevre JF, Weiss E, Trave G: **Formation of soluble inclusion bodies by hpv e6 oncoprotein fused to maltose-binding protein.** *Protein Expr Purif* 2001, **23**(1):22-32.
27. Yao J, Liu Y, Liang HG, Zhang C, Zhu JZ, Qin X, Sun M, Qu SS, Yu ZN: **The effect of zinc(II) on the growth of E-coli studied by microcalorimetry.** *Journal of Thermal Analysis and Calorimetry* 2005, **79**(1):39-43.

28. Zanier K, Nomine Y, Charbonnier S, Ruhlmann C, Schultz P, Schweizer J, Trave G: **Formation of well-defined soluble aggregates upon fusion to MBP is a generic property of E6 proteins from various human papillomavirus species.** *Protein Expr Purif* 2007, **51**(1):59-70.
29. Elharakany AA, Halim FMA, Barakat AO: **Dissociation-Constants and Related Thermodynamic Quantities of the Protonated Acid Form of Tris-(Hydroxymethyl)-Aminomethane in Mixtures of 2-Methoxyethanol and Water at Different Temperatures.** *Journal of Electroanalytical Chemistry* 1984, **162**(1-2):285-305.
30. Nurmemmedov E, Thunnissen M: **Expression, purification, and characterization of the 4 zinc finger region of human tumor suppressor WT1.** *Protein Expr Purif* 2006, **46**(2):379-389.
31. Rouzina I, Pant K, Karpel RL, Williams MC: **Theory of electrostatically regulated binding of T4 gene 32 protein to single- and double-stranded DNA.** *Biophys J* 2005, **89**(3):1941-1956.
32. Lundback T, van Den Berg S, Hard T: **Sequence-specific DNA binding by the glucocorticoid receptor DNA-binding domain is linked to a salt-dependent histidine protonation.** *Biochemistry* 2000, **39**(30):8909-8916.
33. Xu MQ, Paulus H, Chong S: **Fusions to self-splicing inteins for protein purification.** *Methods Enzymol* 2000, **326**:376-418.
34. Chang JY, Alkan SS, Hilschmann N, Braun DG: **Thrombin specificity. Selective cleavage of antibody light chains at the joints of variable with joining regions and joining with constant regions.** *Eur J Biochem* 1985, **151**(2):225-230.
35. Rodriguez PL, Carrasco L: **Improved factor Xa cleavage of fusion proteins containing maltose binding protein.** *Biotechniques* 1995, **18**(2):238, 241-233.
36. Dougherty WG, Parks TD, Cary SM, Bazan JF, Fletterick RJ: **Characterization of the catalytic residues of the tobacco etch virus 49-kDa proteinase.** *Virology* 1989, **172**(1):302-310.
37. Nagai K, Thogersen HC: **Generation of beta-globin by sequence-specific proteolysis of a hybrid protein produced in Escherichia coli.** *Nature* 1984, **309**(5971):810-812.
38. Dougherty WG, Carrington JC, Cary SM, Parks TD: **Biochemical and mutational analysis of a plant virus polyprotein cleavage site.** *EMBO J* 1988, **7**(5):1281-1287.
39. Phan J, Zdanov A, Evdokimov AG, Tropea JE, Peters HK, 3rd, Kapust RB, Li M, Wlodawer A, Waugh DS: **Structural basis for the substrate specificity of tobacco etch virus protease.** *J Biol Chem* 2002, **277**(52):50564-50572.



40. Mohanty AK, Simmons CR, Wiener MC: **Inhibition of tobacco etch virus protease activity by detergents.** *Protein Expr Purif* 2003, **27**(1):109-114.
41. Tubb MR, Smith LE, Davidson WS: **Purification of recombinant apolipoproteins A-I and A-IV and efficient affinity tag cleavage by tobacco etch virus protease.** *J Lipid Res* 2009, **50**(7):1497-1504.
42. Fang L, Jia KZ, Tang YL, Ma DY, Yu M, Hua ZC: **An improved strategy for high-level production of TEV protease in Escherichia coli and its purification and characterization.** *Protein Expr Purif* 2007, **51**(1):102-109.
43. Liew OW, Ching Chong JP, Yandle TG, Brennan SO: **Preparation of recombinant thioredoxin fused N-terminal proCNP: Analysis of enterokinase cleavage products reveals new enterokinase cleavage sites.** *Protein Expr Purif* 2005, **41**(2):332-340.
44. Chang JY: **Thrombin specificity. Requirement for apolar amino acids adjacent to the thrombin cleavage site of polypeptide substrate.** *Eur J Biochem* 1985, **151**(2):217-224.
45. Lichty JJ, Malecki JL, Agnew HD, Michelson-Horowitz DJ, Tan S: **Comparison of affinity tags for protein purification.** *Protein Expr Purif* 2005, **41**(1):98-105.
46. Nallamsetty S, Kapust RB, Tozser J, Cherry S, Tropea JE, Copeland TD, Waugh DS: **Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease in vivo and in vitro.** *Protein Expr Purif* 2004, **38**(1):108-115.
47. Eliseev R, Alexandrov A, Gunter T: **High-yield expression and purification of p18 form of Bax as an MBP-fusion protein.** *Protein Expr Purif* 2004, **35**(2):206-209.
48. Rubio V, Shen Y, Saijo Y, Liu Y, Gusmaroli G, Dinesh-Kumar SP, Deng XW: **An alternative tandem affinity purification strategy applied to Arabidopsis protein complex isolation.** *Plant J* 2005, **41**(5):767-778.
49. Mao H: **A self-cleavable sortase fusion for one-step purification of free recombinant proteins.** *Protein Expr Purif* 2004, **37**(1):253-263.
50. Lorentsen RH, Fynbo CH, Thogersen HC, Etzerodt M, Holtet TL: **Expression, refolding, and purification of recombinant human granzyme B.** *Protein Expr Purif* 2005, **39**(1):18-26.
51. Liu XQ: **Protein-splicing intein: Genetic mobility, origin, and evolution.** *Annu Rev Genet* 2000, **34**:61-76.
52. Humphries HE, Christodoulides M, Heckels JE: **Expression of the class 1 outer-membrane protein of Neisseria meningitidis in Escherichia coli and purification using a self-cleavable affinity tag.** *Protein Expr Purif* 2002, **26**(2):243-248.
53. Butt TR, Edavettal SC, Hall JP, Mattern MR: **SUMO fusion technology for difficult-to-express proteins.** *Protein Expr Purif* 2005, **43**(1):1-9.

54. Abdullah N, Chase HA: **Removal of poly-histidine fusion tags from recombinant proteins purified by expanded bed adsorption.** *Biotechnol Bioeng* 2005, **92**(4):501-513.
55. Wagner FW, Wilkes SH, Prescott JM: **Specificity of *Aeromonas* aminopeptidase toward amino acid amides and dipeptides.** *J Biol Chem* 1972, **247**(4):1208-1210.
56. Kenig M, Peternel S, Gaberc-Porekar V, Menart V: **Influence of the protein oligomericity on final yield after affinity tag removal in purification of recombinant proteins.** *J Chromatogr A* 2006, **1101**(1-2):293-306.
57. Westra DF, Welling GW, Koedijk DG, Scheffer AJ, The TH, Welling-Wester S: **Immobilised metal-ion affinity chromatography purification of histidine-tagged recombinant proteins: a wash step with a low concentration of EDTA.** *J Chromatogr B Biomed Sci Appl* 2001, **760**(1):129-136.





## Chapter 3: Metal exchange experiments of zinc finger proteins

### 1. Introduction

#### 1.1 Investigation of zinc binding sites by metal exchange experiments

Metal-substitution experiment may be used not only to investigate metal binding sites of proteins but also to evaluate protein quality after protein purification. For zinc finger proteins, in particular, investigation of zinc binding sites is absolutely critical before structure determination. The zinc binding properties of zinc fingers provide a convenient way to probe the structural integrity using spectroscopic properties. This approach was used in early work on metal binding proteins by substitution of zinc by a cobalt ion. Whereas zinc ( $d^{10}$ ) has no specific absorption properties,  $Co^{2+}$  ( $d^7$ ) displays absorption in the visible spectrum ( $\lambda=550\sim 880\text{nm}$ ) allowing the easy detection of the metal coordinating species. This approach was also applied widely in the characterization of structural zinc binding domains in various proteins including zinc finger proteins [1, 2] although the binding affinities are 3-4 orders of magnitude lower for cobalt than for zinc [3]. In NMR studies, metal substitution experiments are usually performed using cadmium. Among the eight natural isotopes of cadmium, only two ( $^{111}\text{Cd}$  and  $^{113}\text{Cd}$ ) possess spin 1/2 which provide interesting probes. The affinity of metal ions for cysteine sulfur atoms increase in the following order:  $Zn^{2+} < Cd^{2+} < Hg^{2+}$ . Therefore,  $Cd^{2+}$  may compete with and displace  $Zn^{2+}$  in kinetically labile cysteine-containing metal sites such as those found in zinc finger proteins.[4, 5] Cadmium substitution induces chemical shift changes that can be easily monitored using  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra. In particular, amide protons involved in hydrogen bonds to the cysteine sulfur atoms are the most sensitive to metal substitution. This effect can be used to identify residues which are in proximity to the metal binding sphere. More sophisticated and precise information may be obtained using NMR active cadmium nuclei such as  $^{113}\text{Cd}$ . Its spin 1/2 allows the design of correlation experiments between the  $Cd^{2+}$  ion and the  $H\beta$  protons of the coordinating cysteine residues. Furthermore, the chemical shift of the cadmium is sensitive to the nature and the geometry of the binding site, providing useful information, although this kind of study is impaired by the lack of either theoretical or experimental data allowing interpretation of the observed cadmium frequencies.

## 1.2 Kinetics of metal exchange

For metal-containing proteins, metal exchange from a binding site may affect protein function where this is related to the intrinsic properties of metal ions.[6] The measurement of exchange rates provides an additional source of information on zinc binding proteins. Indeed, this exchange rate depends on several factors intimately linked to the nature of the protein such as the fraction of an "open state" allowing the exchange to take place or the charge distribution at the protein surface. The former feature is related to the dynamic behavior and the plasticity of the protein. Thus, while it remains difficult to relate a measured rate to a specific feature of the protein, the comparison of exchange rates measured for a protein and the same protein affected by a slight perturbation, such as a single point mutation, allows the detection of subtle changes in the protein's dynamics. While possible functional aspects of metal coordination in proteins are often ignored, it is expected that such comparisons should shed light on the role of metal coordination in some biological systems. In some specific systems, such as for proteins involved in zinc homeostasis, metal exchange properties are directly related to the protein function [7], but for many others, the possible functional role of metal exchange properties is unknown.

## 2. Materials and methods

### 2.1 Replacement of Zn<sup>2+</sup> by Cd<sup>2+</sup>

#### 2.1.1 Materials

Protein	ARDBD wild type/T575A mutant	ATXN7 and ATXN7 L3
Sample preparation	For WT, 3.5µl of 20mM CdCl <sub>2</sub> -EDTA was added to Zn <sup>2+</sup> -ARDBD WT. For T575A mutant, 2.5µl of 20mM CdCl <sub>2</sub> -EDTA was added to Zn <sup>2+</sup> -ARDBD T575A.	For ATXN7, 6.75µl of 20mM CdCl <sub>2</sub> -EDTA was added to Zn <sup>2+</sup> -ATXN7. For ATXN7L3, 2µl of 20mM CdCl <sub>2</sub> -EDTA was added to Zn <sup>2+</sup> -ATXN7L3.
Sample concentration	WT: 80µM; T575A mutant : 60µM	ATXN7 : 150µM; ATXN7 L3 : 60µM
Experiment temperature	303K	298K

#### 2.1.2 Methods

All exchanges were initiated by addition of a three-fold excess of Cd<sup>2+</sup>-EDTA in protein samples on an ice bath. This ensured that the metal exchange occurs smoothly and allows the kinetics to be followed by NMR. A series of <sup>1</sup>H-<sup>15</sup>N HSQC spectra were recorded every 20 minutes at 303K (AR-DBD) and 298K (ATXN7 and ATXN7 L3), respectively. Since the metal exchange is slow on the NMR time scale, the exchange leads to the gradual appearance of a new set of correlation peaks corresponding to the

$\text{Cd}^{2+}$  loaded proteins. As the relative intensities of the Cd- and Zn- specific peaks are proportional to the populations of the two complexes, the metal exchange kinetics can be followed by plotting peak intensities as a function of experimental time. Due to the slow exchange regime, the assignments of correlations in the cadmium-loaded protein may not be derived directly from those of the zinc-loaded protein and a  $^{15}\text{N}$ -edited NOESY spectrum is necessary to confirm sequential assignments. The chemical shift difference between cadmium-loaded and zinc-loaded proteins is calculated according to:

$$\Delta\delta_{(\text{Zn,Cd})} = \sqrt{10 * (\delta_{\text{HN}(\text{Cd})} - \delta_{\text{HN}(\text{Zn})})^2 + (\delta_{\text{N}(\text{Cd})} - \delta_{\text{N}(\text{Zn})})^2}$$

where  $\delta_{\text{HN}}$  and  $\delta_{\text{N}}$  indicate amide proton and nitrogen chemical shift, respectively.

## 2.2 Measurements of kinetics

To detect rapid reactions, samples were mixed rapidly within the NMR sample tube and a series of time-dependent NMR experiments were recorded. Peak intensities in spectra were measured by CARA and exponential decay rates were obtained from a non-linear, least-squares fit using the Levenberg–Marquardt algorithm implemented in matlab.  $\text{Zn}^{2+}$ – $\text{Cd}^{2+}$  exchange kinetics were interpreted using the following equations:

$$I_{\text{Zn}}(t) = Ae^{-R_{\text{exp}} t};$$

$$I_{\text{Cd}}(t) = A(1 - e^{-R_{\text{exp}} t});$$

where  $R_{\text{exp}}$  and A (two parameters) are fitted.

In addition, Monte Carlo simulations were used to estimate the statistical error on the parameters.

## 2.3 $^{15}\text{N}$ relaxation rate measurement for AR-DBD WT and T575A mutant

Only  $R_2$  relaxation rate measurements were recorded. The relaxation-dependent magnetization decay was sampled at 5, 10, 20,40,60,80,100,120 and 140 ms. For error estimation, the spectrum with a relaxation delays of 40ms were recorded twice. Peak intensities were measured by CARA and exponential decay rates were obtained from a non-linear, least-squares fit using the Levenberg–Marquardt algorithm implemented in matlab.

### 3. Results and discussion

#### 3.1 Metal exchange experiments for AR-DBD wild type and T575A mutant

The exchange of coordinated  $Zn^{2+}$  ions for  $Cd^{2+}$  was measured by recording a series of  $^1H$ - $^{15}N$  HSQC spectra after the addition of an excess of  $Cd^{2+}$ -EDTA to the protein sample. Fig.1 shows the superposition of  $^1H$ - $^{15}N$  HSQC spectra of AR-DBD wild type (WT) bound to zinc and 3 hours after addition of  $CdCl_2$ -EDTA. Upon replacement of  $Zn^{2+}$  by  $Cd^{2+}$ , the signals of several amide groups in close proximity to the metal-binding sites, such as residue C595 and C614, disappeared and reappeared at new positions, whereas residues such as C559 and C576 were unaffected (Fig.1). The chemical shift changes for the amide group of each residue are plotted in Fig.1B. 73 of 77 amide groups for residues were identified in HSQC spectra and retained for further analysis. Seven of the eight cysteine residues involved in zinc coordination were observed in the spectra but no peaks could be detected for C611 in the spectrum of the Cd-bound form. Most amide groups show small chemical shift changes upon replacement of  $Zn^{2+}$  by  $Cd^{2+}$ . The largest chemical shift changes were observed for amide groups in close proximity to one of the sulphur atoms ( $\leq 3\text{\AA}$ ) involved in metal coordination, such as C562, C576, C579 and C614 (Fig.2A) that form N-H $\cdots$ S hydrogen bonds.

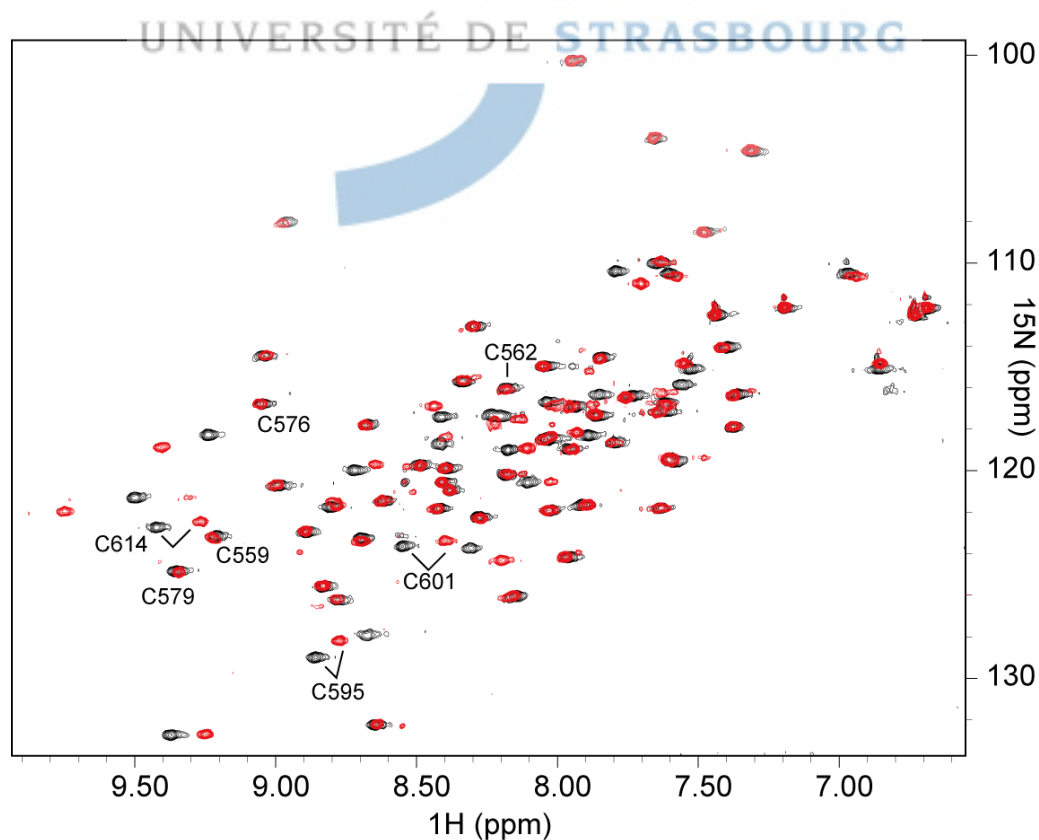


Fig.1 Superposition of  $^1H$ - $^{15}N$  HSQC spectra of AR-DBD WT bound to zinc (black) and cadmium (red).

The chemical shift changes that occur upon metal ion substitution for WT and T575A mutant domains are shown in Fig.2A. The difference in values for WT and T575A domain are shown in Fig.2B and are mapped to the structure of the WT domain using a color code (Fig.3A). Residue S578 showed the largest difference in chemical shift change between WT and T575A mutant domains (indicated by a star in Fig.2A). Residues T558, N561, C569 and C614 also revealed larger chemical shift changes in WT (Fig.2A and 2B). S578 lies at the head of the first  $\alpha$ -helix and adjacent to C579 which is one of the zinc binding ligands (Fig.3A). In particular, C569 which is not bound to zinc also shows larger chemical shift changes in WT domain (Fig.3A) than in the T575A mutant domain. On the other hand, C562, I603 and D604 have larger chemical shift changes in T575A mutant domain (Fig.2B) than those in WT domain. In the three-dimensional structure of the domain, C562 and D604 are located in loop regions, indicating a close contact with each other (Fig.3A). Residues in the second  $\alpha$ -helix show small fluctuations in chemical shift changes between WT and T575A mutant (Fig.3B).



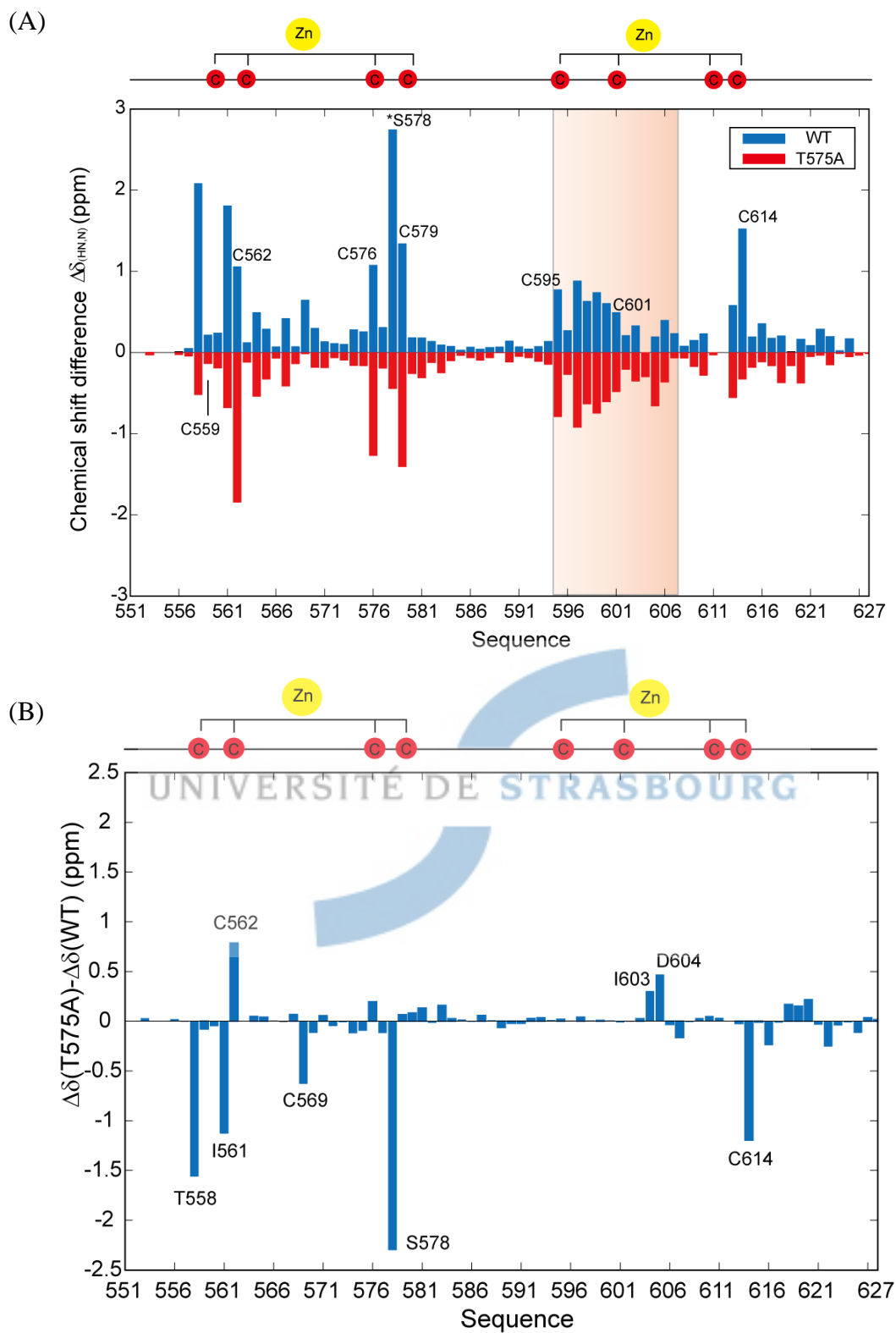
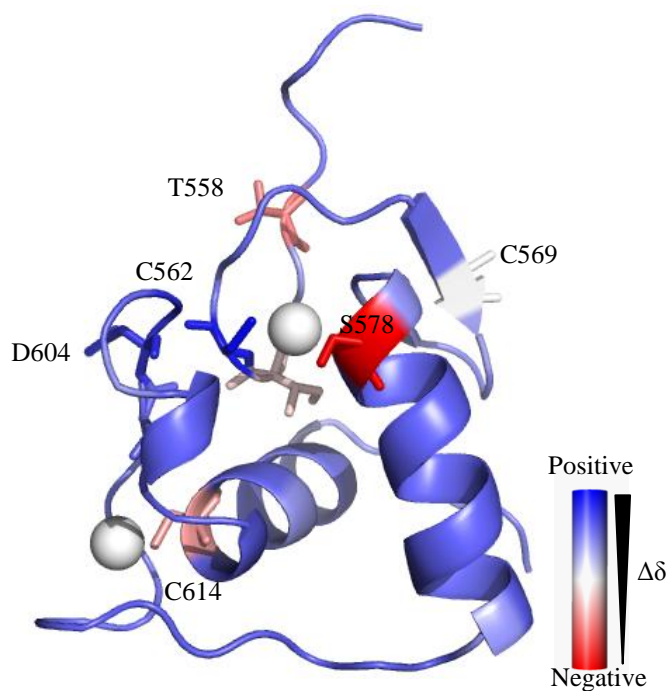


Fig.2 (A) Plot of chemical shift changes of AR-DBD wild type (WT) and T575A mutant bound to zinc and cadmium. (B) The differences in chemical shift changes between AR-DBD WT and T575A mutant upon metal exchange. Values were derived by subtracting the values in plot A. The topology of zinc coordination is showed at the top of each plot.

(A)



(B)

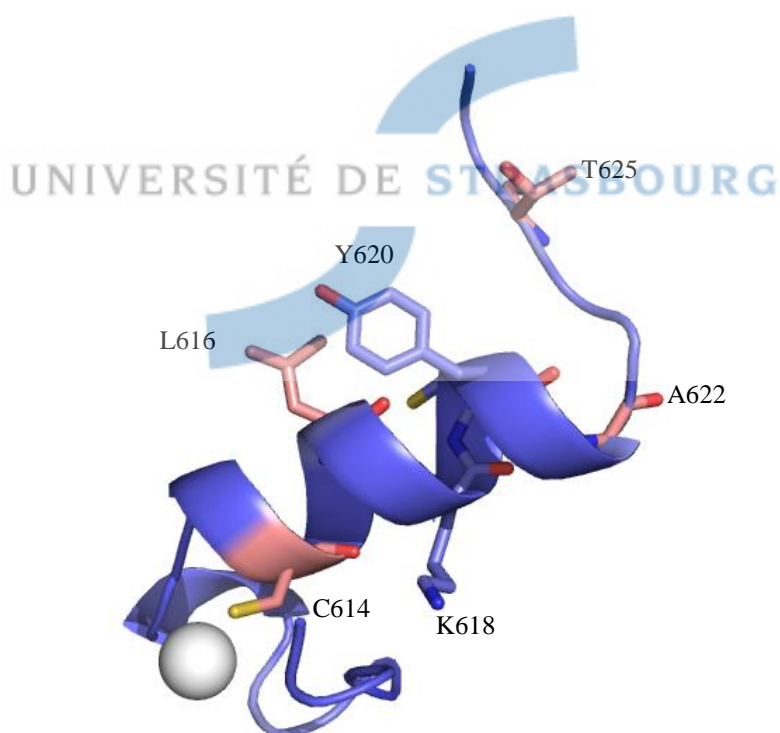


Fig.3 (A) The backbone structure of AR-DBD WT colored by the difference in chemical shift changes between WT and T575A mutant domains upon metal ion substitution in Fig.2B. Residues with increased negative values are shown as blue sticks on the ribbon structure. (B) Ribbon representation of the C-terminal region of AR-DBD WT. Several residues in the second  $\alpha$ -helix including C614, L616, A622 and T625 show larger chemical shift changes in WT (pink sticks) than those in T575A mutant domain. On the contrary, residues K618 and Y620 show larger chemical shift changes in the T575A mutant domain (blue sticks) than in WT domain.



### 3.1.2 Exchange kinetics

The chemical shift changes that occur upon metal ion substitution allow the effects of metal exchange to be followed at the level of individual residues and thus the metal exchange kinetics for the two metal-binding sites can be extracted. However, the exchange kinetics must be slow enough in order to be followed by NMR. For AR-DBD, the two zinc binding sites have distinct zinc exchange rates, but the zinc ion in the second zinc binding site exchanges too rapidly to be followed in our experimental set-up (Fig.4). As a result, only the kinetics of metal ion exchange for the first zinc binding site could be measured. Fig.5B and C show how rates could be extracted from the intensities of residues adjacent to the first site in the metal exchange process. In the T575A mutant, the zinc ion in the first zinc binding site exchanged more quickly than in WT by a factor of two (Table 1). This kinetic difference between WT and mutant domains may be related to different dynamic properties of the metal binding sites in the two domains. Hence, the measurement of  $^{15}\text{N}$  relaxation rates for the two proteins in the presence of  $\text{Cd}^{2+}$  was undertaken.

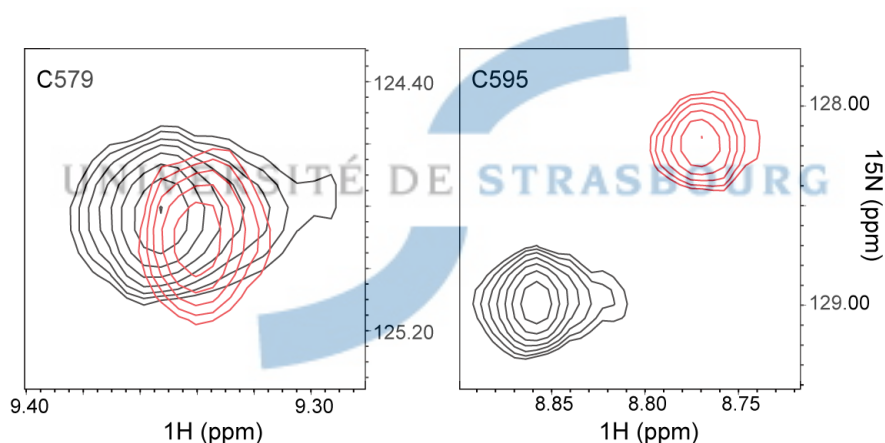
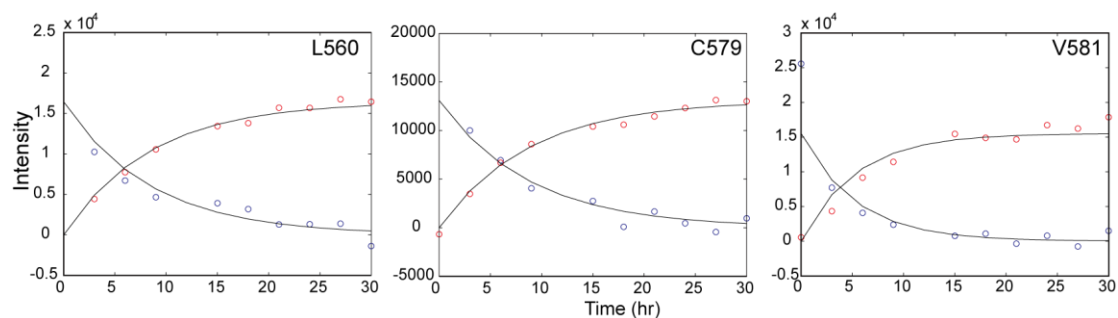


Fig.4 Two peaks C579 (in the first zinc binding site ) and C595 ( in the second zinc binding site ) were selected from the superposition of  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of AR-DBD WT bound to  $\text{Zn}^{2+}$  (black) and after replacement of  $\text{Zn}^{2+}$  by  $\text{Cd}^{2+}$  (red).

## (A) WT



## (B) T575A

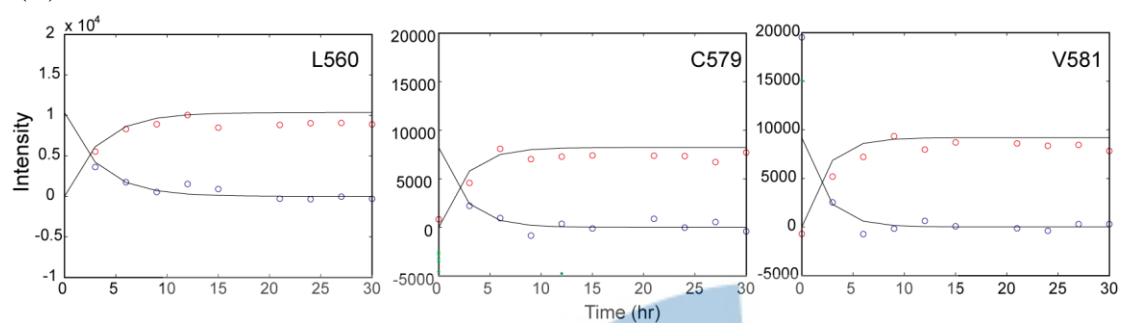


Fig.5 Decay of intensities of Zn<sup>2+</sup>-AR-DBD and growth of intensities of Cd<sup>2+</sup>-AR-DBD for WT (A) and T575A mutant (B) domains.

UNIVERSITÉ DE STRASBOURG

Table 1 Metal exchange rates for AR-DBD WT and T575A mutant derived from increasing and decreasing cross-peaks in HSQC spectra upon exchange of Zn<sup>2+</sup> by Cd<sup>2+</sup>

	WT	T575A mutant
	Rate ( $10^{-5} \text{ s}^{-1}$ )	Rate ( $10^{-5} \text{ s}^{-1}$ )
L560	$3.2 \pm 0.2$	$9.2 \pm 0.5$
C579	$5.1 \pm 0.5$	$12.3 \pm 1$
V581	$5.0 \pm 0.6$	$9.7 \pm 1$

### 3.1.3 Dynamic behavior of AR-DBD upon metal-exchange process

The previous results showed that the metal exchange rate for the first zinc binding site of T575A mutant domain is faster than that in WT. To investigate the effect on the dynamic behavior of the WT and T575A mutant domains of replacement of  $Zn^{2+}$  by  $Cd^{2+}$ , a series of  $^{15}N$  transverse relaxation measurements were recorded. Fig.6 shows profiles of  $R_2$  values for AR-DBD WT and T575A bound to  $Zn^{2+}$  and  $Cd^{2+}$ . The  $R_2$  profiles of AR-DBD WT bound to  $Zn^{2+}$  and  $Cd^{2+}$  are quite similar, indicating a lack of dramatic change in protein dynamics. The first zinc binding site does not show differences in AR-DBD bound to  $Zn^{2+}$  and  $Cd^{2+}$ . However, residues T602, I603, N610 and C611 reveal  $R_2$  values higher than those of the baseline values, indicating the contribution of conformational exchange on the  $\mu$ s-ms time scale (blue areas in Fig.6). Residue I603 shows higher  $R_2$  in  $Cd^{2+}$  than that in  $Zn^{2+}$ -ARDBD WT whereas  $R_2$  values of I603 are similar in  $Zn^{2+}$  and  $Cd^{2+}$ -T575A mutant. In addition, increased  $R_2$  values are also observed for D604 in both  $Zn^{2+}$ -WT and T575A mutant. The  $R_2$  value of D604 is higher in  $Zn^{2+}$ -T575A mutant than in the  $Cd^{2+}$ -bound T575A mutant domain whereas in WT the  $R_2$  values are quite similar (pink areas in Fig.6). Another significant difference is that residue K605 revealed a higher  $R_2$  value in  $Zn^{2+}$  than that in  $Cd^{2+}$ -ARDBD T575A mutant. This phenomenon cannot be evaluated in the WT domain because the cross-peak of K605 could not be detected in the  $^1H$ - $^{15}N$  HSQC spectrum of  $Cd^{2+}$ -ARDBD. The differences in exchange contribution upon metal exchange within the second zinc binding site that are observed for WT and mutant proteins may be due to either a change of intermediate time scale dynamics on slight difference in the chemical shifts. Relaxation dispersion experiments may be performed to distinguish between these two hypotheses. These time-consuming experiments have not yet been recorded for these proteins, and therefore, the nature of the change leading to these subtle differences remains to be investigated.

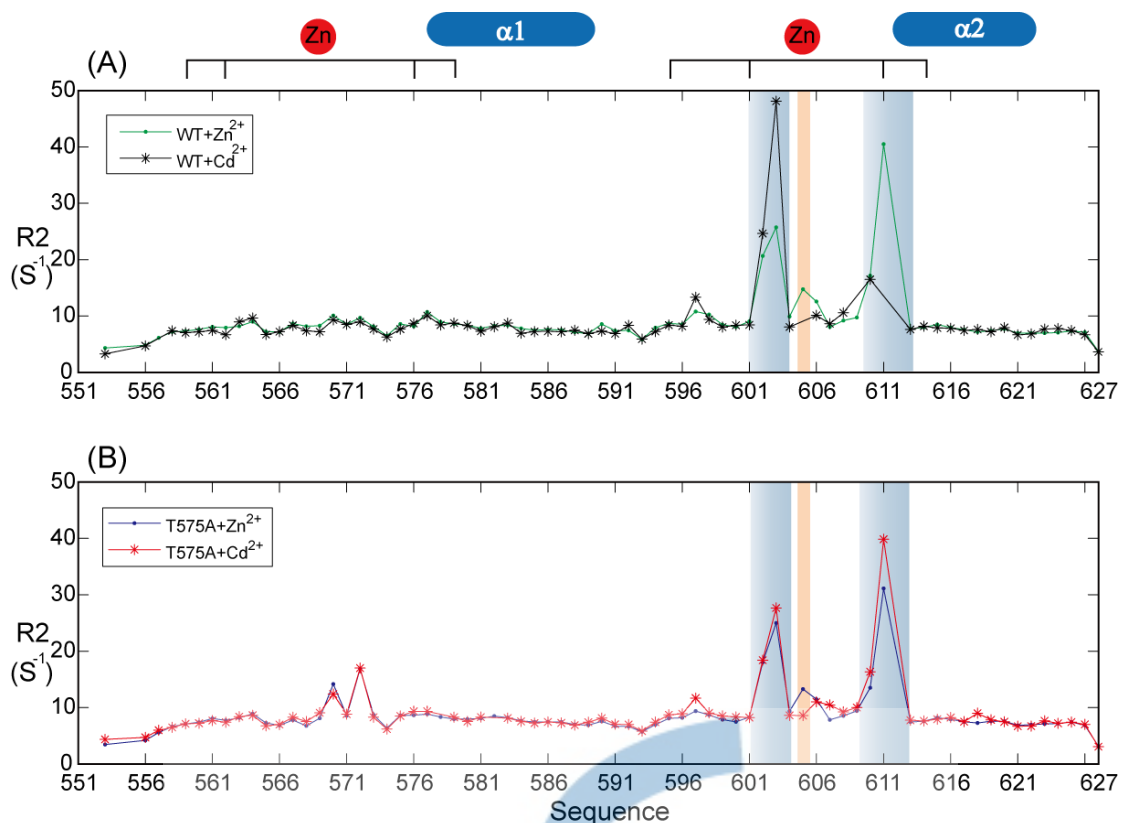


Fig.6  $R_2$  relaxation measurements for AR-DBD (A) WT and (B) T575A mutant bound to  $Zn^{2+}$  and  $Cd^{2+}$ .

### 3.1.4 Conclusions

Investigation of metal binding sites by metal-exchange experiments is a convenient method for the study of zinc finger proteins. NMR spectroscopy provides a sensitive probe for characterizing the metal-binding sites in zinc finger proteins and evaluating conformational changes that occur upon metal exchange. The comparisons of exchange rates for AR-DBD WT and T575A mutant revealed some significant kinetic differences. Most notably, the first zinc binding site of the T575A mutant exchanges more quickly than that in WT. This difference may be related to distinct dynamic behavior of the two proteins. In  $^{15}N$  transverse relaxation measurements, the baselines of the profiles for  $Zn^{2+}$ -AR-DBD and  $Cd^{2+}$ -AR-DBD showed similarity in overall motions in both the WT and T575A mutant domains bound to  $Zn^{2+}$  and  $Cd^{2+}$ . The first zinc binding site does not show significant differences in dynamic properties in  $Zn^{2+}$ - and  $Cd^{2+}$ -AR-DBD whereas the second zinc binding site reveals conformational exchange at the  $\mu$ s-ms time scale in both  $Zn^{2+}$ - and  $Cd^{2+}$ -AR-DBD. To characterize further the difference in conformational exchange terms between  $Zn^{2+}$ - and  $Cd^{2+}$  bound AR-DBD, investigation using relaxation dispersion experiments is required.

## 3.2 Metal exchange experiments on ATXN7 and ATXN7 L3

### 3.2.1 Zn<sup>2+</sup>-Cd<sup>2+</sup> exchange of ATXN7

Fig.7A shows the superposition of the <sup>1</sup>H-<sup>15</sup>N HSQC spectra of ATXN7 bound to zinc and to cadmium. Upon replacement of Zn<sup>2+</sup> by Cd<sup>2+</sup>, the signals of a number of amide groups disappeared and reappeared in new positions. Fig.7B shows the chemical shift changes that occur upon metal ion substitution. 64 of the 72 expected cross-peaks of amide groups could be assigned in the HSQC spectra (Fig.7A). Residues adjacent to the metal binding site such as C347, V349, T359, R360, L362~L369 showed the largest chemical shift changes ( $\Delta\delta > 0.2$ ppm) (Fig.7B shaded region). The largest chemical shift change was observed for T366, indicating the possible formation of a T366 N-H...C364 O hydrogen bond (Fig.11A). In NOESY spectra, the amide proton resonance of T366 has a strong NOE cross-peak to the H $\alpha$  resonance of C364.



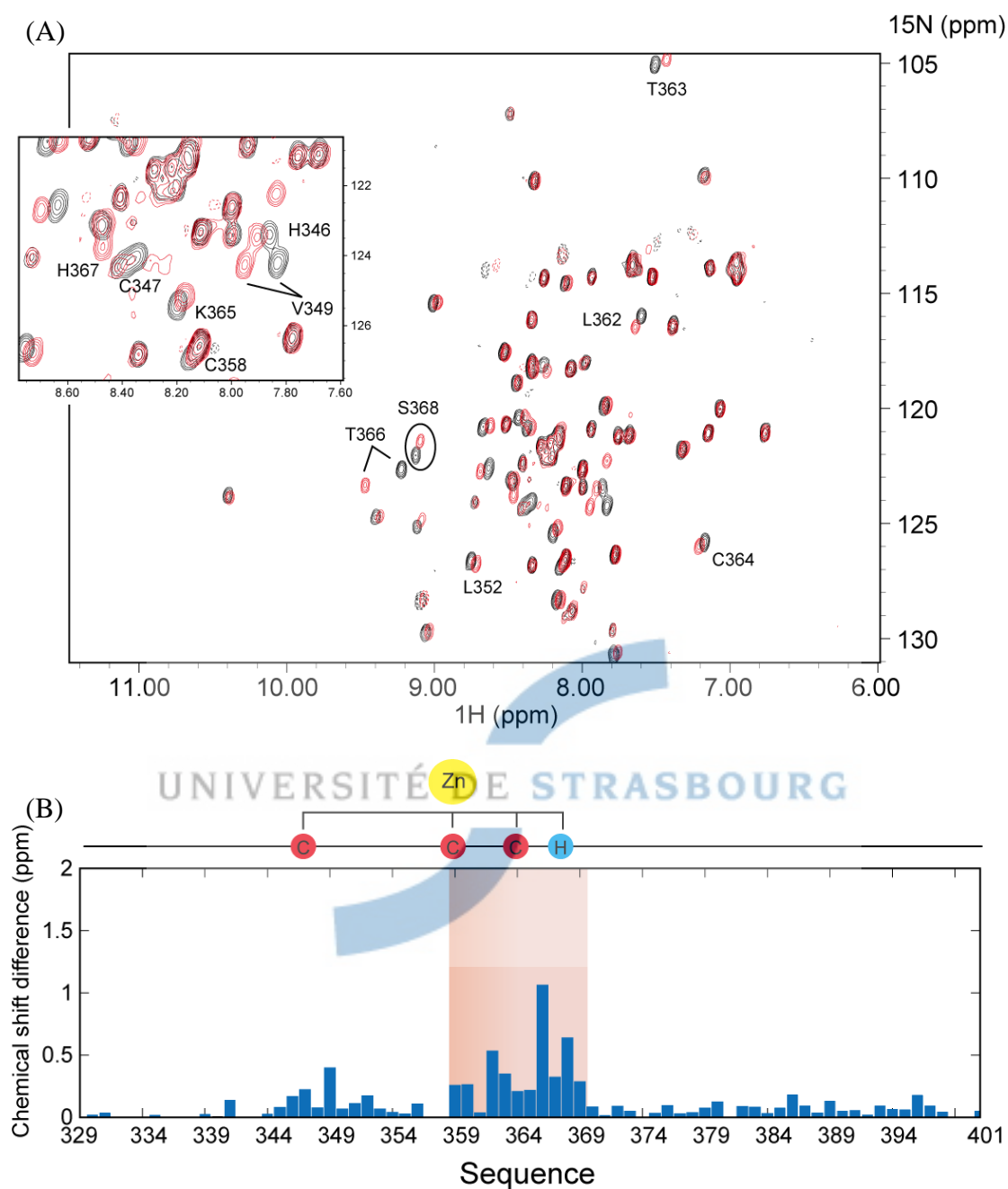


Fig.7 (A)  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of ATXN7 bound to  $\text{Zn}^{2+}$  (black) and  $\text{Cd}^{2+}$  (red). (B) A plot of chemical shift changes in ATXN7 upon replacement of  $\text{Zn}^{2+}$  by  $\text{Cd}^{2+}$ . The zinc binding topology is shown above the figure.

### 3.2.2 $\text{Zn}^{2+}$ - $\text{Cd}^{2+}$ exchange for ATXN7 L3

Fig.8B shows the chemical shift changes that occur upon metal ion substitution in ATXN7 L3. 75 out of 80 peaks were observed in  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra (Fig.8A) and retained for the further analysis. Residues C209, C220, C226 and H229 showed the largest chemical shift changes, reflecting their involvement in zinc coordination.

Several residues adjacent to these zinc coordinating residues such as V211, I212, E214, R222, L224, R225, Q228 and T230, also revealed large shifts ( $\Delta\delta > 0.2$  ppm) (the shaded regions in Fig.8B). In addition, part of the unstructured region (residue 266-271) also revealed a slight perturbation upon metal exchange. In relaxation measurements, these residues display some degree of restriction of their motions with a correlation time quite different (1ns) from that of the core region ( $\sim 5$ ns). This slight effect upon metal exchange suggests that this apparently unstructured part is not completely independent of the structural core and is also affected by events at the metal binding site.

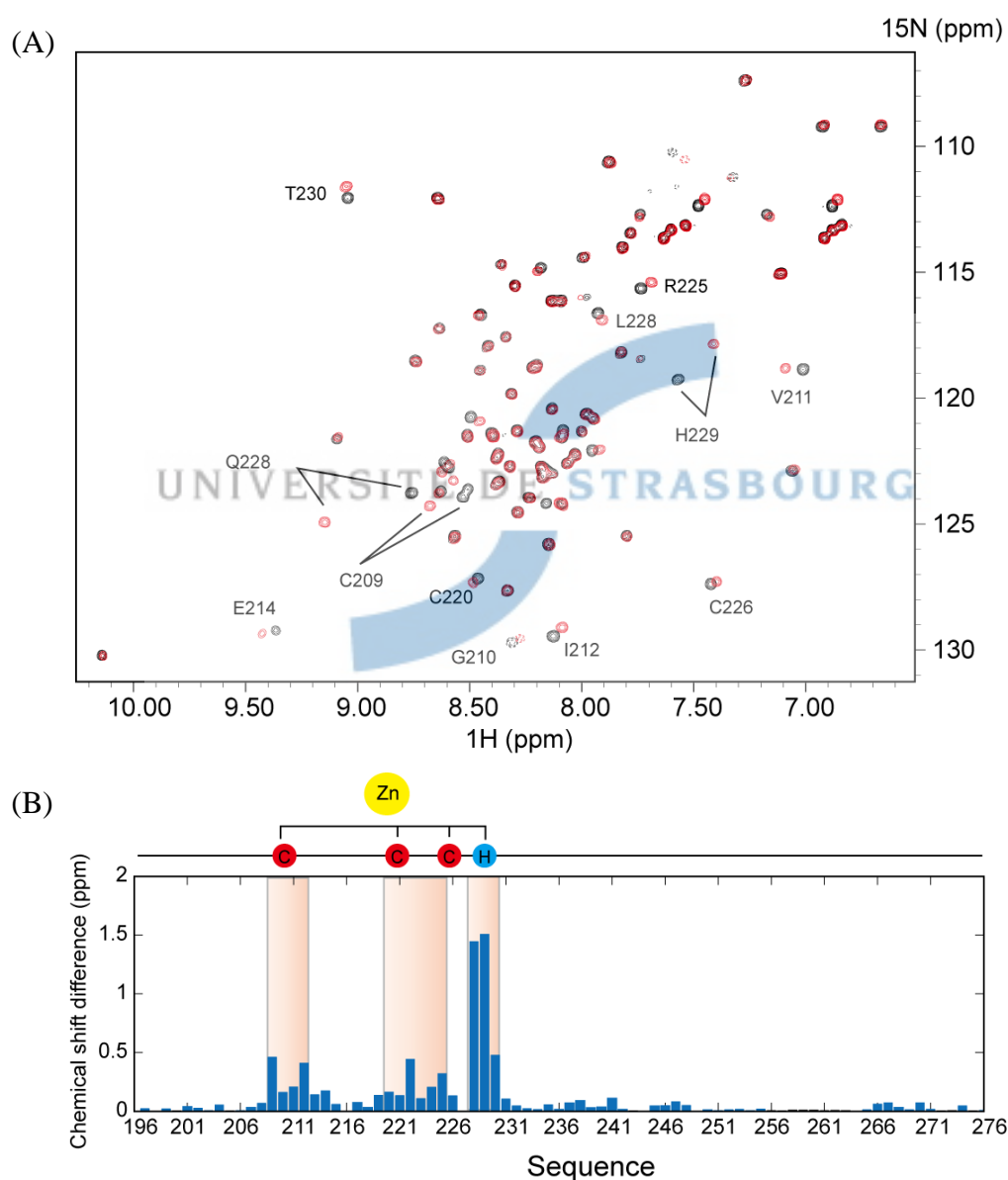


Fig.8 (A)  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of ATXN7 L3 bound to  $\text{Zn}^{2+}$  (black) and replaced by  $\text{Cd}^{2+}$  for 2 hours (red). (B) Chemical shift changes for ATXN7 L3 upon replacement of  $\text{Zn}^{2+}$  by  $\text{Cd}^{2+}$ .



### 3.2.3 Comparison of kinetics for ATXN7 and ATXN7 L3

Since ATXN7 and ATXN7 L3 have similar zinc coordination scaffolds, but different organization of the peripheral helices, we were interested to establish whether this could lead to a difference in zinc exchange kinetics. Experimental conditions for metal exchange for  $\text{Zn}^{2+}$ -ATXN7 and  $\text{Zn}^{2+}$ -ATXN7 L3 were identical. Fig.10 shows representative kinetic curves obtained for ATXN7 and ATXN7 L3. The zinc ion in ATXN7 L3 was found to exchange about 8 times faster than that in ATXN7 (table 2). The primary sequences of ATXN7 and ATXN7 L3 in the vicinity of the zinc coordinating sphere display significant differences between the third cysteine residue and the histidine residue involved in zinc coordination (Fig.11). For ATXN7, the sequence in this loop region is CKTH while for ATXN7 L3 it is CPQH (highlighted by the red box in Fig.9). This sequence difference induces a change in the conformation of the zinc coordination site. T366 in ATXN7 has a strong NOE cross-peak with C364 (Fig.11A) whereas the constrained side chain of P227 in ATXN7 L3 prevents Q228 from being close to C226. The side-chain orientation of Q228, pointing away from the zinc coordination center, (Fig.11B) and the difference in distance between the amide proton and the zinc ion (T366 (4.6Å) and Q228 (5.9Å)), leads to more “open” loop in ATXN7 L3 compared to ATXN7. Interestingly, T366 of ATXN7 and Q228 of ATXN7 L3 show the biggest chemical shift changes upon metal ion substitution. This similarity in the effect of metal exchange but difference in the primary sequence may result in a distinct dynamic behavior between ATXN7 and ATXN7 L3 which requires the further investigation.

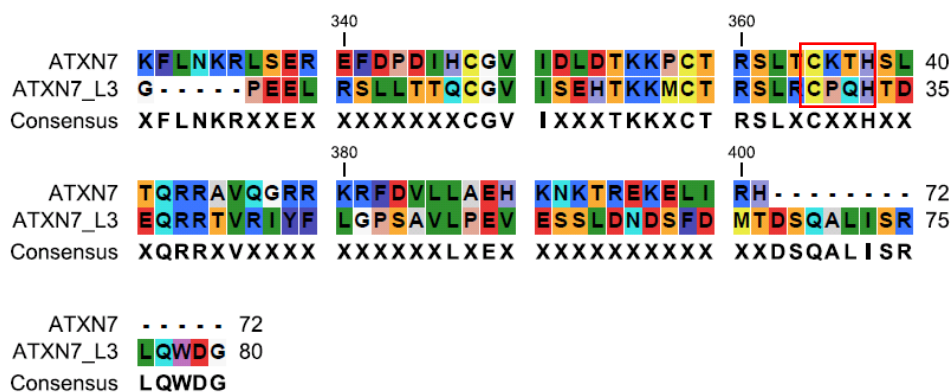


Fig.9 Sequence alignment of ATXN7 and ATXN7 L3.

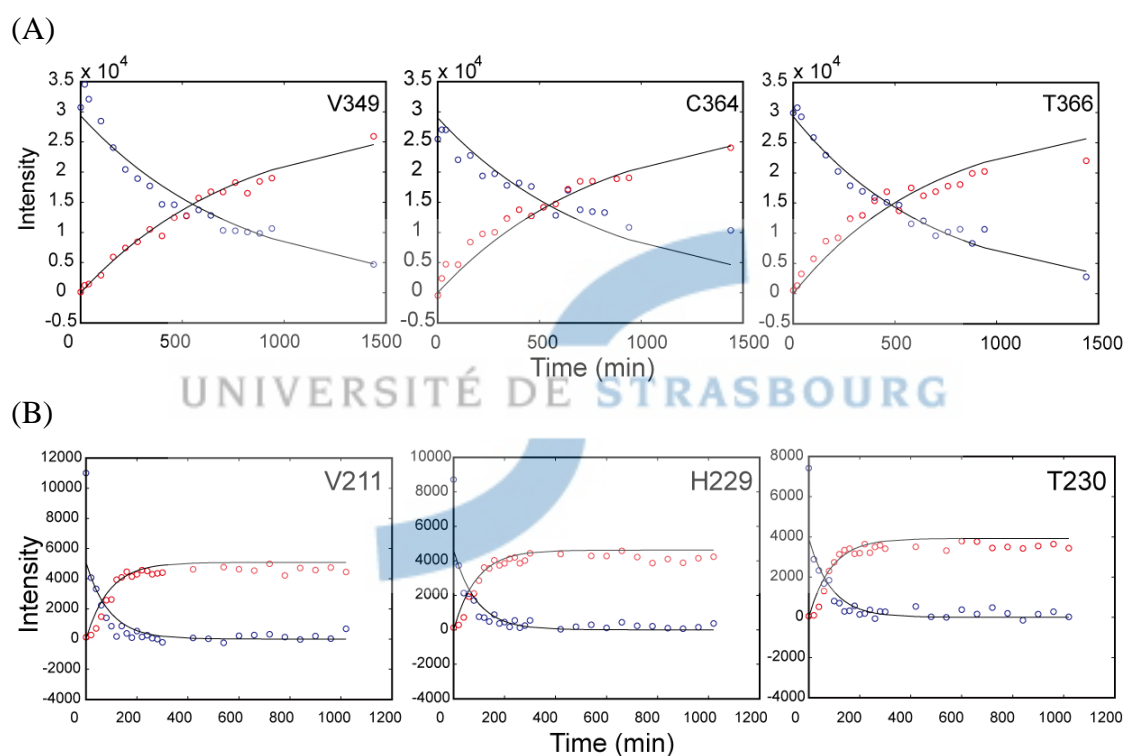


Fig.10 Kinetic curves for (A) ATXN7 and (B) ATXN7 L3. Decay and growth curves extracted from intensities of cross-peaks in ATXN7 and ATXN7 L3 during metal exchange processes.

Table 2 Metal exchange rates for ATXN7 and ATXN7 L3 derived from increasing and decreasing cross-peaks in  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra upon exchange of  $\text{Zn}^{2+}$  by  $\text{Cd}^{2+}$

	ATXN7		ATXN7 L3
	Rate (10 <sup>-5</sup> s <sup>-1</sup> )		Rate (10 <sup>-5</sup> s <sup>-1</sup> )
V349	3.00±0.05	V211	20.0±0.8
C364	2.30±0.04	H229	21.0±0.7
T366	2.50±0.04	T230	19.5±0.8

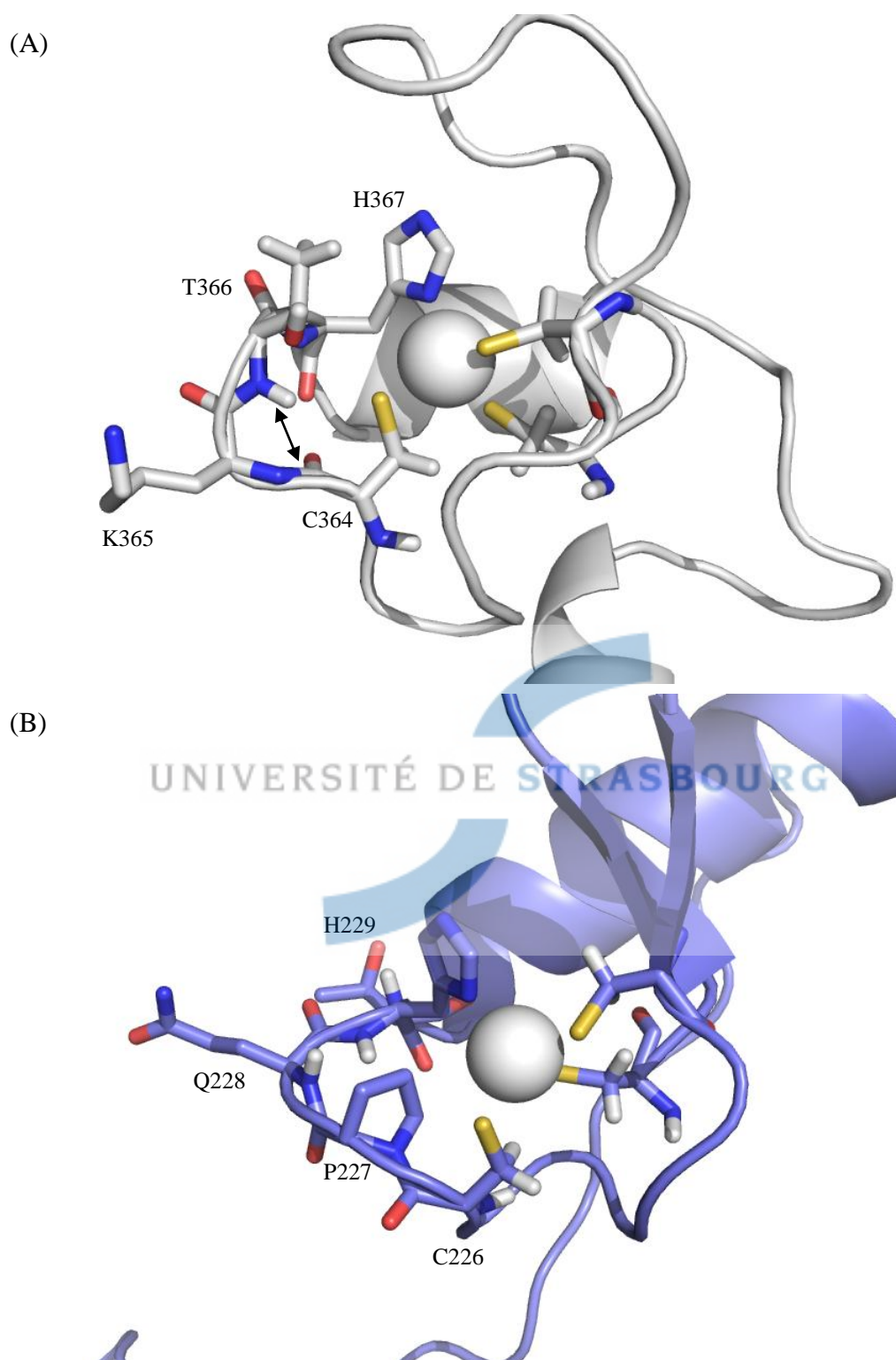


Fig.11 (A) The zinc coordination site of ATXN7. The largest chemical shift change was observed for T366 which forms an N-H...O hydrogen bond with C364. (B) Conformation of the zinc coordination site of ATXN7 L3.

### 3.2.4 Conclusions

The metal exchange rate for ATXN7 L3 is considerably faster than that of ATXN7. Although the chemical shift changes for ATXN7 and ATXN7 L3 upon metal exchange show similarity in the zinc coordination environment, differences in the primary sequence of the zinc binding regions between ATXN7 and ATXN7 L3 result in distinct conformations in the loop region of the zinc coordination sites. This significant difference allows distinct dynamic behavior of Cd<sup>2+</sup>-ATXN7 and ATXN7 L3.

### References

1. Michael SF, Kilfoil VJ, Schmidt MH, Amann BT, Berg JM: **Metal-Binding and Folding Properties of a Minimalist Cys2his2 Zinc Finger Peptide**. *Proceedings of the National Academy of Sciences of the United States of America* 1992, **89**(11):4796-4800.
2. Roehm PC, Berg JM: **Sequential metal binding by the RING finger domain of BRCA1**. *Biochemistry* 1997, **36**(33):10240-10245.
3. Krizek BA, Merkle DL, Berg JM: **Ligand Variation and Metal-Ion Binding-Specificity in Zinc Finger Peptides**. *Inorganic Chemistry* 1993, **32**(6):937-940.
4. Henehan CJ, Pountney DL, Zerbe O, Vasak M: **Identification of Cysteine Ligands in Metalloproteins Using Optical and Nmr-Spectroscopy - Cadmium-Substituted Rubredoxin as a Model [Cd(Cyss)4]2- Center**. *Protein Science* 1993, **2**(10):1756-1764.
5. Houben K, Wasielewski E, Dominguez C, Kellenberger E, Atkinson RA, Timmers HTM, Kieffer B, Boelens R: **Dynamics and metal exchange properties of C4C4 RING domains from CNOT4 and the p44 subunit of TFIIF**. *Journal of Molecular Biology* 2005, **349**(3):621-637.
6. Outten CE, O'Halloran TV: **Femtomolar sensitivity of metalloregulatory proteins controlling zinc homeostasis**. *Science* 2001, **292**(5526):2488-2492.

## Chapter 4: Dynamic studies of AR-DBD and SCA7 domains of ATXN7 and ATXN7 L3

### 1. Introduction

Macromolecular functions are, in many cases, dependent on excursions to excited states and motional behavior on different time scales. Molecular dynamics describe the interconversion rate between different conformational states that are accessible to the molecule. These different conformations are separated by energy barriers of varying heights, leading to large variations of interconversion rates, from picoseconds to seconds. If NMR spectroscopy is unable to determine the number of states that are accessible to a given molecular system, this method allows the very accurate measurement of interconversion rates, with atomic resolution. That is, dynamic information can be obtained at specific sites of a macromolecule providing unique insight into the functional importance of internal and molecular motions. For instance, protein dynamics have been shown to contribute to the stability of functional states and play an important role in catalysis.[1] As a result, NMR studies of protein dynamics bridge the gap between a static protein description (which can be viewed as a ground state description) and its dynamic features (which are related to excited states of the system), enabling a deeper understanding of the protein's mechanisms.[2]

#### 1.1 Protein dynamics time-scales

Fig. 1 shows the characteristic time scales of different types of molecular motions. NMR spectroscopy is characterized by a variety of observables that can be measured on a single system (resonance frequencies, relaxation rates, coupling constant, etc.). These observables are sensitive to motions occurring on different time scales, thus giving access to a wide range of type of motion such as local folding-unfolding equilibria, allosteric transitions or catalytic reactions.[3] One distinguishes generally three different time scales: (i) pico-second to nano-second time scale referred to the Larmor time scale, (ii) millisecond time scale referred as the spectral time-scale and (iii) milliseconds to second time-scale corresponding to the so-called relaxation time-scale. Dynamic process slower than seconds can, of course, be followed in real-time by recording spectra at different times. It should be noted that the lower limit of this third time scale has recently been pushed back by the advent of fast multi-dimensional spectroscopies, allowing 2D spectra to be recorded in a few seconds.[4]

### 1.1.1 Pico- to nano time-scale motions

Sub-nano second time-scale motions are related to the dynamic features of the protein backbone encoded by its amino-acid composition. Loop regions of proteins, and unfolded (or disordered parts) or fully disordered proteins can be viewed as large ensembles of conformational states that interconvert on the pico-second time scale, due to thermal fluctuations. Although some general rules have been established to link the amino-acid composition of the chain to the disordered nature of the protein, or protein segment, considerable research effort is currently underway to decipher the laws governing high frequency molecular motions in proteins. These motions are of primary importance in molecular recognition phenomenon, since disorder-to-order transitions is one of the mechanisms used to imprint specific recognition properties on a protein chain. First identified as a fundamental mechanism for protein-DNA interactions [5], the modulation of fast protein dynamics has since been observed in numerous systems. More and more examples of the involvement of these motions in allosteric communication within one protein domain or between distinct domains have been also reported.[6] Rotational diffusion of proteins occurs through nano-second time-scale motions and their description allows access to information on the molecular shape and its changes upon various events (ligand binding, partial unfolding, oxidative stress, etc. ...). These Larmor time-scale motions induce random transitions between spin states leading to nuclear spin relaxation. Relaxation measurements are therefore widely used to obtain information on these fast motions. It should also be mentioned that, since the global diffusion of the protein leads to a complete loss of correlation of spin states, molecular diffusion represents a "barrier", preventing slower motions (on the nanosecond to microsecond time-scale) to be described. Another difficulty encountered in describing protein motions is the lack of a suitable formalism allowing the possible correlations between motions occurring at different regions of the protein or at slightly different time scales to be taken into account.

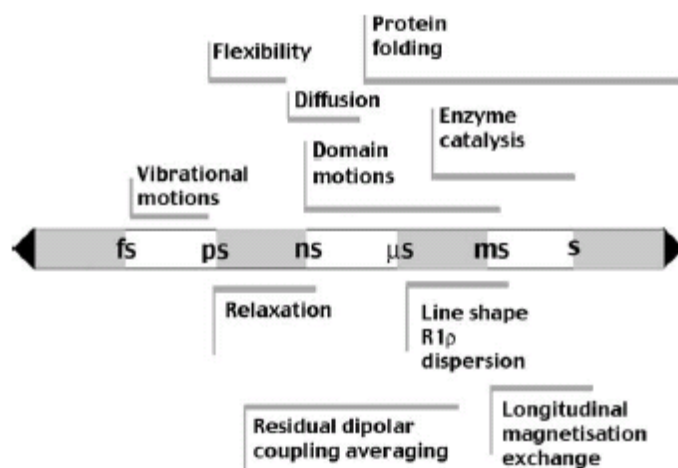


Fig.1 Time scales of molecular motions and NMR parameters.[3]

### 1.1.2 Micro- to millisecond time-scale motions

Motions occurring on this time-scale have a direct impact on the NMR spectra since they interfere with the measurement of resonance frequencies, leading to a broadening of the peaks. Therefore, the occurrence of this type of motion in proteins was reported as soon as the first protein NMR spectra were recorded. These motions, often referred as "conformational exchange" shed light very early on the dynamic nature of proteins, and occur in many biochemical systems such as ligand binding, global or local folding-unfolding processes, oligomerization and self assembly properties or enzymatic reactions. The effect of these motions on recorded spectra depends both on the populations of the distinct states and how the interconversion rates differ from the frequency shifts between the exchanging states. In many cases, the degradation of the spectrum quality induced by these motions has prevented further study. However, the combination of specific biochemical preparation of samples, isotopic labeling and the use of relaxation dispersion methods has recently emerged as a very powerful tool to investigate the functional importance of these motions.[7]

### 1.1.3 Millisecond to second time-scale motions

This time scale represent the upper limit of most motions which have been mentioned above since rate limited steps of enzyme catalysis, off-rate kinetic constants or slow unfolding can occur on this time scale. The difference between this so-called relaxation time scale and the spectral time scale described above is that the frequencies of the individual states of the system are resolved in the spectrum ("slow exchange regime"), allowing their specific observation. The observable of choice to investigate this time scale is longitudinal relaxation measurements since ( $R_1$ , NOE, ...) since the lifetime of longitudinal relaxation increases with molecular weights reaching



hundred of milliseconds and above for large molecular systems. Various experiments exploit this property, many being related to the study of small molecules binding to large macromolecules such as proteins or protein complexes.

## **1.2 General principles of protein motion studies using nuclear spin relaxation**

Various protein motions affect spin relaxation through random modulation of magnetic interactions. Provided that these modulations contain some contributions at characteristic frequencies of the observed spin system, random transitions will be induced and contribute to the evolution of the system towards thermal equilibrium. Among all these interactions, dipole-dipole interactions are the most important, followed by chemical shift anisotropy, whose importance depends on the magnetic field. In proteins, proton spins constitute a complex network of interacting spins, rendering the interpretation of relaxation properties very complex. This is why the measurement of  $^{15}\text{N}$  relaxation parameters was established as a very popular method to probe the dynamic features of the protein backbone.  $^{15}\text{N}$  labeling of protein is relatively easy provided that the protein can be obtained by heterologous expression in *E. coli*. Moreover, amide  $^{15}\text{N}$  relaxation is dominated by the interaction with the covalently attached amide proton, and this couple can be considered as an isolated two-spin system, simplifying both relaxation measurements experiments and their interpretation to describe protein motions. One pitfall of this approach resides in its limitation to the protein backbone, whose motions do not always fully account for the functional properties of a given protein. In order to overcome this limitation, methyl relaxation approaches have been developed to provide information about amino-acid side chain dynamics.[8] Typically,  $^{15}\text{N}$  protein relaxation studies include the measurement of  $^{15}\text{N}$  longitudinal relaxation rate ( $R_1 = 1/T_1$ ), transverse relaxation rate ( $R_2 = 1/T_2$ ) and the heteronuclear nuclear Overhauser effect ( $^1\text{H}$ - $^{15}\text{N}$  NOE) using  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments in which the intensities of peaks decreased as a function of a delay placed at a point in the sequence when the relevant relaxation process is active. The interpretation of relaxation rates in term of protein motions is performed using different strategies described below.

## **1.3 Interpretation of relaxation parameters**

Since protein motions may be extremely complex, a full description would require a large number of parameters. Moreover, the interpretation of relaxation data requires the choice of a model for motions, which constitutes a severe limitation recognized in 1982 when Lipari and Szabo published their "model free" approach.[9] This approach separates the overall tumbling of the protein (rotational diffusion) from the internal motions assuming a lack of correlation between these two motions. This assumption

is based on the fact that internal and global motions occur a different time scales (tens to hundred of picoseconds for internal motions and nanoseconds for global motions respectively). Assuming this oversimplified view of motions, this model allows a very good description of the rotational diffusion of the protein, and the estimate of an order parameter describing the angular amplitude of fast internal motions. Recognizing that the model-free approach actually requires some assumptions as to the nature of the motions, and therefore a model, further attempts to interpret protein relaxation data proposed the analysis of the values of the spectral density functions which can be calculated directly from the relaxation data. This so-called spectral density mapping approach yields accurate information about the prevalence of motions at the specified frequencies.[10] These two approaches have proven to be complementary and are often associated. The plots of spectral density functions values were used as an initial screen of the motional behavior of the protein, guiding further analysis using more elaborated models such as that from Lipari-Szabo. It should be mentioned that more accurate and complete description of the motional behavior of the protein requires relaxation data to be measured under varying physical parameters such as temperature, solvent etc... Temperature variation for instance allows access to the local heat capacity, which is of particular interest in understanding the ligand binding properties of some systems. Unfortunately, this is often prevented by the excessive amount of measurement time required for such studies.

### 1.3.1 Spectral density mapping

For a two-spin system, assuming that only dipole-dipole interactions and chemical shift anisotropy are effective for relaxation, the relaxation rates  $R_1$ ,  $R_2$  and  $^1\text{H}$ - $^{15}\text{N}$  NOE are dependent on linear combinations of the spectral density function evaluated at five critical frequencies,  $J(0)$ ,  $J(\omega_N)$ ,  $J(\omega_H + \omega_N)$ ,  $J(\omega_H)$ , and  $J(\omega_H - \omega_N)$ . The spectral density function results from the Fourier transform of the time-dependent correlation function  $C(t)$  which describes how the orientation of a N-H vector loses its orientation with time and is equal to :

$$J(\omega) = \frac{2}{5} \frac{\tau_c}{(1 + \omega^2 \tau_c^2)} \quad (1)$$

The value of  $J(\omega)$  at a given frequency provides an accurate estimate of the importance of protein motion at this frequency with no assumption on the motional model. This is why, in an early work, Peng and Wagner suggested insight into protein dynamics could be gained by determining the spectral density at as many as possible points on the spectral density curve.[10] The practical implementation of this approach, termed "spectral density mapping", uses an approximation (high frequency

approximation) to reduce the number of back calculated values to three, equal to the number of relaxation parameters that can be measured from a single set of experiments, and consists in solving the following set of linear equations using  $J(0)$ ,  $J(\omega_N)$ ,  $J(\omega_H)$  as unknown :

$$R_1 = \frac{1}{4} d^2 [3J(\omega_N) + 7J(\omega_H)] + C^2 J(\omega_N) \quad (1)$$

$$R_2 = \frac{1}{8} d^2 [4J(0) + 3J(\omega_N) + 13J(\omega_H)] + \frac{1}{6} C^2 [4J(0) + 3J(\omega_N)] \quad (2)$$

$$\text{NOE} = 1 + \frac{1}{4} T_1 d^2 (\gamma_H / \gamma_N) [5J(\omega_H)] \quad (3)$$

Where  $d = (\mu_0 h \gamma_N \gamma_H / 8\pi^2) \langle 1/r_{NH}^3 \rangle$ ;  $c = \omega_N \Delta\sigma / \sqrt{3}$  ;  $\mu_0$ : permeability of vacuum;  $h$ : Planck's constant;  $\gamma_N$  and  $\gamma_H$ : the gyromagnetic ratios of  $^{15}\text{N}$  and  $^1\text{H}$  nuclei, respectively;  $\Delta\sigma$ : chemical shift anisotropy of the  $^{15}\text{N}$  spin.

The advantage of the “reduced” spectral density mapping approach is that it uses the same relaxation parameters as the model-free formalism, allowing the results obtained using the two approaches to be compared.

Based on the equations listed above, plots of spectral density function at frequencies 0,  $\omega_N$  and  $\omega_H + \omega_N$  can be analyzed as follows:

1. The baseline in the  $J(0)$  spectral density profiles shows the global behavior of the molecule.
2. Values of  $J(\omega_H)$  and  $J(\omega_N)$  affect all three relaxation parameters, whereas  $J(0)$  only influences  $R_2$ . Therefore, conformational exchange will translate into elevated values of  $J(0)$  accounting for the additional contribution,  $R_{ex}$  to the observed transverse relaxation rate. As a result, variations of  $J(0)$  profile can be used to detect the presence of molecular motion on the  $\mu\text{s}$ - $\text{ms}$  time scale.
3. Increased motions close to either  $^{15}\text{N}$  or  $^1\text{H}$  Larmor frequency enhance  $R_1$  relaxation of the  $^{15}\text{N}$  nucleus, although  $R_1$  is most sensitive to changes in  $J(\omega_H)$ .
4. Reduced values of  $^1\text{H}$ - $^{15}\text{N}$  NOE result from fast molecular tumbling on the ps-ns time scale since the NOE depends only on  $J(\omega_H + \omega_N)$  and  $J(\omega_H - \omega_N)$ . Commonly, low values of NOE are observed in the tails of the proteins or in loop regions.

In addition to the analysis of spectral density function values against sequence, parametric plots of  $J(\omega_N)$  or  $J(\omega_H + \omega_N)$  as a function of  $J(0)$  have proven very useful to obtain an overview of the dynamic events affecting a protein backbone.[11] The partition of points in different area of the plot provides an immediate indication of the type of motions that affects the backbone. Points gathering along the solid line, which represents values of the spectral density function calculated for rigid body motion, are characterized by a lack of internal motions whereas points aligned along a line joining two positions given by two values of the correlation time (one for internal motion and

another one for global tumbling) will be indicative of different levels of flexibility.

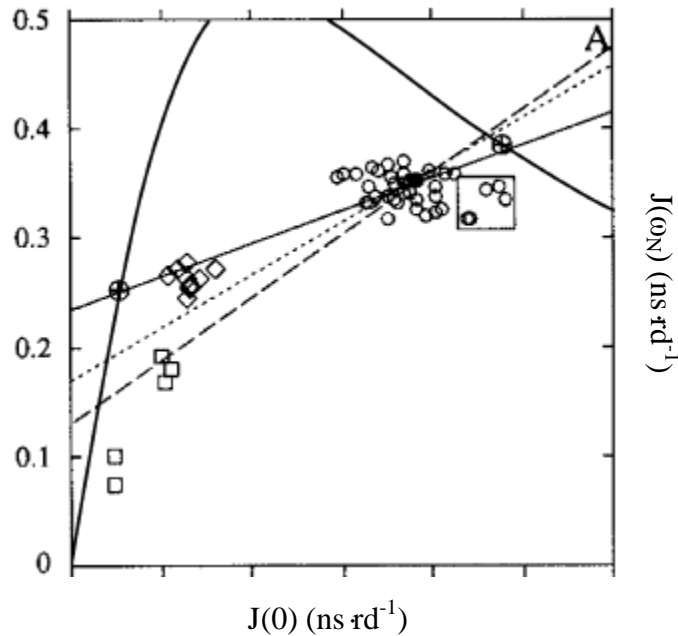


Fig.2 Spectral density plots for the DNA-binding domain of FruR.[11] The solid curve represents values of the spectral density calculated at specific frequencies, assuming isotropic rigid body. Linear regression fitting to the data measured at  $^1\text{H}$  frequency gives the correlation time for sets of residues. Residues involving conformational exchange were excluded from the fit (boxed).

UNIVERSITE DE STRASBOURG

### 1.3.2 Lipari-Szabo Model free analysis

As mentioned above, the Lipari Szabo model-free formalism was proposed as the simplest way to obtain a model for the spectral density function with a minimal number of parameters. This function is written as a sum of two spectral density functions describing the fast internal motions and the overall tumbling of the protein, respectively:

$$J(\omega) = S^2 J^0(\omega) + (1 - S^2) J^i(\omega)$$

$J^0(\omega)$  and  $J^i(\omega)$  being the spectral density functions describing the global and internal motions respectively that are characterized by two correlation times (one for global motion and the second for internal motion). These motions, which occur on different time scales, are assumed to be uncorrelated.  $S^2$  is an order parameter, which describes the amplitude of fast internal motions. Completely restricted internal motions will yield a  $S^2$  equal to 1, and completely unrestricted motions are characterized by a  $S^2$  equal to 0. The value of  $S^2$  could be calculated from the following expression :[12]

$$S^2 = \frac{4\pi}{5} \sum_{M=-2}^2 \langle Y_{2,M}(\theta, \varphi) \rangle \Big|_0^{\tau_c} \langle Y_{2,M}^*(\theta, \varphi) \rangle \Big|_0^{\tau_c}$$

where  $Y_{2,M}(\theta(t),\varphi(t))$  are the spherical harmonics functions which are averaged over all values of the spherical coordinates  $\theta$  and  $\varphi$  of the inter-nuclear vector that are explored by internal motions during a time  $\tau_c$ .  $Y_{2,M}^*(\theta(t),\varphi(t))$  are the conjugate of the spherical harmonics functions. It should be mentioned that  $\langle Y_{2,M}(\theta(t),\varphi(t)) \rangle$  represents an ensemble average over all possible orientations of a given vector. For relaxation, and in particular high frequency motions, the effective value of the order parameter will be limited by the range of orientations explored within a delay which is lower than the global correlation time.

## 1.4 Alternative methods to probe protein dynamics

In addition to relaxation measurements using  $^{15}\text{N}$  and  $^1\text{H}$  as probes, other methods can also be used to evaluate the dynamic behaviors of macromolecules including residual dipolar couplings and HET-SOFAST experiments.

### 1.4.1 Residual dipolar couplings (RDC)

Residual dipolar couplings were introduced in the late 1990s as a valuable tool to obtain structural information on proteins. The method relies on the ability to induce a weak alignment of the molecules using oriented media. [13] There are several types of orienting media such as Pf1 filamentous phages, polyethylene glycol (PEG) or phospholipid bicelles and strained polyacrylamide gel.[14] Partial alignment of the molecule leads to the incomplete averaging of anisotropic interactions such as dipolar interactions which could be measured as variable contributions to scalar coupling constants. These contributions depend on the orientation of the vector describing the inter-nuclear interaction relative to the magnetic field. If  $\theta$  is the angle between the inter-nuclear vector and the magnetic field, the contribution to the scalar coupling constant will have the following dependency:

$$\text{RDC}(\theta) = A(3\cos^2\theta - 1)$$

where  $A$  is a constant which depends on the alignment strength and the type of interaction. Assuming a rigid molecule, the measurement of five independent RDC allows the definition of the orientation of the molecule relative to the laboratory frame (which is described by the eigen vectors of the orientation tensor) and the determination of the degree of alignment (described by the eigenvalues of the tensor). Other measurements can then be used as a cross-validation check of the structure, or alternatively, to refine the structure using various approaches.[15] Once the parameters of the alignment tensor are known, the RDC values are expressed within a molecular frame using the following expression:

$$D(\theta,\varphi) = D_a[(3\cos^2\theta - 1) + 1.5R\sin^2\theta\cos 2\varphi]$$

where  $D_a$  is the axial component of the alignment tensor (in Hz),  $R$ ; the rhombicity

and  $(\theta, \varphi)$  the polar coordinate of the inter-nuclear vector in the alignment frame. Protein internal motions induce a decrease of RDC values when compared to a rigid molecule with the same alignment properties. This effect is described by an order parameter which, as for fast internal motions, has a value of 1 when no motion is present and 0 for an isotropic motion. There is however a key difference between these order parameters, since the order parameter derived from the RDC measurements is averaged by motions on time-scales up to the inverse of the RDC coupling, that is by motions up to the millisecond time scale.[12]

$$S_{\text{RDC}}^2 = \frac{4\pi}{5} \sum_{M=-2}^2 \langle Y_{2,M}(\theta, \varphi) \rangle \Big|_0^{\text{ms}} \langle Y_{2,M}^*(\theta, \varphi) \rangle \Big|_0^{\text{ms}}$$

Therefore, the interpretation of RDC couplings in term of motions provides a way to probe motions that are slower than the molecular tumbling correlation time and the comparison of  $S_{\text{RDC}}^2$  and  $S_{\text{LS}}^2$  allows the detection of motions on the nano-second to millisecond time scale.[12] However, this method requires the measurement of several values (at least five) of the RDC for the same interaction (inter-nuclear vector) in different orienting media, a requirement that is not always compatible with the studied protein. Recently, RDC measurements of disordered proteins have proven to be a very good tool to obtain information on the flexibility of a disordered peptide chain.[16] It was in particular shown that the restriction of local disorder induced by a particular sequence of amino-acids (such as a bulky hydrophobic residue) could be detected by RDC values, opening new perspectives to establish sequence-motion relationships that are currently lacking in bioinformatics analysis of protein sequences.

#### 1.4.2 HET-SOFAST experiment

A slight modification of the fast acquisition heteronuclear correlation SOFAST experiment led to an interesting way to probe local dynamics and proton density of a protein.[17] This experiment is named HET-SOFAST, and its biggest advantage is that it can be used to detect changes of motions along the peptide chains within a short time. In the laboratory, it has proven to be very useful to preview the overall motions of molecules before starting more thorough relaxation measurements.

The SOFAST experiment is a  $^1\text{H}$ - $^{15}\text{N}$  HMQC heteronuclear correlation experiment which exploits the fast selective  $T_1$  relaxation properties of the amide protons of a peptide chain, which, by accelerating the return to thermal equilibrium of the longitudinal magnetization, allows a significant reduction of the relaxation delay. The selective excitation of amide protons is performed using two band selective pulses for excitation and refocusing. The short selective relaxation time is due to the ability of



the neighboring protons to efficiently relax the excited amide proton and further propagate the magnetization to the solvent. This characteristic relies both on efficient cross-relaxation rates (indicative of rigid inter-proton vectors) and on a high density of neighboring protons. The Heterogeneity-SOFAST (HET-SOFAST) experiment probes the above mentioned features of the peptide chain by recording two  $^1\text{H}$ - $^{15}\text{N}$  SOFAST, one of which being modified in order to perturb the network of aliphatic spins by applying a band selective excitation pulse prior to the relaxation time (which is set to 100 to 200 ms). Peak intensities from the two experiments are compared by computing the intensity ratio. For rigid amide protons, buried in the core of the protein, this ratio is expected to be significantly lower than 1 ( $\lambda_{\text{noe}} < 1$ ) whereas it will be nearly equal to 1 for an amide proton located within a flexible loop ( $\lambda_{\text{noe}} \sim 1$ ).

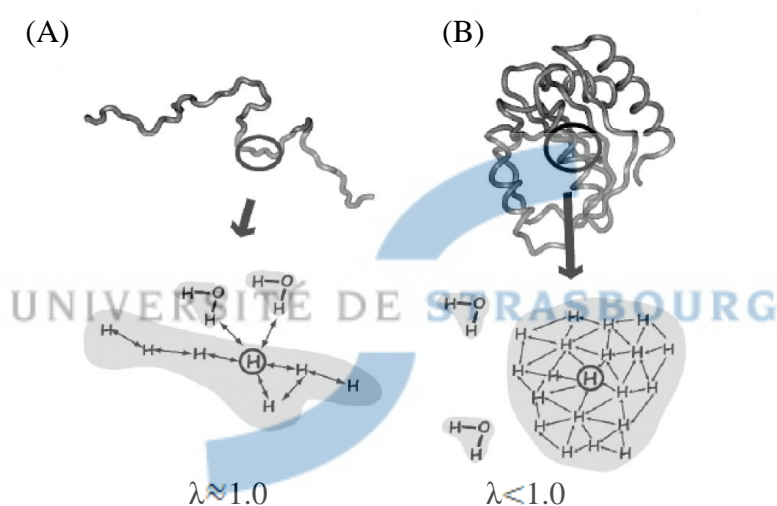


Fig.3 Conformations of polypeptide chains. (A) An elongated polypeptide chain is characterized by a high degree of contact of the amide groups to the water. The ratio of the peak intensities  $\lambda$  is about 1. (B) The globular well-folded protein is identified by low water accessibility of the amide groups and the values of  $\lambda$  are below 1.[17]

## 2. Materials and methods

### 2.1 Relaxation measurements and $^1\text{H}$ - $^{15}\text{N}$ NOE experiments

$R_1$ ,  $R_2$  relaxations and  $^1\text{H}$ - $^{15}\text{N}$  NOE experiments were acquired at 600MHz. The experiments for AR-DBD wild type and T575A mutant were recorded at 300K while those for ATXN7 and ATXN7 L3 were recorded at 295K. The relaxation delays were set to 0, 100, 200, 400, 600, 800, 1000, 1200, 1400 and 1600 ms for longitudinal relaxation and to 5, 10, 20, 40, 60, 80, 100, 120 and 140 ms for transverse relaxation measurements respectively. For error estimation, the spectra with relaxation delays of 400 ms ( $T_1$ ) and 40 ms ( $T_2$ ) were recorded twice. The  $^1\text{H}$ - $^{15}\text{N}$  NOEs were acquired by



recording two spectra with and without proton saturation. The 2D data sets contained 2048 real data points for 256  $t_1$  blocks for AR-DBD wild type and T575A mutant. For ATXN7 and ATXN7 L3, the number of real data points was set to 4096. Eight scans were acquired per free induction decay in  $T_1$  and  $T_2$  relaxation measurements except for  $^1\text{H}$ - $^{15}\text{N}$  NOE where 54 scans were accumulated. Cross peak intensity measurements on 2D maps was performed using CARA.  $R_1$  and  $R_2$  values were determined by fitting the decaying intensities with a two-parameter exponential function. The NOE values were calculated using the ratio of the cross peaks measured in the spectra recorded with and without proton saturation.

### 2.1.1 Analysis of the diffusion tensor from relaxation measurement

The diffusion tensor were analyzed using TENSOR.[18] The principle axes of the rotational diffusion tensor were determined using experimental values of  $R_1$ ,  $R_2$  and  $^1\text{H}$ - $^{15}\text{N}$  NOEs. The order parameters ( $S^2$ ) were calculated using the Lipari-Szabo model-free formalism for anisotropic tumbling

### 2.2 HET-SOFAST experiments

HET-SOFAST experiments were recorded for  $^{15}\text{N}$ -labeled ATXN7 and ATXN7 L3 proteins at 295K using a Bruker DRX 600MHz spectrometer. Parameters were as follows:

Proton carrier frequency	5401.17Hz
Contact time	250ms
Band width	4200Hz -4200Hz (Off set)
Power and frequency of the perturbation pulse	1180 $\mu\text{s}$ @0db; 2250 $\mu\text{s}$ @24db; 2000 $\mu\text{s}$ @14.4dB Shaped pulse power SP3=13.5dB
Shape of pulse	SPNAM1=pc9_2_200.bb; SPNAM2=reburp.1000 SPNAM3=ISNOB5.200
Number of scans	16
Acquisition points (F1xF2)	1024x200
Duration of the experiments	26min 24sec

Two spectra were recorded: with and without the  $\text{H}^{\text{sat}}$  inversion pulse. Peaks intensities were measured in CARA and the ratio of peaks intensities ( $\lambda = I^{\text{sat}}/I^{\text{ref}}$ ) were used for further analysis.

## 2.3 Residue dipolar couplings (RDC) measurements

### 2.3.1 Sample preparation

RDC measurements were performed using strained polyacrylamide gels as orienting medium.[14] 6~8% polyacrylamide gels were made within a dedicated 3-mm cylinder polymerization chamber by mixing liquid polyacrylamide and polymerization initiator. The gel was then washed at least three times with deionized water to remove excess polyacrylamide, and immersed into a concentrated protein solution which diffuses into the gel. To account for the dilution by the alignment medium, a concentrated stock solution of protein is required and concentrations of 600  $\mu\text{M}$  were used. The diffusion process is allowed to last for at least 16 hours before transfer of the gel into a dedicated NMR tube with an inner diameter of  $\sim 4$  mm using a specific device shown in Fig.4. One important aspect of this process is to preserve the homogeneity of the sample during the transfer into the NMR tube.

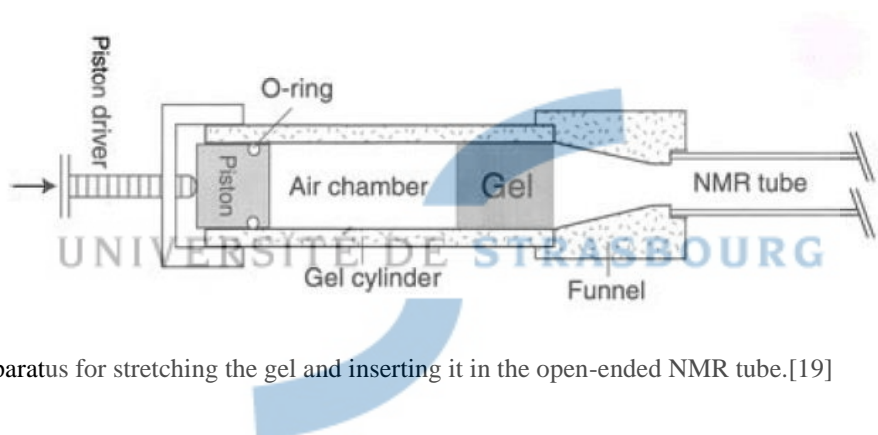


Fig.4 Apparatus for stretching the gel and inserting it in the open-ended NMR tube.[19]

### 2.3.2 Experiments for RDC measurements

RDC were measured on  $^1\text{H}$ - $^{15}\text{N}$  vectors using an IPAP  $^1\text{H}$ - $^{15}\text{N}$  HSQC modified to eliminate the signal from acrylamide.[20] IPAP experiments allow alleviation of the overlap problems originating from the doubling of the number of peaks in F1-  $^{15}\text{N}$  coupled HSQC spectra. [20] This experiment differs from a standard HSQC by the introduction of an additional delay during  $t_1$  evolution. This delay allows  $J_{\text{NH}}$  to evolve during  $1/2J_{\text{NH}}$  leading to antiphase signal. The successive addition and subtraction of the modified HSQC with a regular F1-  $^{15}\text{N}$  coupled HSQC spectra allows each component of the doublet to be obtained and an accurate measurement of its frequency to be made. The frequency difference between the two components of the doublet is used to get the value of the  $J_{\text{NH}}$  coupling constant. This value is subtracted from the  $J_{\text{NH}}$  coupling constant measured in absence of gel to obtain the actual value of the RDC.

### 3. Results

#### 3.1 Dynamics of ARDBD wild type and T575A mutant

Dynamics of AR-DBD WT and T575A mutant were probed by recording a series of  $^{15}\text{N}$  relaxation experiments at 303K. 72 residues were observed in  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra. Fig.5 shows the profiles of the  $^{15}\text{N}$   $R_1$ ,  $R_2$  relaxation rates, and  $^1\text{H}$ - $^{15}\text{N}$  NOE together with the spectral density functions at 0,  $\omega_{\text{N}}$  and  $\omega_{\text{H}}$  frequencies for AR-DBD wild type (WT) and T575A mutant.

##### 3.1.1. Micro- to milliseconds time-scale motions

Analysis of the  $R_2$  relaxation rates and  $J(0)$  values measured for the AR-DBD WT and T575A mutant immediately suggests the presence of motions occurring on the  $\mu\text{s}$ -ms time scale. The residues affected by such motions are localized in specific parts of the protein domain and the effect is dominant in the second zinc finger (ZNII). In particular, residues T602, I603 and C611 have increased  $J(0)$  values in both WT and T575A (colored in yellow in figure 6). T602 and I603 are located in a loop region close to the ZNII coordination site while C611 is involved directly in zinc coordination (Fig.5A). The overall profile of  $J(0)$  is found to be very similar between the wild type and mutant proteins, with the exception of small fluctuations of  $J(0)$  that are noticeable in the region adjacent to the first zinc binding module of the mutant protein (Fig.6B). Indeed, residues H570 and G572 in T575A mutant reveal a higher value of  $J(0)$  compared to the baseline values, suggesting a possible difference in conformational exchange motions between WT and T575A on the  $\mu\text{s}$ -ms time scale. The presence of the histidine residue suggested that this equilibrium may be due to an exchange between protonated and deprotonated forms of the histidine imidazole ring [21] and prompted us to measure the pKa of H570 by measuring the frequencies of the non-exchangeable protons of the imidazole ring as a function of pH. This study allowed us to detect a shift of the pKa of the histidine by 0.5 pH unit upon mutation providing a possible molecular basis to explain the biological effect of this mutation.

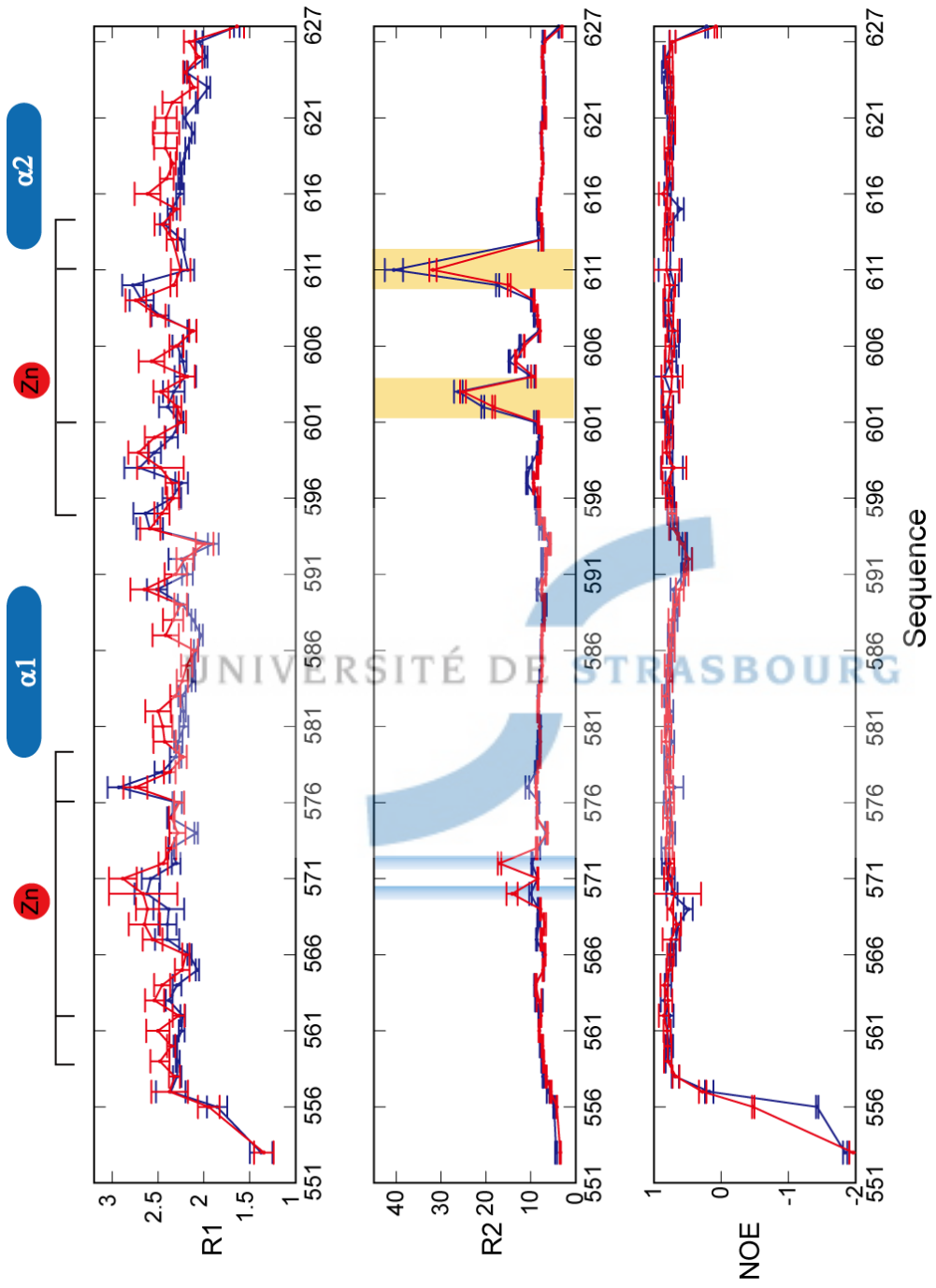


Fig.5 (A) Experimental values of relaxation rates for AR-DBD wild type (blue) and T575A mutant (red).

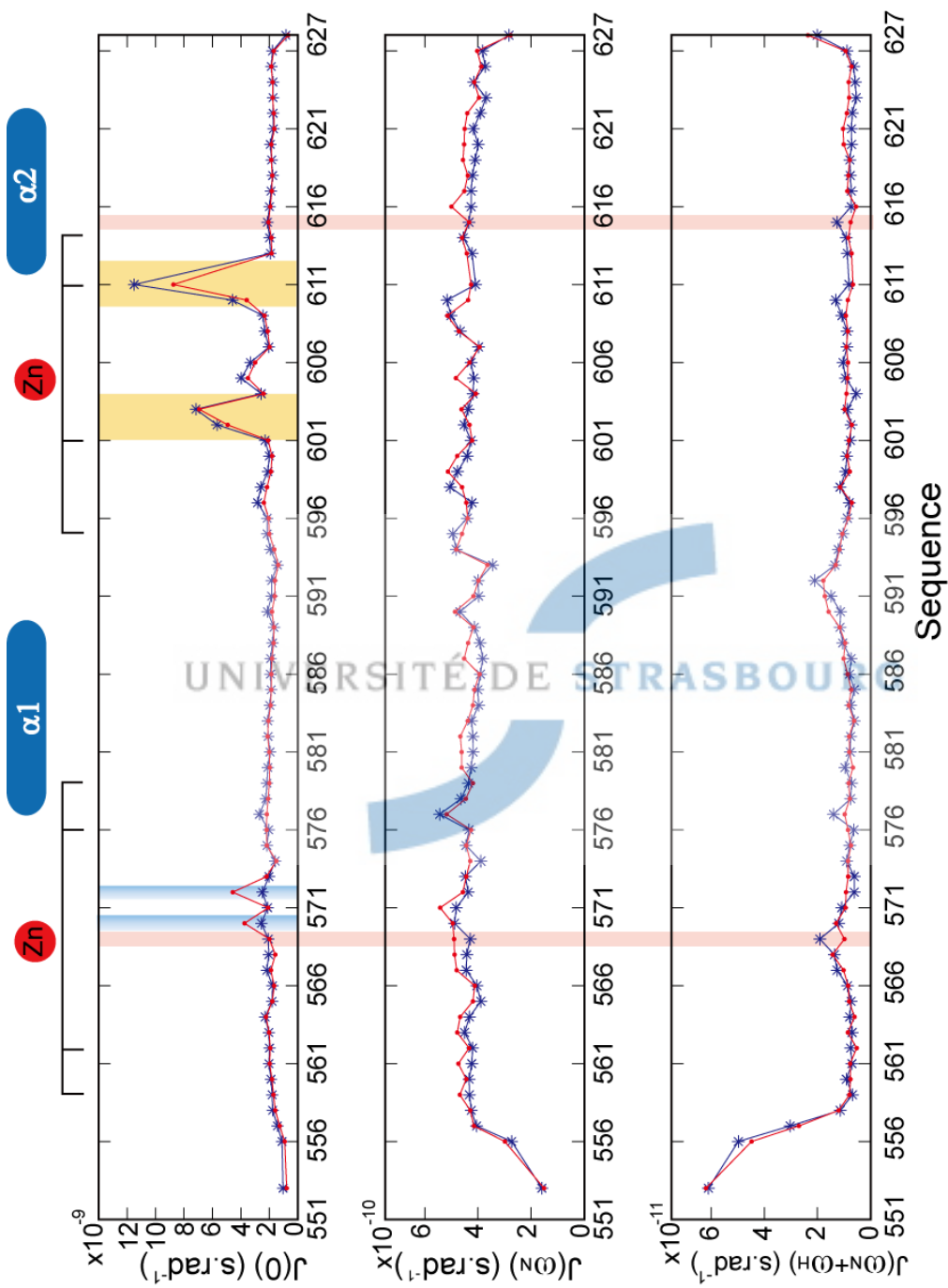


Fig.5 (B) Spectral density functions at 0,  $\omega_N$ , and  $\omega_{H+\omega_N}$  for AR-DBD wild type (blue) and T575A mutant (red). Residues T602, I603 and C611 show larger  $J(0)$  in both WT and T575A mutant (orange) whereas H570 and G522 possess increased  $J(0)$  in T575A mutant (blue).

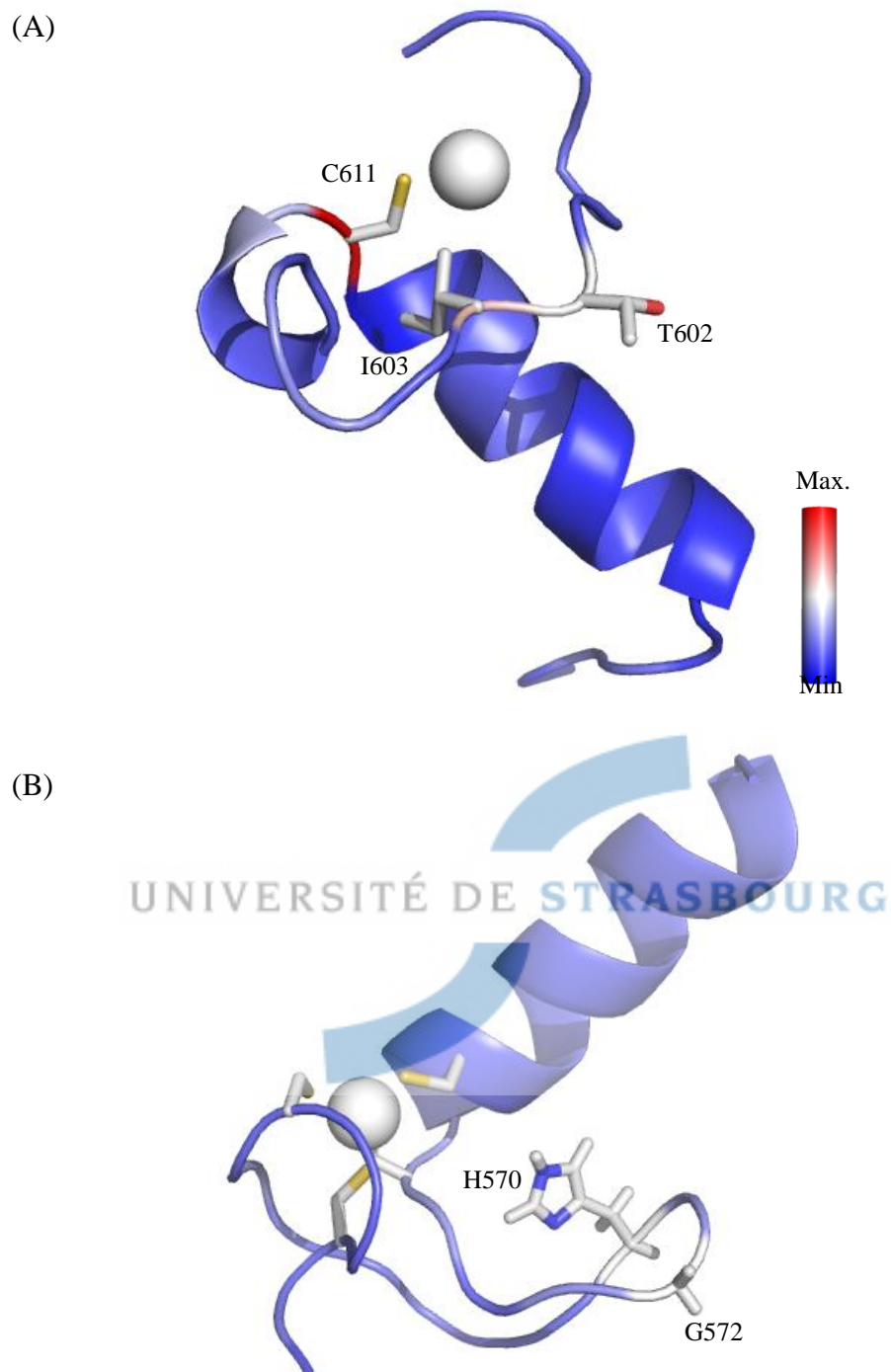


Fig.6 (A) Structure of the second zinc module of AR-DBD WT. C661 possesses a maximum  $J(0)$  value while T602 and I603 also show higher  $J(0)$  values, suggesting the conformational exchange on the  $\mu$ -ms time scale. (B) Structure of the first zinc module of AR-DBD T575A mutant. H570 and G522 are located in the loop region and have increased  $J(0)$  value. Residues involved in zinc coordination are labeled as the pink sticks.

### 3.1.2 ps-ns time-scale motions

$^1\text{H}$ - $^{15}\text{N}$  NOE together with  $J(\omega_{\text{H}}+\omega_{\text{N}})$  values were used to analyse fast internal motions in WT and T575A mutant of AR DBD (Fig.5). Whereas high frequency motions within the N-terminal tail of the domain are characterized by the negative  $^1\text{H}$ - $^{15}\text{N}$  NOE measured for N-terminal residues such as M553, no such effect could be observed for the C-terminal tail indicating a more rigid structure. Indeed, the C-terminal tail of the DBD domain is structured with an  $\alpha$ -helix, explaining the lack of flexibility. Residues in the loop region such as C569, K592, R607 and R608 also display reduced  $^1\text{H}$ - $^{15}\text{N}$  NOE values indicative of enhanced flexibility. However, it is not clear whether the slight differences of  $^1\text{H}$ - $^{15}\text{N}$  NOE observed between WT and T575A mutant proteins reveal significantly different flexibility. In order to further our analysis of internal motions, and detect possible differences between the two domains, a “model-free” analysis of relaxation data was performed using the TENSOR2 program.[18] This program was also used to gain information about the anisotropic diffusion properties of both domains. The order parameters ( $S^2$ ) obtained for each residue of AR-DBD WT and T575A mutant proteins are shown in Fig.7A and 7B respectively and mapped onto the 3D structures in Fig.6A. Most order parameters are close to 1 indicating an overall rigidity of the domain. Lower values of  $S^2$  are found in loops and tail regions of the domain indicating higher flexibility.

Some slight differences are observed in the  $S^2$  values measured for WT and T575A mutant as shown in the difference plot (Fig.7C). These differences are observed for residues H567, C569 and R615 which consistently display higher mobility in the WT protein (red areas in Fig.8B). In addition, fluctuating differences were also apparent in the loop regions between WT and T575A mutant (green region in Fig.7C), indicating distinct motional behavior.



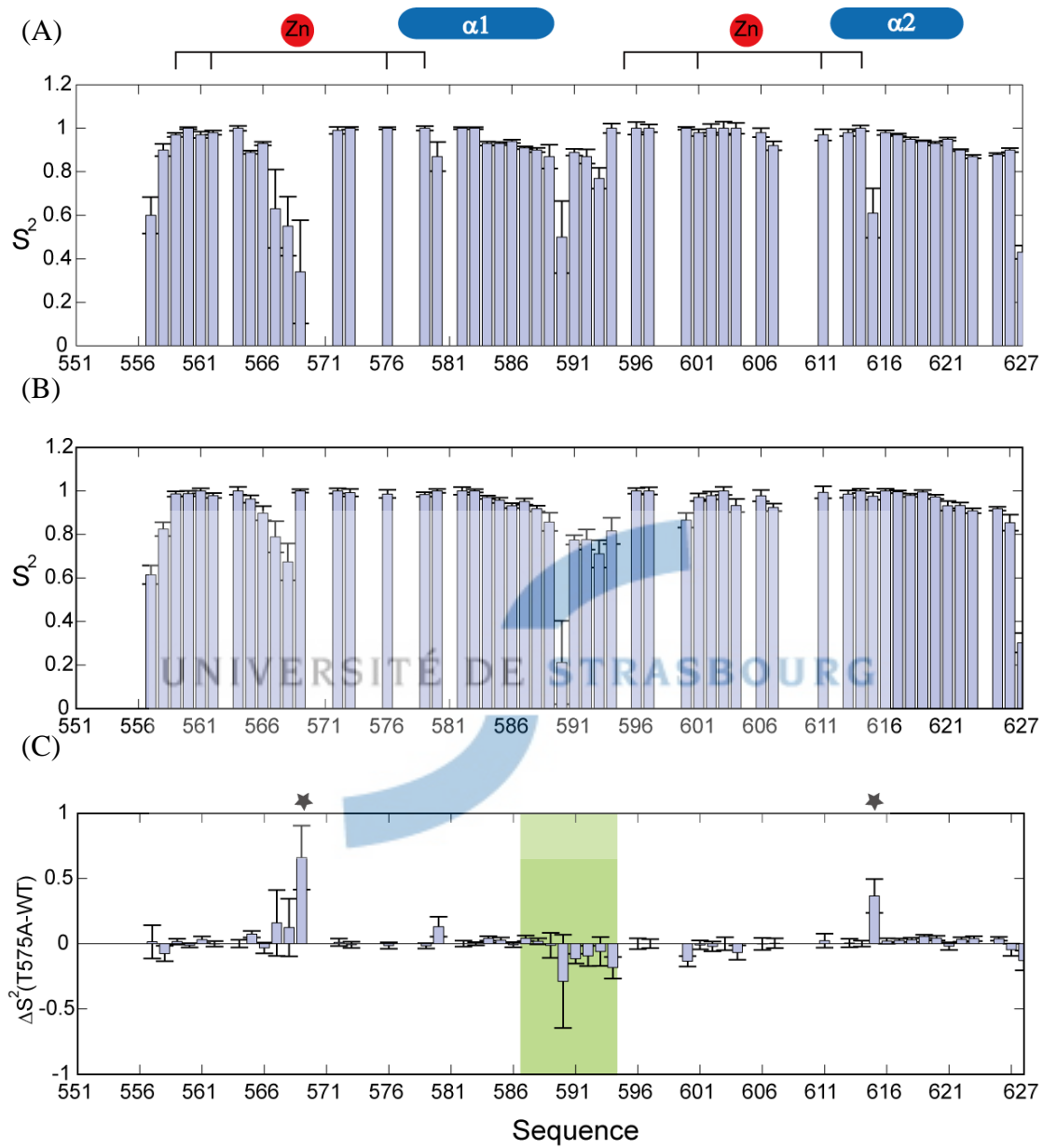


Fig.7 The order parameters of AR-DBD (A) WT and (B) T575A mutant. (C) Difference of order parameter between WT and T575A mutant.

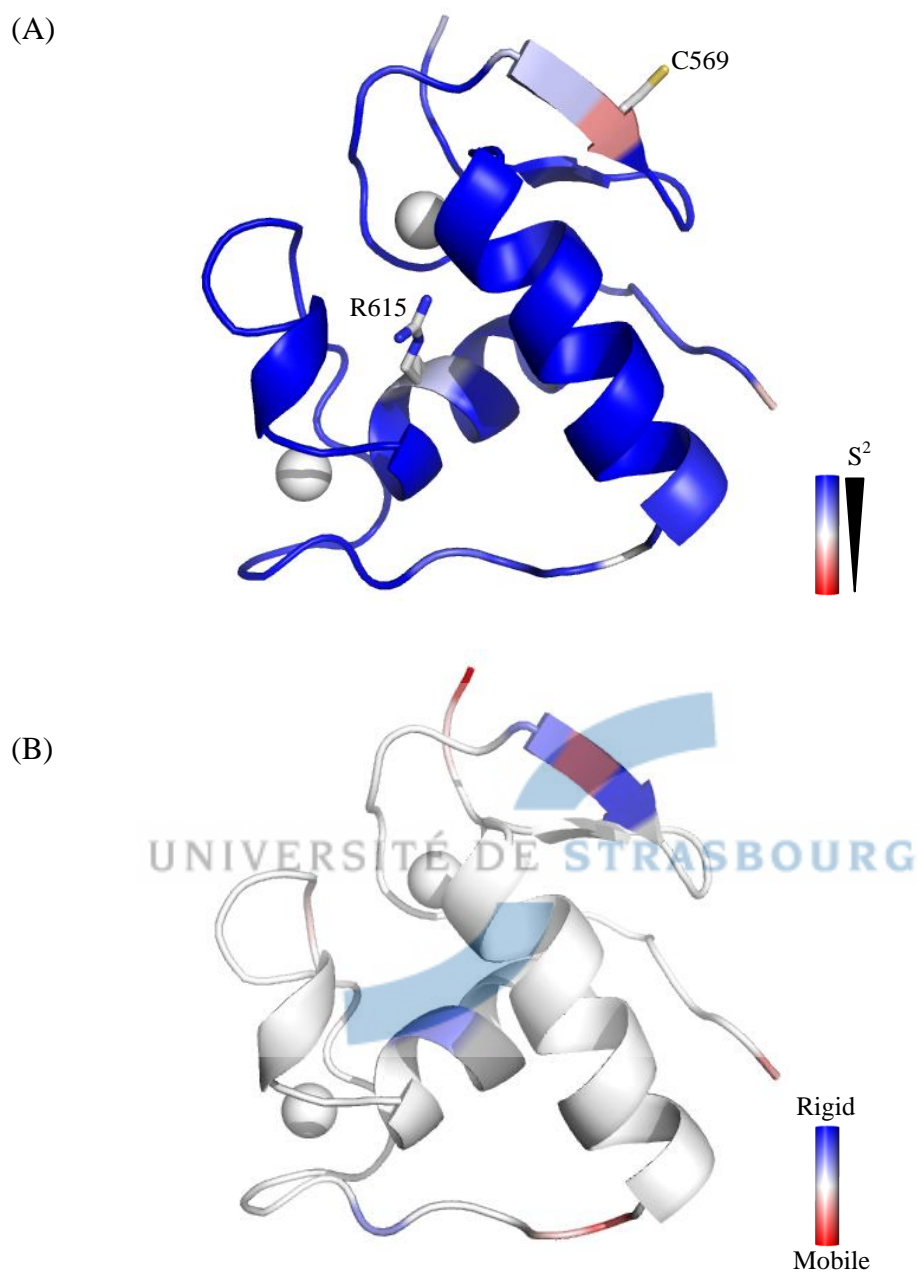
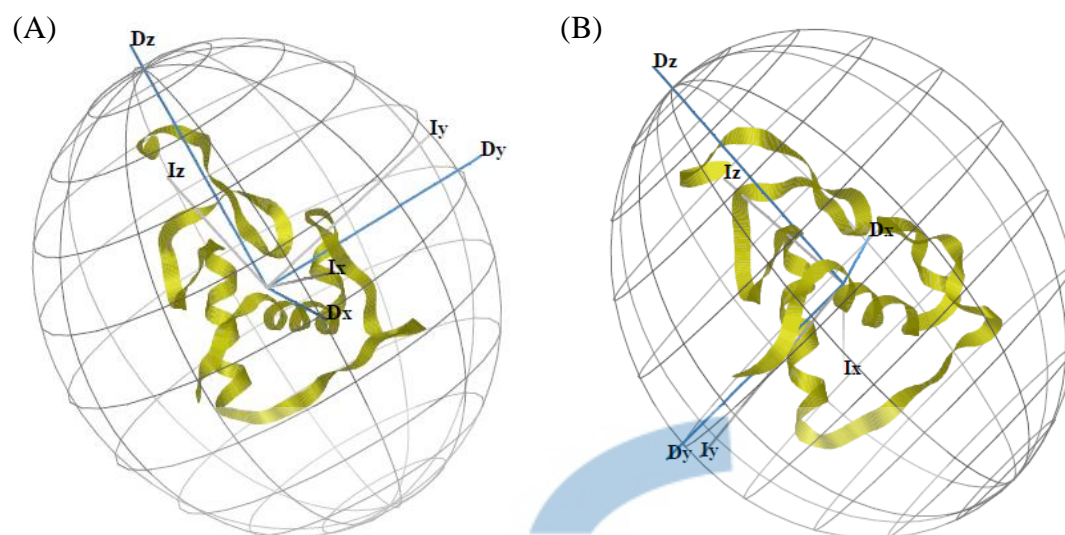


Fig.8 (A) Structure of AR-DBD WT. The backbone is colored using the scale of order parameter ( $S^2$ ). (B) Difference of mobility between WT and T575A mutant. The color of backbone represents difference of order parameters between WT and T575A mutant (Fig.7C). The red parts represent the positive values in Fig.7C, indicating higher mobility in WT and contrast with the blue areas with the negative values in Fig.7C, showing higher mobility in T575A mutant.

Results of  $^{15}\text{N}$  relaxation measurements also provide information on the rotational diffusion tensor. Fig.9 shows the rotational diffusion tensors of AR-DBD WT and T575A mutant. The orientation of rotational diffusion tensors for both WT and T575A mutant domains show considerable similarity; in particular, the z axis of diffusion tensor is nearly parallel to the N-terminal tails. Moreover, the internal tensors of AR-DBD are quite close to the diffusion tensors (Fig.9 axes  $I_x$ ,  $I_y$  and  $I_z$ ), indicating high correlations between the static and motional structures. The orientation of N-H vector in the motional structure reflects the slight mobility compared with that in the static structure and the slight difference in orientation between rotational diffusion tensor and internal tensor. Fig.10 shows the comparison of  $R_2/R_1$  derived from the experimental values and back calculation from the structures (model1~model5). Values from back calculation for most residues fit with the experimental values suggesting that the orientations of the NH vectors for each residue with respect to the diffusion tensor is close to those in the refined structures. However, several experimental data points are far from the back calculated values, indicating occurrence of conformational exchange. These results conclude that the orientation of refined structure is quite consistent with the motional structure in solution, not only occurs in the WT domain but also in the T575A mutant domain.

UNIVERSITÉ DE STRASBOURG



UNIVERSITÉ DE STRASBOURG

Fig.9 Rotational diffusion tensor of AR-DBD (A) WT and (B) T575A mutant.  $D_x$ ,  $D_y$  and  $D_z$ : axes of rotational diffusion tensor;  $I_x$ ,  $I_y$  and  $I_z$ : internal tensor.

Table1 Diffusion tensor parameters from  $R_2/R_1$  data for AR-DBD

Tensor	$D_{xx} (10^{-9} s^{-1})$	$D_{yy} (10^{-9} s^{-1})$	$D_{zz} (10^{-9} s^{-1})$	$\tau_c$ (ns)	$\chi_2$
Wild type	2.99	3.53	3.81	4.9	113.6
T575A mutant	3.06	3.58	4.03	4.6	271.9

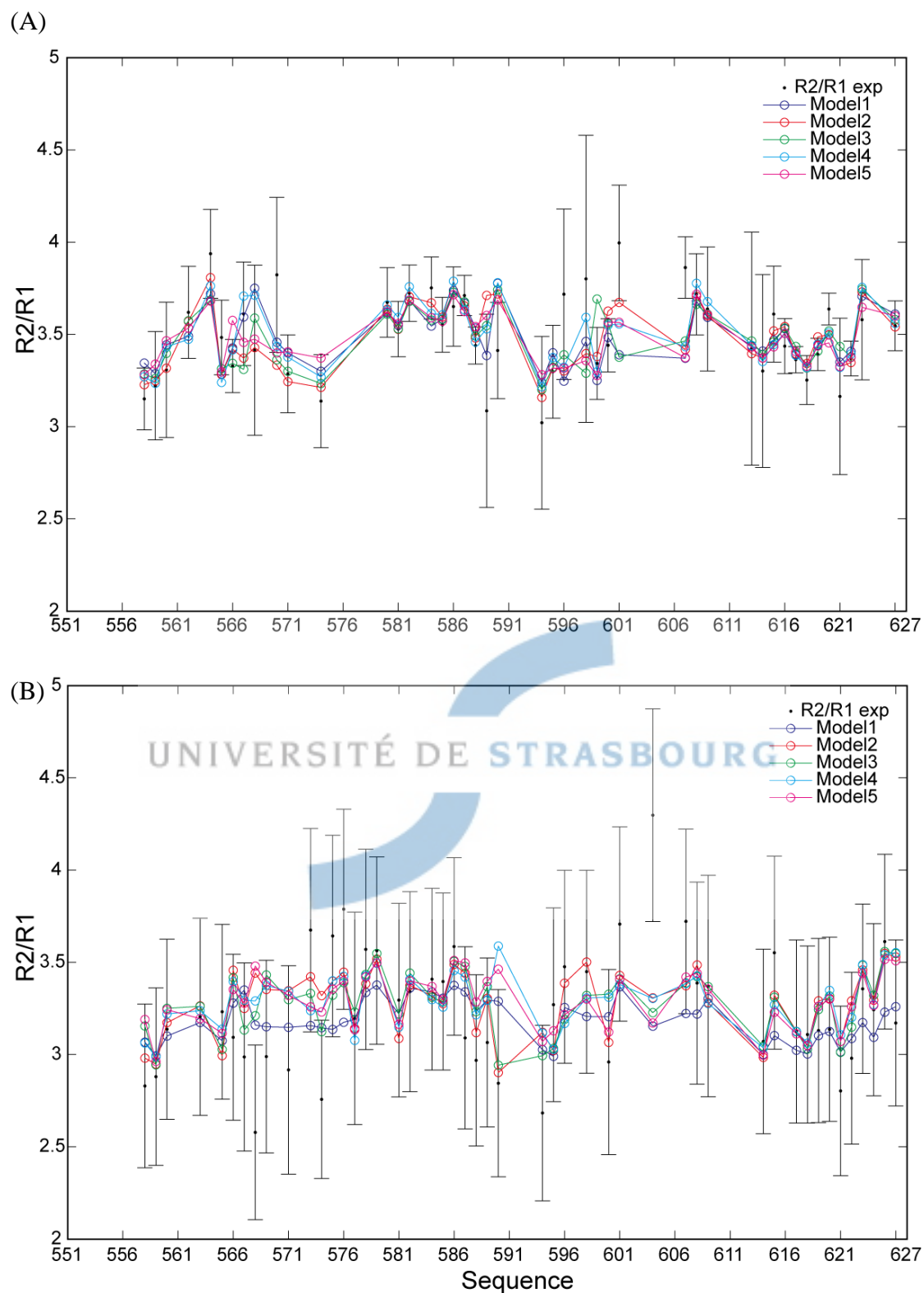


Fig.10 Comparison of  $R_2/R_1$  ratios derived from experimental values (black dots) with back calculated values from structures for AR-DBD (A) WT and (B) T575A mutant. Model 1~ model5 are ordered from the lower energy to higher energy.

### 3.1.3 Conclusions

The  $^{15}\text{N}$  relaxation measurements showed several similarities in dynamics for AR-DBD WT and T575A mutant. The profiles of baseline in  $J(0)$  for WT and T575A mutant domains suggest that internal motions for both are on the same time scale ( $\mu\text{s}$ -ms time scale). The appearances of reduced  $^1\text{H}$ - $^{15}\text{N}$  NOE in the N- and C-terminal regions for WT and T575A reveals the fast motions on the ps-ns time scale. Residues in proximity to the second zinc binding site such as T602 and I603 show conformational exchange on the  $\mu\text{s}$ -ms time scale in both WT and T575A mutant domains. On the other hand, the rotational diffusion tensors show similar orientation of the NH vectors between WT and T575A mutant domain, in particular the z axis of the diffusion tensor is nearly parallel to the N-terminal tails in both WT and T575A mutant domains.

However, several differences in dynamic behavior are shown between AR-DBD WT and T575A mutant. First, residues H570 and G572 reveal higher  $J(0)$  values in T575A mutant than in WT domains, indicating slight fluctuation of  $\mu\text{s}$ -ms time-scale motions for the T575A mutant domain. Second, the order parameters from Lipari-Sazbo model-free analysis reveal distinct internal motions between WT and T575A mutant domains. Residues C569, in close proximity to the residues H570 and G572, and R615, at the C-terminal end of helix  $\alpha_2$ , show higher mobility in the WT domain than in the T575A mutant domain and the loop regions also reveal slight difference between WT and T575A mutant domains. These results suggest that dynamic behavior of WT and T575A mutant domains are distinct despite high structural similarity. In particular, the imidazole ring of H570 reveals a different pKa between WT and T575A mutant domains, providing a molecular basis for the effect of the mutation.

## 3.2 Dynamics of ATXN7 zinc finger domains

Our work on the second domain of Ataxin 7 allowed the discovery of two novel zinc finger folds (see chapter 6). Remarkably, the two domains differ in the organization of the secondary structure elements around a conserved zinc coordination sphere. This difference was not initially anticipated from the analysis of the primary sequences which displayed a high level of similarity, with the exception of regions located at the domain boundaries. In order to further our understanding of this structural domain, we were interested to learn whether the different organization of secondary structure elements leads to distinct dynamic properties.

### 3.2.1 Relaxation measurements of ATXN7 and ATXN7 L3 SCA7 domains

$^{15}\text{N}$  heteronuclear relaxation measurements were performed at 600 MHz using  $^{15}\text{N}$

labeled samples. For ATXN7 domain,  $R_1$ ,  $R_2$  and  $^1\text{H}$ - $^{15}\text{N}$  NOE values could be obtained with good precision for 67 residues out of 72 (Fig.11A), whereas data for 74 residues out of 84 (Fig.11B) were obtained for ATXN7 L3 SCA7 domain. A comparative overview of the relaxations rates measured for the two SCA7 domains immediately suggested different motional behavior. Whereas reduced values of  $^1\text{H}$ - $^{15}\text{N}$  NOEs were clearly present for few residues at the N- and C-terminus of ATXN7 SCA7 domain, indicating the occurrence of fast internal motions in these tails (blue areas in  $^1\text{H}$ - $^{15}\text{N}$  NOEs and Fig.11A),  $^1\text{H}$ - $^{15}\text{N}$  NOEs as well as  $R_2$  displayed a clear partition between two parts of the molecule, with the 30 last residues showing low values of  $^1\text{H}$ - $^{15}\text{N}$  NOEs.

The relaxation rates measured for the two domains were translated into spectral density functions values at 0,  $\omega_N$  and  $\omega_N + \omega_H$  frequencies for further study (Fig.14). Types of motions affecting both domains were further analyzed qualitatively using parametric plots of  $J(\omega_N)$  against  $J(0)$ . A number of similarities and differences emerged from the comparison of these plots. As noticed earlier [22], most of the residues are distributed along a line joining two correlation times on the parametric curve. This feature is most striking on the plot obtained for ATXN7 L3, where two groups of residues corresponding to the core domain and the disordered C-terminal tail were clearly seen. The two correlation times deduced from the intersection of the linear correlation and the parametric curve are 1.0 and 6.0 ns. The picture obtained for the ATXN7 SCA7 domain (Fig.12) differs significantly from ATXN7 L3 (Fig.13). The group corresponding to residues with significant internal mobility is more reduced, and the longer correlation time was found to be significantly higher (7.3 ns) than that found for ATXN7 L3, indicating a slower tumbling rate, reflecting a larger core structure (Fig.13).



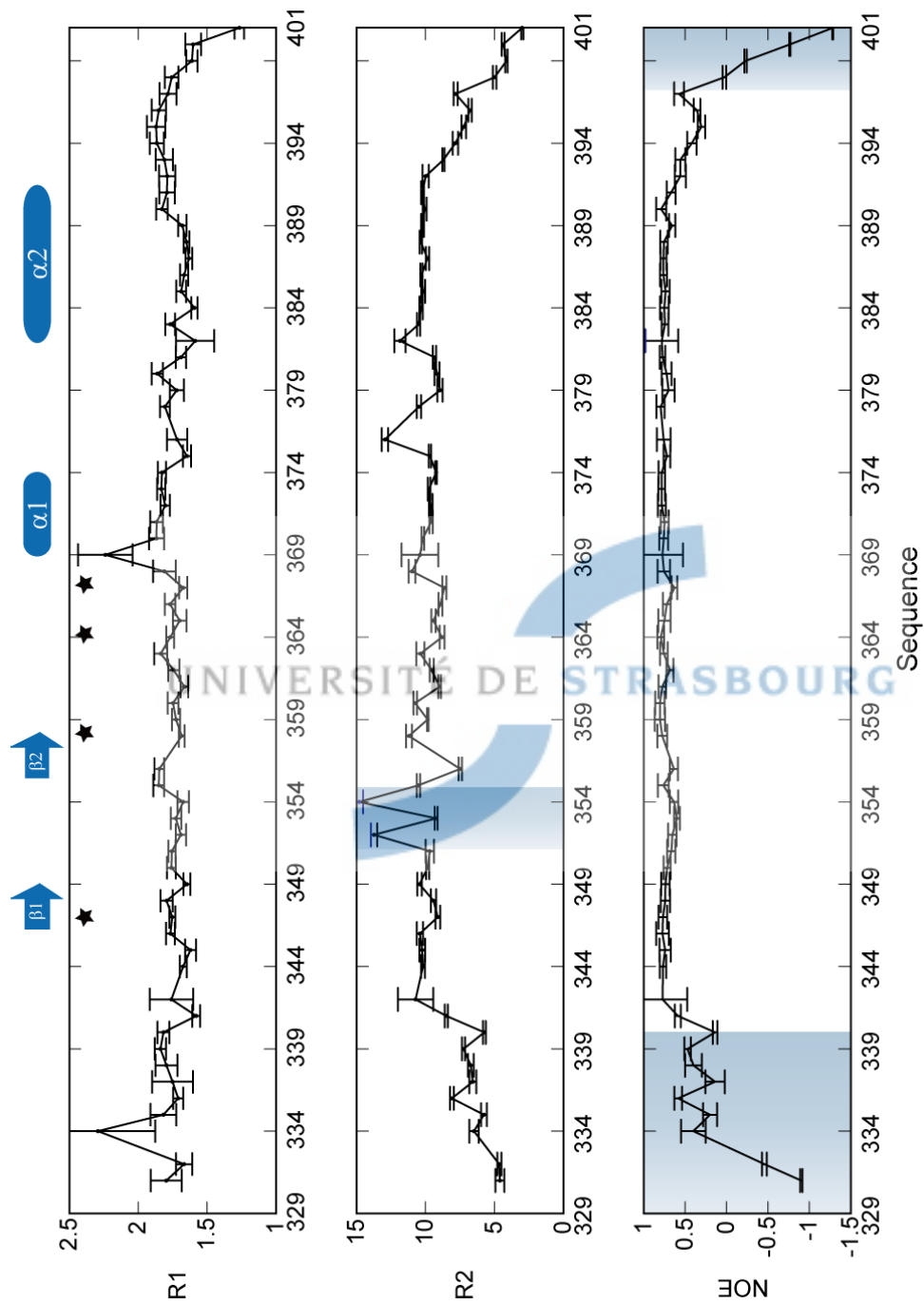


Fig.11 (A) Experimental values of relaxation rates for ATXN7 at 600MHz with respect to protein sequence. The stars indicate residues involved in zinc coordination site.

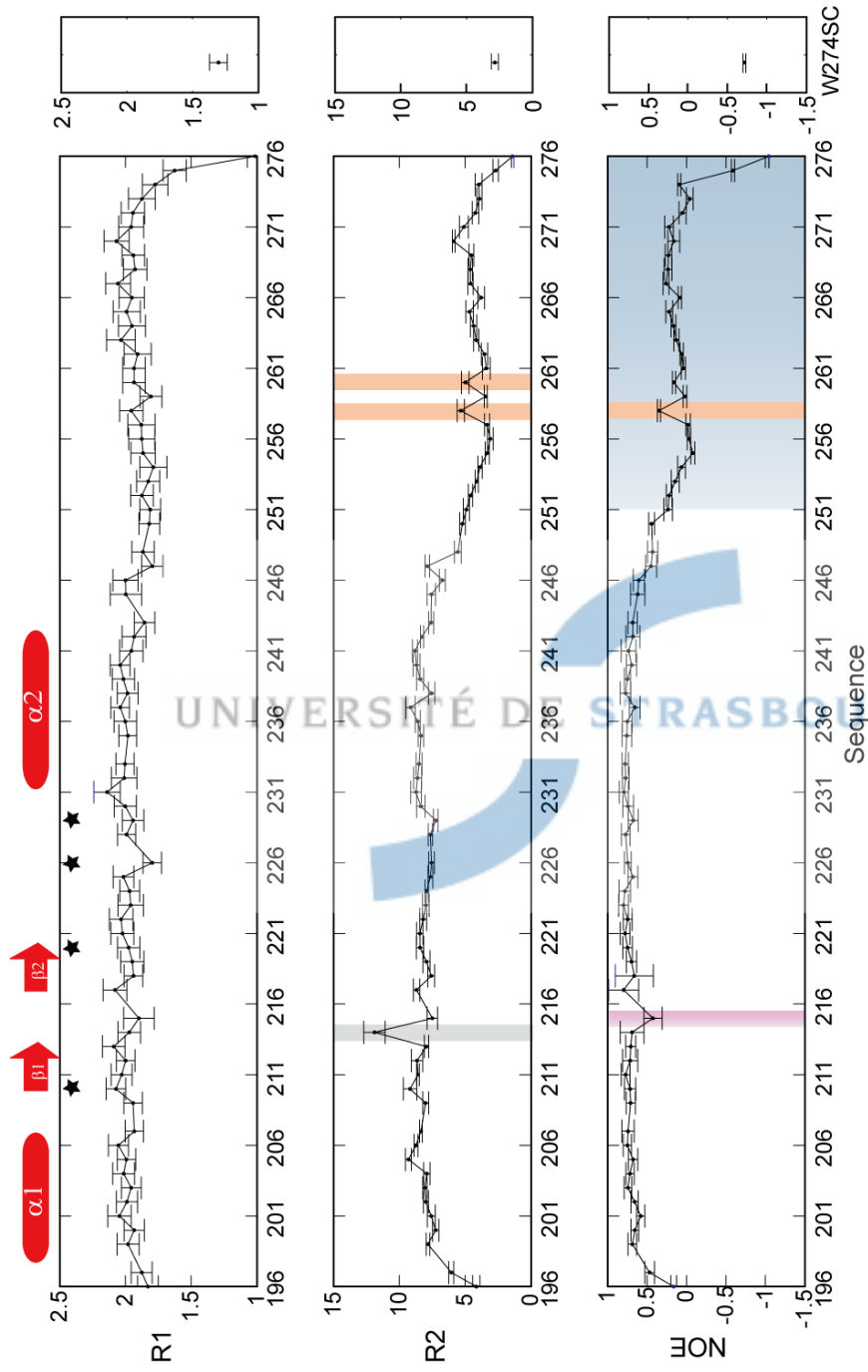


Fig.11 (B) Experimental values of relaxation rates for ATXN7 L3 at 600MHz with respect to protein sequence.

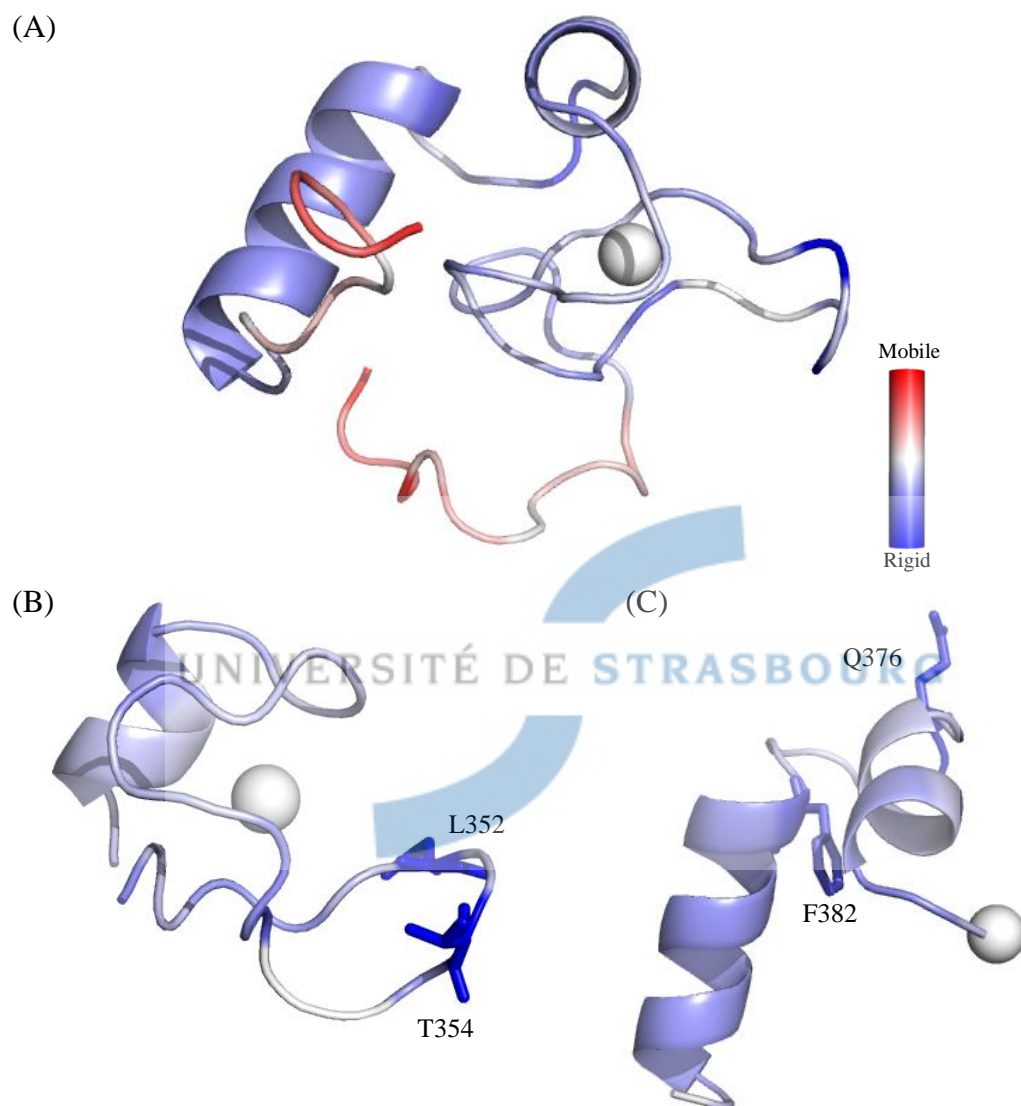


Fig.12 (A) Overall structure of ATXN7. The backbone is colored according to  $J(0)$  values. (B) The zinc coordinating site of ATXN7. The tip of loop (red) has the maximal  $J(0)$  value. (C) The loop region between two  $\alpha$ -helices. Residues Q376 and F382 show elevated  $J(0)$  values, indicating the conformational exchange in the loop region.

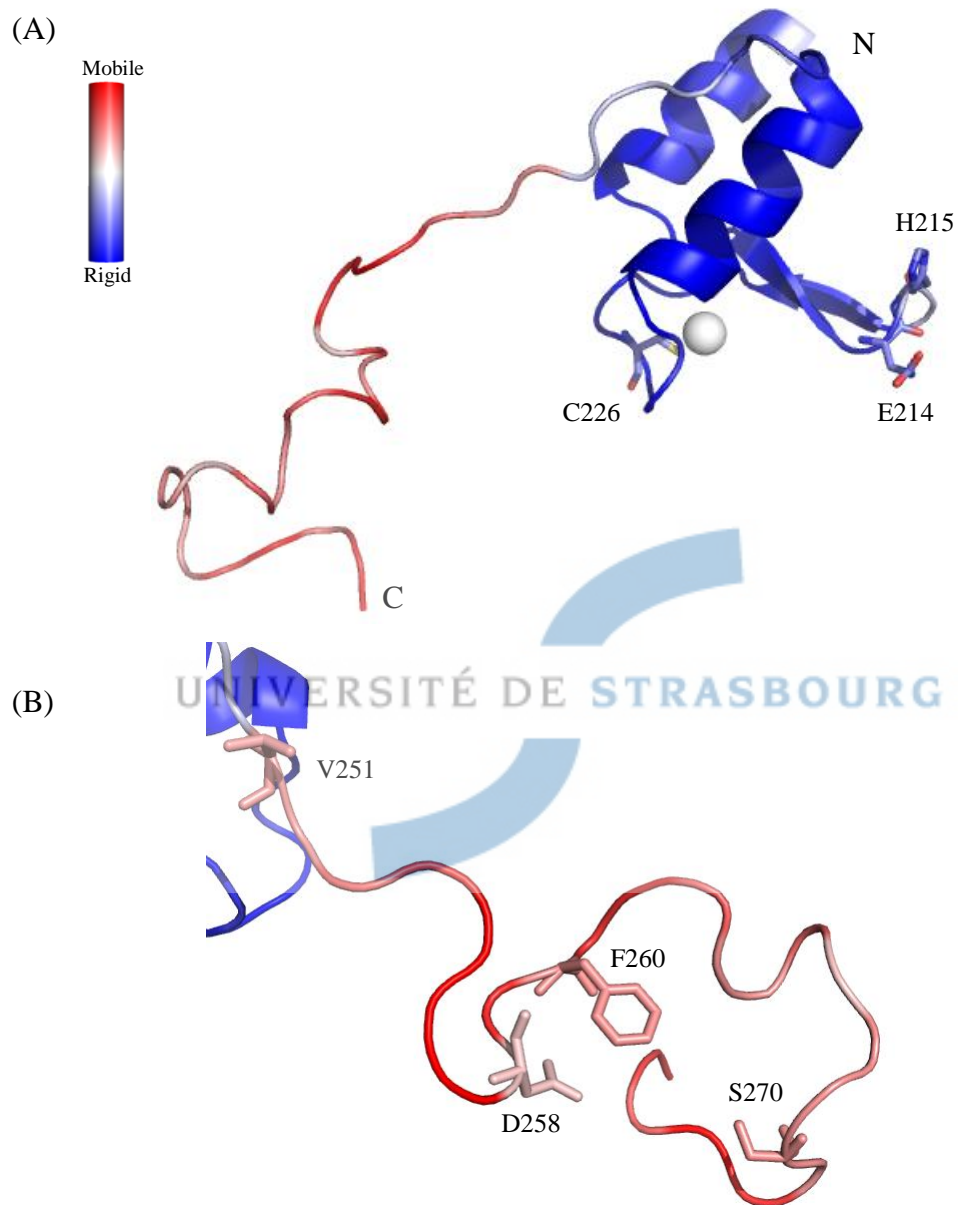


Fig.13 (A) The structure of ATXN7 L3. The backbone is colored according to values of  $^1\text{H}$ - $^{15}\text{N}$  NOEs. The structural core contains two  $\alpha$ -helices whereas the C-terminus is unstructured. (B) The C-terminal tail of ATXN7 L3. The unfolded part starts at V251. Residues D258, F260 and S270 form a fold-like center in the C-terminal tail.

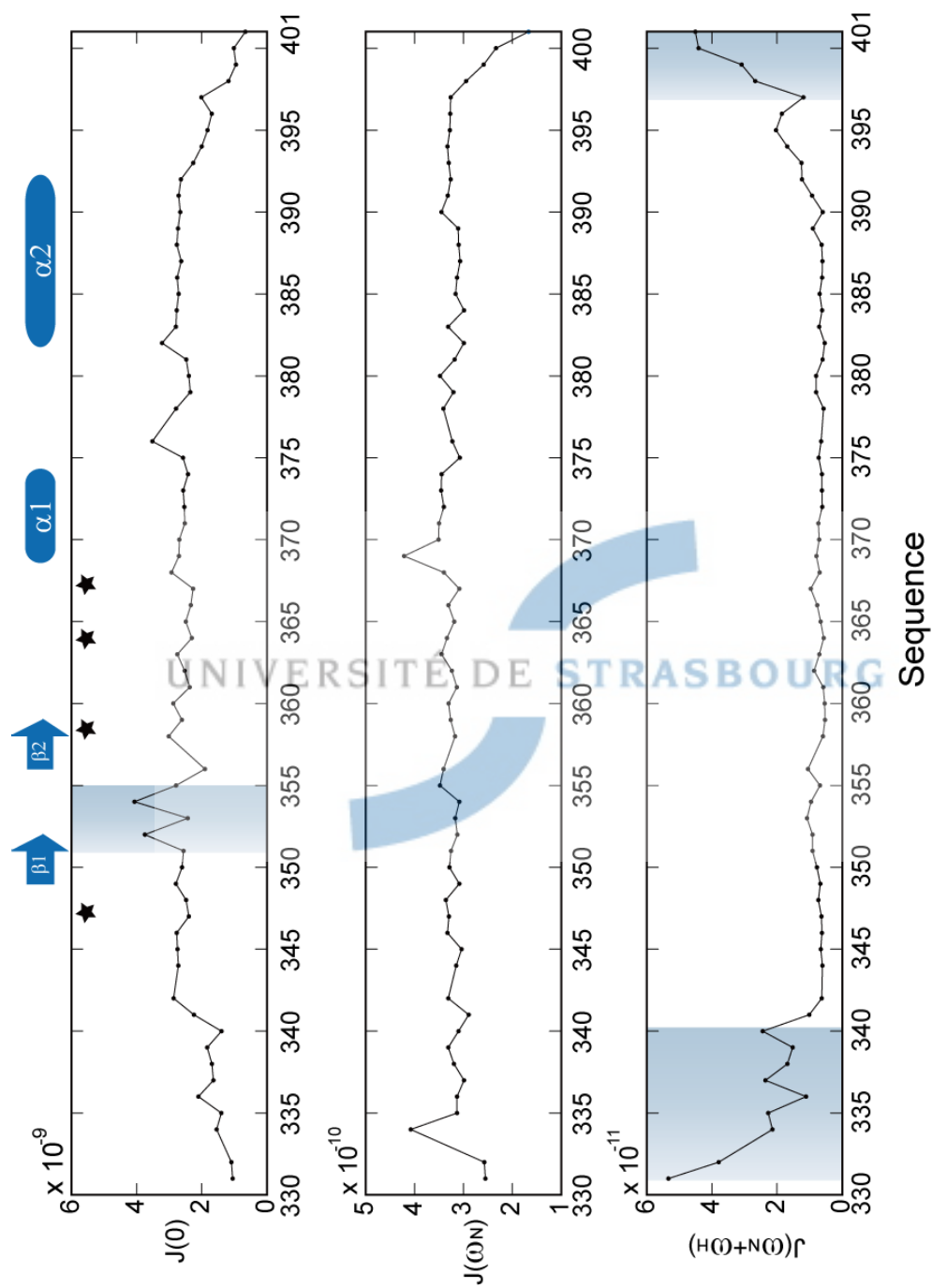


Fig.14 (A) Spectral density functions at 0,  $\omega_N$ , and  $\omega_N + \omega_H$  for ATXN7 at 600MHz with respect to protein sequence. The stars indicate residues involved in zinc coordination site.

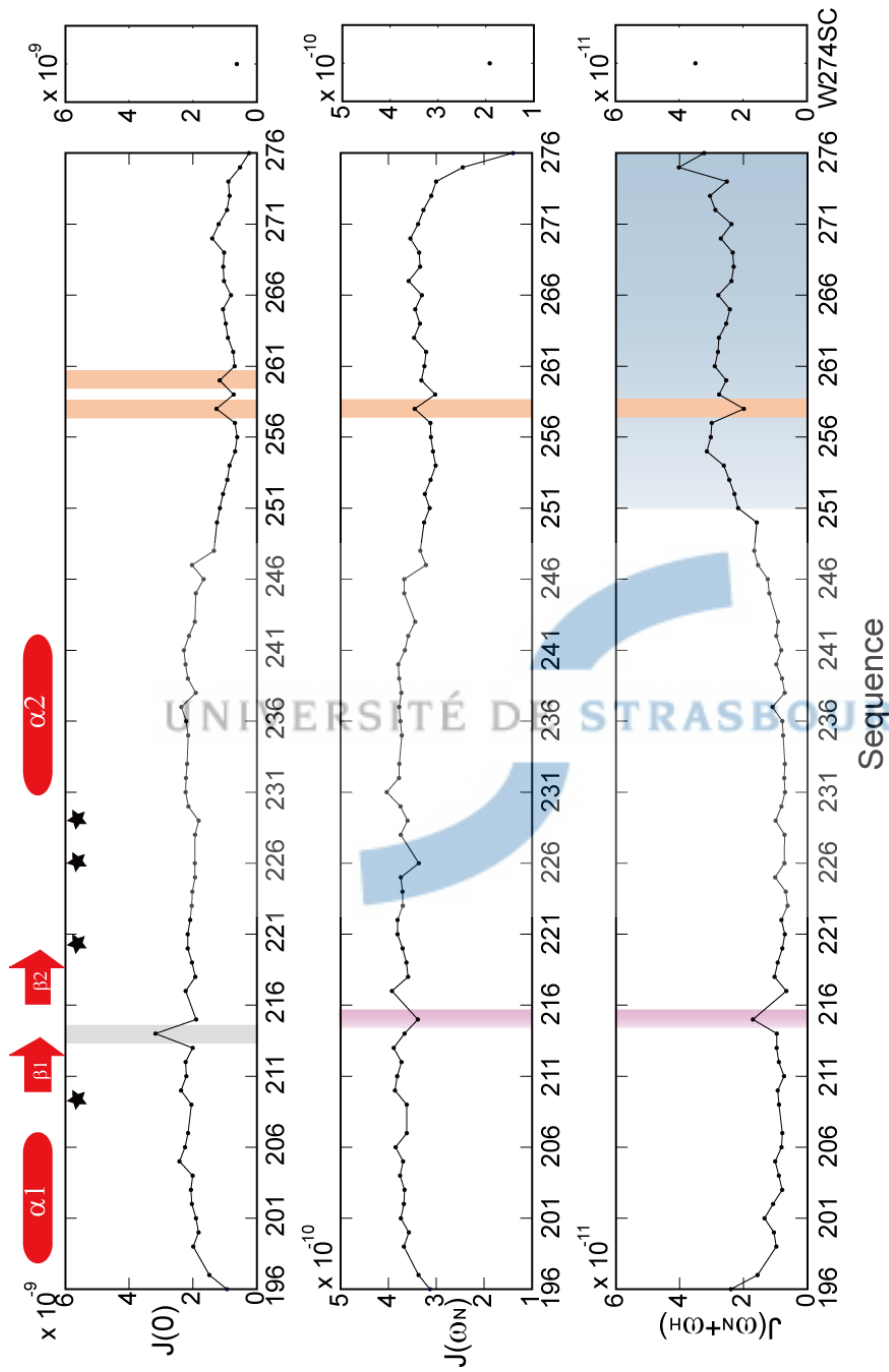


Fig.14 (B) Spectral density functions at  $0$ ,  $\omega_N$ , and  $\frac{\omega_N+\omega_C}{2}$  for ATXN7 L3 at 600MHz with respect to protein sequence. The sequences of secondary structures are shown on the top of figure. The stars represent the residues binding to the zinc ion. The side chain of W274 (W274SC) is also shown in these plots. The stars indicate residues involved in zinc coordination site.

### 3.2.2 Evidences of conformational exchanges

The analysis of  $J(\omega_N)=F(J(0))$  parametric plots (Fig.15A and 16A) revealed a feature that is common to both SCA7 domains. Indeed, several residues of the ATXN7 SCA7 domain showed increased values of  $J(0)$ , and appeared at the extreme right-hand side of the plot, outside the region defined by the isotropic tumbling parametric curve (Shown in green in Fig.15A). Increased values of  $J(0)$  are indicative of the presence of  $\mu$ s-ms time scale conformational exchange. Strikingly, this slow motion contribution was found to affect similar regions of both SAC 7 domains, namely the extremity of the loop between the first two zinc-binding ligands (Fig.15B). Thus exchange contributions were found to contribute to transverse relaxation of L352 and T354 in ATXN7, and in the corresponding position in ATXN7 L3: E214. This observation suggests that the hairpin loop, that protrude into the solvent, and is thus relatively unconstrained, probably sample several local conformations on a slow time scale, a property which is common to both domains, despite the difference in amino-acid composition.

Intermediate ( $\mu$ s to ms) time-scale motions were also found in other parts of the two domains. In ATXN7, residues Q376 and F382 which are located in the loop between two  $\alpha$ -helices and at the head of the C-terminal  $\alpha$ -helix, respectively, clearly showed conformational exchange contribution to their  $J(0)$  values. For ATXN7 L3, the analysis of the  $J(0)$  profile along the protein sequence highlighted some residues with distinct behaviour, in the disordered C-terminal part of the protein, such as D258 and F260, suggesting that high and low frequency motions both contribute to the disorder.



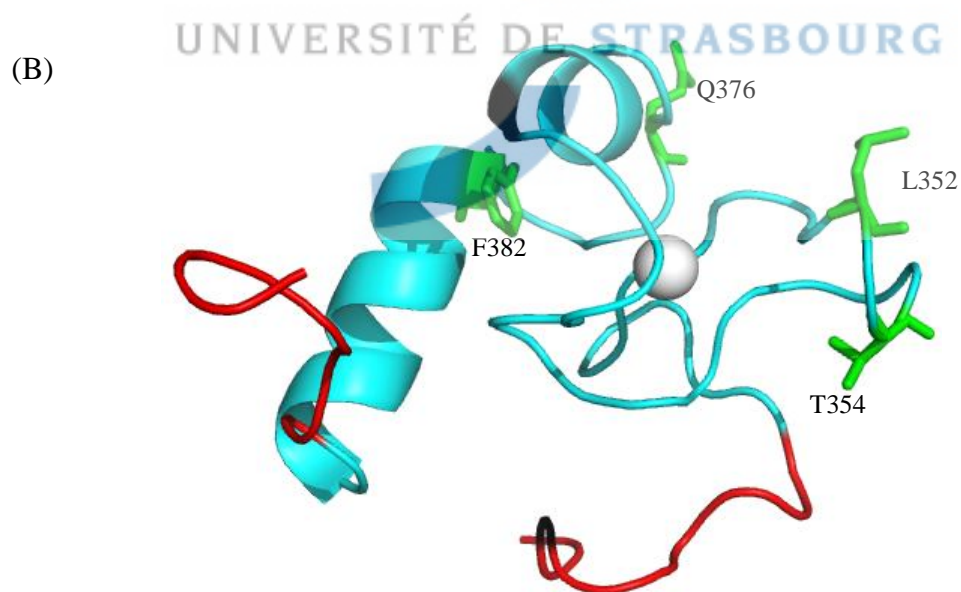
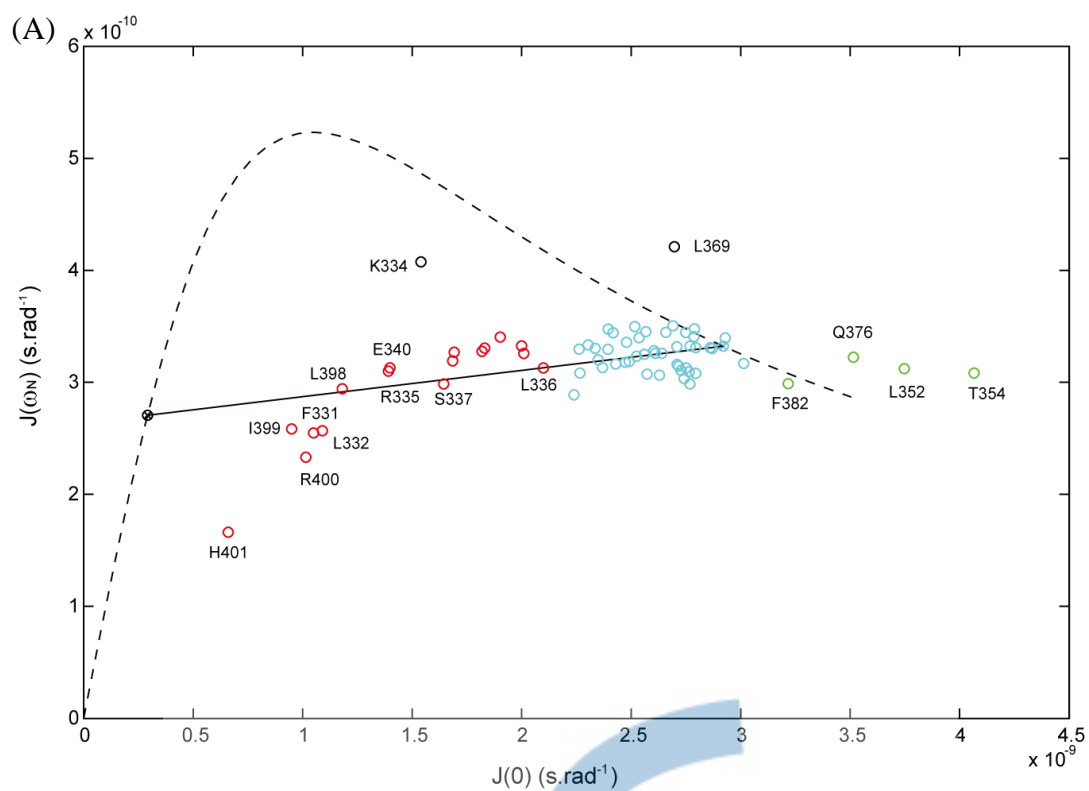


Fig.15 (A) Spectral density plots for ATXN7. The dashed curve represents values of the spectral density calculated at the pairs of  $J(0)$  and  $J(\omega_N)$ , assuming isotropic rigid body rotation. Linear regression fitting to the points of data gives the correlation time for sets of residues. The structural core is represented in cyan; unstructured part in red; slower internal motion in green. (B) Backbone structure of ATXN7. The color of the backbone corresponds to the data points in (A).

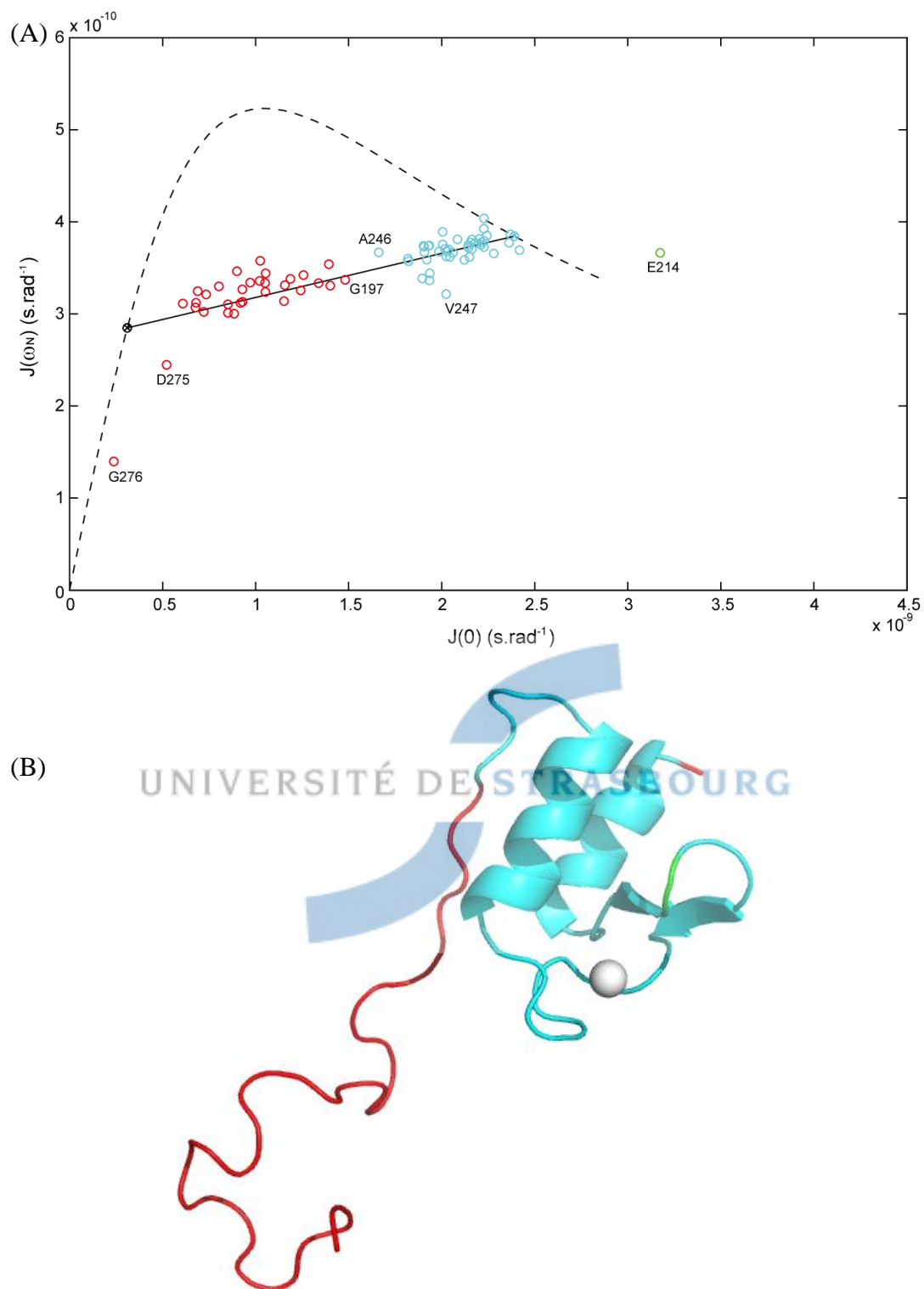


Fig.16 (A) Plots of  $J(\omega_N)$  as a function of  $J(0)$  for ATXN7 L3. The dashed line represents the values calculated at appropriate pairs of frequencies, assuming isotropic rigid body rotation. The data points of the structural core are represented in cyan while those for the unfolded part are shown as red circles. (B) The backbone structure of ATXN7 L3 colored as in A.

### 3.2.3 Analysis of rotational diffusion of SCA7 domains

$R_1$  and  $R_2$  relaxation rates were used as input data for the TENSOR program in order to calculate the orientation of the rotational diffusion tensors. Residues that were observed as rigid on the ps-ns time scale from  $J(\omega_N + \omega_H)$  values and lacking exchange contributions were selected for the calculation. The fully anisotropic diffusion tensors were calculated with a khi2 value of 339.4 for the ATXN7 SCA7 domain and 41.57 for ATXN7 L3. Despite these relatively poor agreements, the orientations of the main axis ( $D_{zz}$ ) of the diffusion tensor were found to be very similar for the two domains, as expected from their structural similarity. Indeed, in both cases, the  $D_{zz}$  axis is nearly parallel to the  $\beta$ -hairpin loop linking the first and second zinc coordinating residues and very close to the direction of the principal axis of the inertia tensor calculated from the rigid parts of the solution structures ( $40.8^\circ$  for ATXN7 and  $40.5^\circ$  for ATXN7 L3)

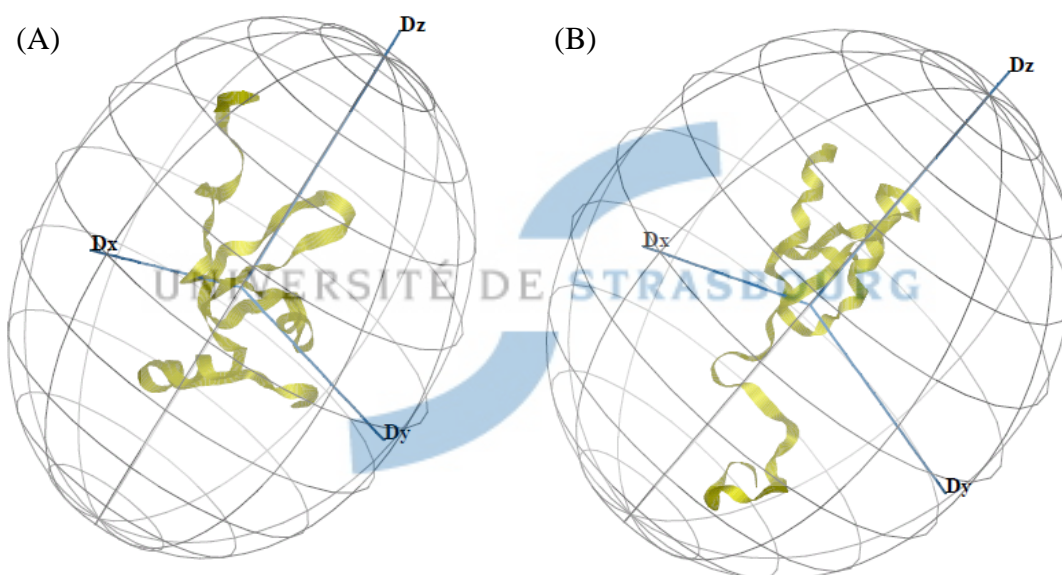


Fig.17 Rotational diffusion tensor of (A) ATXN7 and (B) the core region.

Table 2 Diffusion tensor parameters from  $R_2/R_1$  of ATXN7 and ATXN7 L3

Tensor	$D_{xx}$ ( $10^{-9}s^{-1}$ )	$D_{yy}$ ( $10^{-9}s^{-1}$ )	$D_{zz}$ ( $10^{-9}s^{-1}$ )	$\tau_c$ (ns)	$\chi_2$
ATXN7	2.05	2.45	2.87	6.79	339.4
ATXN7 L3 (core)	2.82	2.95	3.34	5.55	41.6

### 3.2.4 Fast internal dynamics in ATXN7 and ATXN7 L3 domains

The internal dynamics of both SCA7 domains were further investigated by calculating an order parameter ( $S^2$ ) along the sequence using the Lipari-Szabo model. This parameter, whose value ranges between 0 and 1, provides an indication of the amplitude of motions that are faster than the tumbling rate of the protein. Relaxation

rates of most residues could be fitted using one (only  $S^2$ ) or two parameters ( $S^2$  and  $\tau_i$ ) to describe internal motions (colored in sky blue in Fig.18A). Conformational exchange was detected for 17 residues (shown in dark blue in Fig.18A). The average value of  $S^2$  obtained for the core domain (residues F341 to K392) is  $0.8 \pm 0.2$  (Fig.18A). Some residues located within the hairpin loop, such as D353 and K356 exhibited slightly lower order parameters. The profile of  $S^2$  obtained for the N-terminal and C-terminal tails of the domain are significantly different. In the N-terminal part, the  $S^2$  values are distributed around a plateau value of 0.5 from residues 333 to 340, indicating some restriction of these high frequency motions, probably due to interactions between the tail and the core of the domain. In the C-terminal part, the profile of  $S^2$  values displays a continuous decrease from the last ordered core residue (393) to the C-terminal residue, indicating a continuous increase in disorder (Fig.19A).

The profile of order parameters obtained for ATXN7 L3 is remarkably different (Fig.18B). Relaxation rates for most residues could be fitted using one (only  $S^2$ ) or two parameters ( $S^2$  and  $\tau_i$ ) to describe internal motions (colored in sky blue in Fig.18B). Conformational exchange was detected for 26 residues (shown in dark blue in Fig.18B). The average value of  $S^2$  obtained for the core domain (residues G197 to V247) is  $0.9 \pm 0.1$  (Fig.18B). The N-terminal tail is rigid as the domain starts with  $\alpha$ -helix 1, whereas the order parameters display a gradual and continuous decrease from residue 245 to 256, indicating a progressive increase of fast internal motions amplitude. Interestingly, most of the residues in the C-terminal tail were fitted with a two-parameter model ( $S^2$  and  $\tau_i$ ) with an unusually high value of the internal correlation time ( $\tau_i = 1.0$  ns). Moreover, several residues around S270 in the disordered region display increased values of  $S^2$  ( $>0.33$ ) suggesting that local structure may occur within the tail of the ATXN7 L3 protein. Indeed, a set of hydrophobic residues are found in this region, such as F260 and W274, that may well form a molten globule-like structure, to optimize their interaction with the water ( Fig.19B). The transition from the ordered part of the protein to the disordered tail, which is very progressive and smooth, may also be indicative of weak interactions between the disordered tail and the core domain, restricting the disorder. In order to further our understanding of the hydrodynamic behavior of this flexible tail, relaxation data were supplemented with other types of observables that are sensitive to molecular motions.

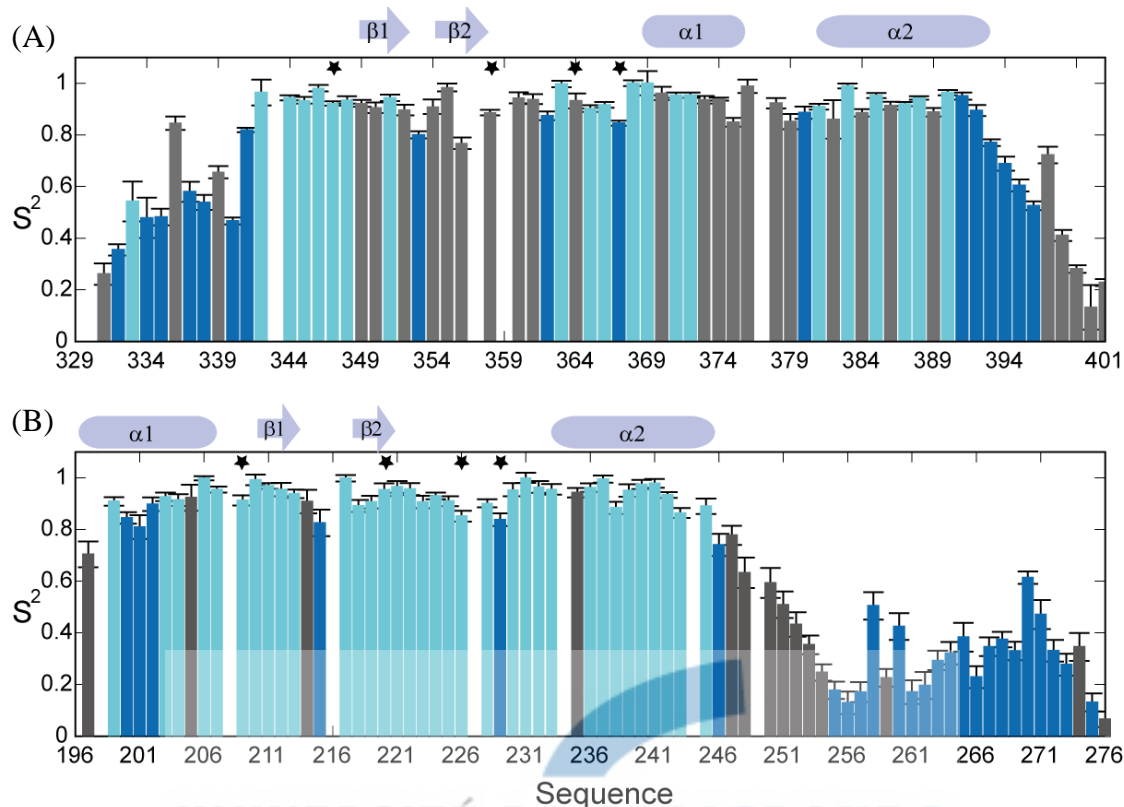


Fig.18 Plots of order parameters of (A) ATXN7 and (B) ATXN7 L3 obtained from the Lipari-Szabo model with variable overall correlation times, versus the sequences. The secondary structures are shown at the top of each plot. The stars represent the residues that coordinate with the zinc ion.

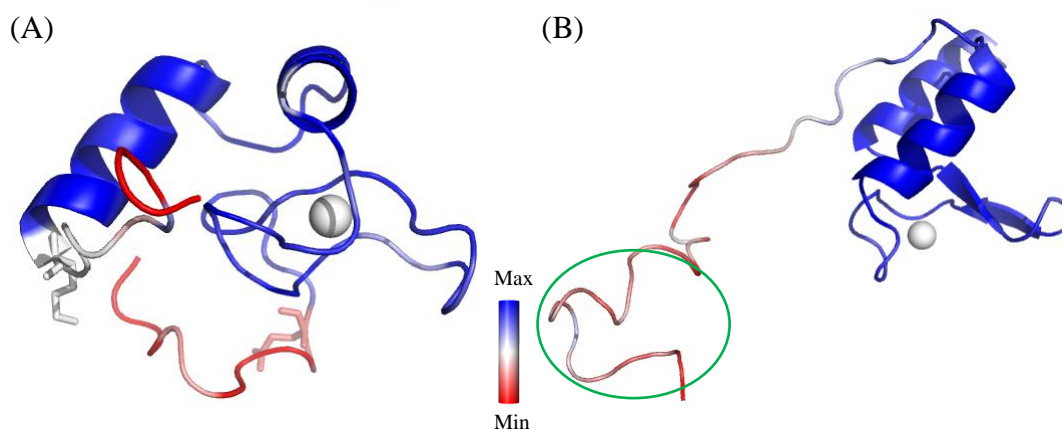


Fig.19 Structures of (A) ATXN7 and (B) ATXN7 L3. The backbone structures are colored according to values of  $S^2$ . For ATXN7 L3, several residues in the unstructured region form another core center which is distinct from the core of the structure (circled). This region was displayed using the results of Lipari-Szabo model-free analysis.

### 3.3 A view on dynamics from the HET-SOFAST experiments

The  $^1\text{H}$ - $^{15}\text{N}$  HET-SOFAST experiment provides a very fast view on the fold of a protein. The observable  $\lambda_{\text{NOE}}$ , which results from the ratio of a  $^1\text{H}$ - $^{15}\text{N}$  correlation peak intensity measured with and without perturbing the magnetization of the aliphatic protons, is sensitive both to the local density of protons and to the dynamics. It was thus interesting to compare the HET-SOFAST profiles obtained for two domains that encompass a common structural motif, but differ in their organization of secondary elements as well as in their dynamic properties.

Fig.20 displays the comparison of  $^1\text{H}$ - $^{15}\text{N}$  HET-SOFAST profiles recorded on ATXN7 and ATXN L3 SCA7 domains. Not surprisingly, the best overlapped is found for the amino-acids located in the zinc coordination site, in particular between the second and the third zinc coordinating cysteine residue. The different folding of the protein chain after the last zinc coordinating residue translates into strikingly different profiles. In ATXN7, solvent exposure of residues within the loop between helices  $\alpha 1$  and  $\alpha 2$  leads to high values of  $\lambda_{\text{NOE}}$ , whereas the corresponding region in ATXN7 L3 which adopts a helical conformation is characterized by low  $\lambda_{\text{NOE}}$  values ( $\lambda_{\text{NOE}} < 0.3$ ). The profile of  $\lambda_{\text{NOE}}$  after the  $\alpha 2$  helix of ATXN7 L3 is particularly informative. This region is reported as mobile from the relaxation data, however, a clear well-shape is observed in this region, with several residues having  $\lambda_{\text{NOE}}$  values below 0.6. The remaining C-terminal region displays  $\lambda_{\text{NOE}}$  values between 0.6 and 0.8, indicative of a disordered peptide chain. A closer view on the slight variations of  $\lambda_{\text{NOE}}$  values in disordered parts of the two domains is also informative. In the C-terminal tail of ATXN7 L3 reduced values of  $\lambda_{\text{NOE}}$  (0.6) are found for residues L268 to S270 corresponding to the site where restriction of motion was detected from the analysis of  $S^2$  values deduced from  $^{15}\text{N}$  relaxation experiments. The same observation holds true for the N-terminal part of ATXN7 domain where few residues display lower  $\lambda_{\text{NOE}}$  values at positions corresponding to slightly restricted internal motion identified from  $S^2$ .

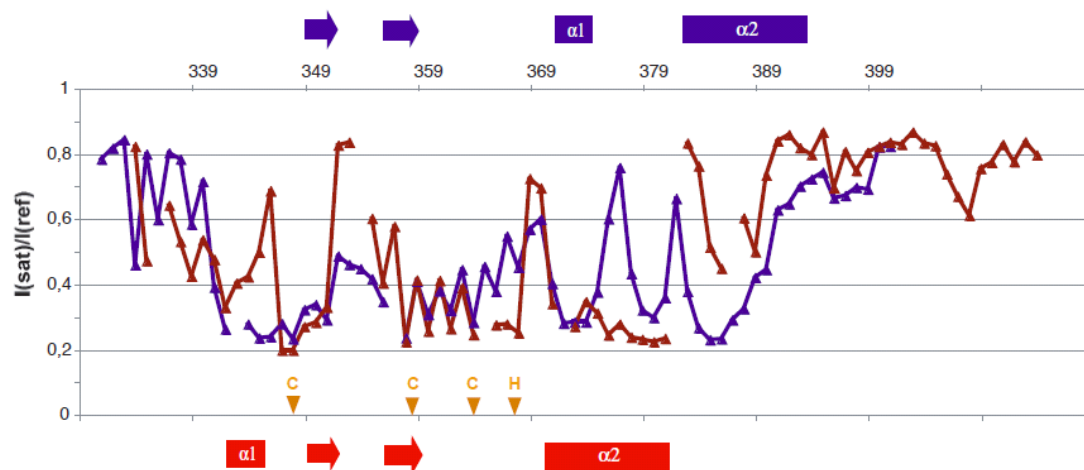


Fig.20 The HET-SOFAST experiments of ATXN7 (blue line) and ATXN7 L3 (red line). The horizontal axis represents the sequence of ATXN7. Secondary structures of ATXN7 and ATXN7 L3 are shown in the top and bottom of the figure, respectively. The *arrows* indicate the positions of residues binding to zinc.

### 3.4 Between order and disorder: combining RDC with dynamics data

Residual Dipolar Couplings are sensitive to the degree of orientation as well as the local order at the position of the residue, allowing motions to be probed over a larger time scale. Fig.21A shows the value of RDCs for each residue in ATXN7 L3. Fluctuation of RDC values between positive and negative values is a signature of a folded domain, where the RDC depends on the orientation of the N-H vector within a frame attached to the core domain. Indeed, a good linear correlation was obtained between experimental RDC values and those recalculated from the NMR structure, which was modeled without taking RDC constraints into account (Fig.21B). Interestingly, such fluctuations are observed for residues located just after the  $\alpha 2$  helix, in the region characterized by a gradual decrease of the relaxation order parameters, indicating an orientation of, at least, the N-terminal region of the tail. Negative RDC values are then observed for a several residues, forming a continuous stretch between residues 251 and 260, which is a signature of disordered regions.[16] Interestingly, residues 265 and 268 display positive RDC values, which correlated to the region where restricted internal dynamics were observed from both HET-SOFAST and relaxation order parameters, providing another clue of a partial folding of this region (Fig.22).



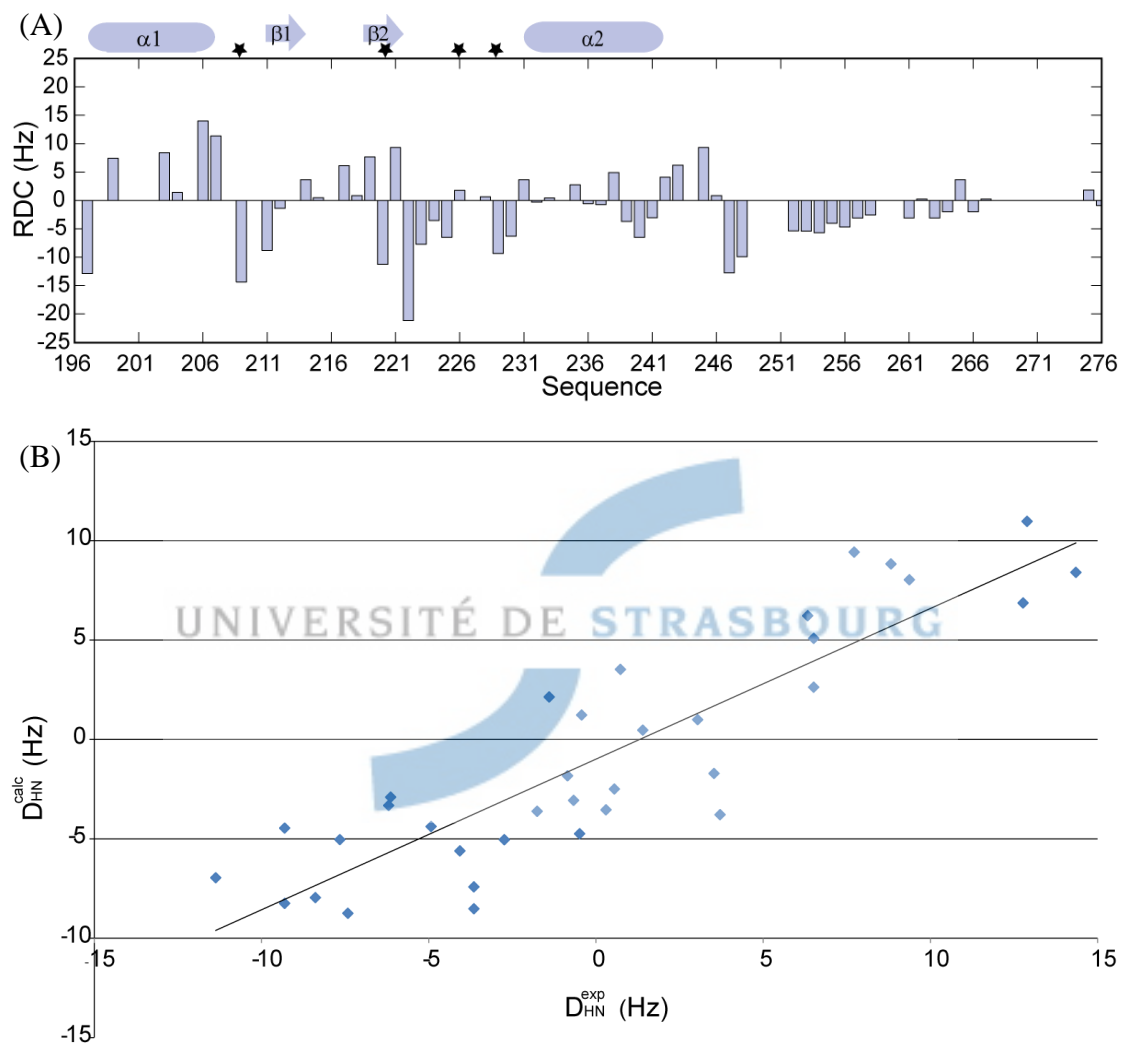


Fig.21 (A) RDCs ( $D_{\text{HN}}^{\text{exp}}$ ) of ATXN7 L3 for each residue. The secondary structure of ATXN7 L3 is shown above the plot. The stars indicate residues involved in zinc coordination. (B) Correlation plot between  $D_{\text{HN}}^{\text{exp}}$  and  $D_{\text{HN}}^{\text{calc}}$  of ATXN7 L3.

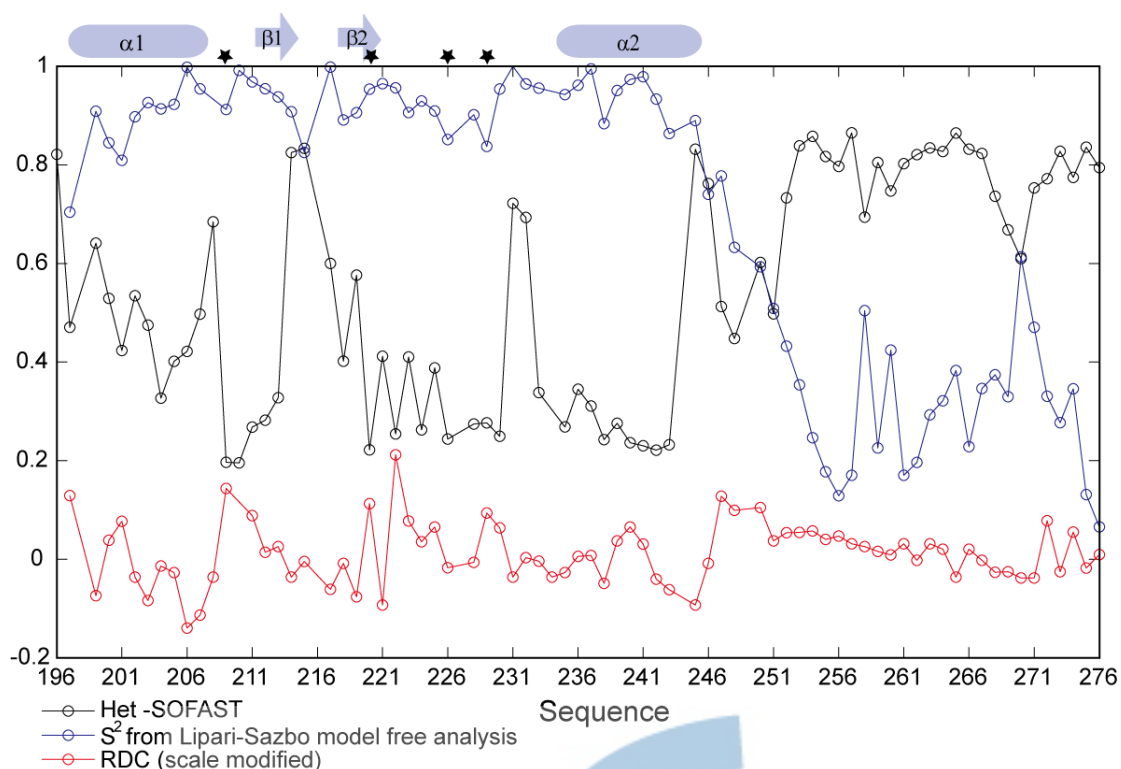


Fig.22 Comparison of three different approaches for the measurement of dynamics. The Het-SOFAST experiments are sensitive to solvent accessibility. The order parameter from Lipari-Sazbo model is sensitive to internal motions faster than the correlation time of global motion. The RDC measurement is sensitive to motions on sub-nanosecond to 50  $\mu$ s time scale.

### 3.5 Conclusions

The ATXN7 and ATXN7 L3 proteins contain a zinc finger domain that is very specific and that is defined by a novel primary sequence signature. Very unexpectedly, these domains differ in the organization of the two helices around an otherwise conserved zinc coordination sphere. We have investigated and compared the dynamic properties of these two domains using both classical  $^{15}\text{N}$  relaxation experiments and more recent approaches including the Het-SOFAST and RDC measurements. The most striking feature that came out of our study is the particular dynamic behavior of the disordered tails that flank the core SCA7 domains. We have found that these tails showed different degrees of disorder, which are interpreted as resulting from transient interactions between these tails and the core domains. Specific dynamic behavior is often the signature of protein interaction spots. Indeed, for ATXN7, preliminary saturation transfer experiments between H2A.H2B dimers allowed us to map the interaction in the N- and C-terminal regions of the ATXN7 SCA7 domain (see Appendix V). For ATXN7 L3, no interacting molecules have yet been identified, but both the sequence conservation properties, together with its dynamic features strongly

suggest that this part of the protein is likely to be involved in intermolecular interactions.

### References

1. Mittermaier A, Kay LE: **Review - New tools provide new insights in NMR studies of protein dynamics.** *Science* 2006, **312**(5771):224-228.
2. Kay LE: **Protein dynamics from NMR.** *Nat Struct Biol* 1998, **5 Suppl**:513-517.
3. Atkinson RA, Kieffer B: **The role of protein motions in molecular recognition: insights from heteronuclear NMR relaxation measurements.** *Progress in Nuclear Magnetic Resonance Spectroscopy* 2004, **44**(3-4):141-187.
4. Gal M, Kern T, Schanda P, Frydman L, Brutscher B: **An improved ultrafast 2D NMR experiment: towards atom-resolved real-time studies of protein kinetics at multi-Hz rates.** *J Biomol NMR* 2009, **43**(1):1-10.
5. Summers MF, South TL, Kim B, Hare DR: **High-resolution structure of an HIV zinc fingerlike domain via a new NMR-based distance geometry approach.** *Biochemistry* 1990, **29**(2):329-340.
6. Egea PF, Rochel N, Birck C, Vachette P, Timmins PA, Moras D: **Effects of ligand binding on the association properties and conformation in solution of retinoic acid receptors RXR and RAR.** *J Mol Biol* 2001, **307**(2):557-576.
7. Sette M, Spurio R, Trotta E, Brandizi C, Brandi A, Pon CL, Barbato G, Boelens R, Gualerzi CO: **Sequence-specific recognition of DNA by the C-terminal domain of nucleoid-associated protein H-NS.** *J Biol Chem* 2009, **284**(44):30453-30462.
8. Tugarinov V, Kay LE: **Relaxation rates of degenerate  $^1\text{H}$  transitions in methyl groups of proteins as reporters of side-chain dynamics.** *J Am Chem Soc* 2006, **128**(22):7299-7308.
9. Lipari G, Szabo A: **Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity.** *Journal of the American Chemical Society* 1982, **104**(17):4546-4559.
10. Peng JW, Wagner G: **Mapping of the spectral densities of N-H bond motions in eglin c using heteronuclear relaxation experiments.** *Biochemistry* 1992, **31**(36):8571-8586.
11. van Heijenoort C, Penin F, Guittet E: **Dynamics of the DNA binding domain of the fructose repressor from the analysis of linear correlations between the  $^{15}\text{N}$ - $^1\text{H}$  bond spectral densities obtained by nuclear magnetic resonance spectroscopy.** *Biochemistry* 1998, **37**(15):5060-5073.

12. Lakomek NA, Lange OF, Walter KF, Fares C, Egger D, Lunkenheimer P, Meiler J, Grubmuller H, Becker S, de Groot BL *et al*: **Residual dipolar couplings as a tool to study molecular recognition of ubiquitin.** *Biochem Soc Trans* 2008, **36**(Pt 6):1433-1437.
13. Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH: **Nuclear Magnetic Dipole Interactions in Field-Oriented Proteins - Information for Structure Determination in Solution.** *Proceedings of the National Academy of Sciences of the United States of America* 1995, **92**(20):9279-9283.
14. Sass HJ, Musco G, Stahl SJ, Wingfield PT, Grzesiek S: **Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes.** *J Biomol NMR* 2000, **18**(4):303-309.
15. Bax A, Grishaev A: **Weak alignment NMR: a hawk-eyed view of biomolecular structure.** *Curr Opin Struct Biol* 2005, **15**(5):563-570.
16. Blackledge M: **Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings.** *Progress in Nuclear Magnetic Resonance Spectroscopy* 2005, **46**(1):23-61.
17. Schanda P, Forge V, Brutscher B: **HET-SOFAST NMR for fast detection of structural compactness and heterogeneity along polypeptide chains.** *Magn Reson Chem* 2006, **44** Spec No:S177-184.
18. Dosset P, Hus JC, Blackledge M, Marion D: **Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data.** *Journal of Biomolecular Nmr* 2000, **16**(1):23-28.
19. Bax A: **Weak alignment offers new NMR opportunities to study protein structure and dynamics.** *Protein Sci* 2003, **12**(1):1-16.
20. Ottiger M, Delaglio F, Bax A: **Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra.** *J Magn Reson* 1998, **131**(2):373-378.
21. Hass MAS, Thuesen MH, Christensen HEM, Led JJ: **Characterization of mu s-ms dynamics of proteins using a combined analysis of N-15 NMR relaxation and chemical shift: Conformational exchange in plastocyanin induced by histidine protonations.** *Journal of the American Chemical Society* 2004, **126**(3):753-765.
22. Lefevre JF, Dayie KT, Peng JW, Wagner G: **Internal mobility in the partially folded DNA binding and dimerization domains of GAL4: NMR analysis of the N-H spectral density functions.** *Biochemistry* 1996, **35**(8):2674-2686.

Altered specificity of a mutated androgen receptor in prostate cancer is associated with a change in the surface potential of the DNA-binding domain

Ying-Hui Wang, Audrey Monge, Jocelyn Céraline, Bruno Kieffer and R. Andrew Atkinson

<sup>1</sup>Département de Biologie et de Génomique Structurales, Institut de Génétique et de Biologie Moléculaire et Cellulaire (Université de Strasbourg / CNRS UMR 7104, INSERM U964), 1 rue  
UNIVERSITÉ DE STRASBOURG  
Laurent Fries, 67404 Illkirch, France.

<sup>2</sup>Equipe Signalisation et Cancer de la prostate (EA4438), Université de Strasbourg, Faculté de Médecine, 1 place de l'Hôpital, BP426 - F67091 Strasbourg cedex

Address correspondence to: Bruno Kieffer, IGBMC, 1 rue Laurent Fries, 67404 Illkirch, France. Fax: +33368854718; E-mail: [kieffer@igbmc.fr](mailto:kieffer@igbmc.fr) or Andrew Atkinson, IGBMC, 1 rue Laurent Fries, 67404 Illkirch, France. Fax: +33368854718; E-mail: [atkinson@igbmc.fr](mailto:atkinson@igbmc.fr)

## Introduction

The androgen receptor (AR) is a ligand-activated transcription factor and a member of the superfamily of nuclear receptors [1,2]. It mediates the biological actions of the male sexual hormone dihydrotestosterone (DHT), including the normal development of the prostate gland [3]. Since AR is also implicated in the progression of prostate cancer, conventional therapy includes androgen ablation, by castration alone or combined with anti-androgens. After an initial clinical response, tumour growth will eventually recur. A number of molecular events are thought to be involved in escape from androgen ablation therapy, including mutations that alter the function of expressed AR [4-6].

AR shares a common structural and functional architecture with other members of the super-family of nuclear receptors. The DNA-binding domain (DBD) and ligand-binding domain (LBD) are preceded by an unusually long A/B domain. AR binds to specific response elements as a homodimer, under the control of dihydrotestosterone binding to the C-terminal LBD. Certain mutations can be rationalised readily in terms of effects on this structural/functional organisation [7]. Thus, in the LBD, the L701H mutation allows activation by adrenocorticoids while mutations of W741 and T877 allow recognition of the anti-androgens bicalutamide and flutamide, respectively, as agonists. Indeed, the T877A mutation is most frequently found in flutamide-treated patients and is thought to arise in response to the selective pressure of the treatment. Nonsense mutations or aberrant pre-mRNA splicing, on the other hand, leads to the expression of a truncated AR that lacks a LBD and is thus constitutively active [8-14]. Yet other mutations affect interactions with co-activator or co-repressor proteins.

The DNA-binding domain interacts with response elements in the promoters of target genes, when AR homo-dimerises in a DNA-dependent manner. However, while AR and other members of the same subgroup of the nuclear receptor super-family recognise the same inverted repeats in such promoters, specificity may arise from binding to direct repeats. AR DBD is expected to bind to inverted repeats in a “head-to-head” manner, as observed for the glucocorticoid receptor. In the only crystal structure of the AR DBD, the homo-dimer binds to direct repeats in the same manner, rather than the “head-to-tail” arrangement observed for the vitamin D receptor binding to direct repeats. Further specificity is thought to derive from the recruitment of co-activator or co-repressor proteins in the presence of agonist or antagonist ligands, respectively. The DBD itself is formed by two C4 zinc fingers in which the first set of four cysteine residues bind one zinc ion while the second set bind another zinc ion (Figure 1A). Within the DBD, residues are conserved to varying degrees and this can be rationalised to some extent in terms of conserved structural (*e.g.*, zinc-binding, hydrophobic core) and functional (*e.g.*, DNA-binding) properties.

Given the primary function of the DNA-binding domain, it is not surprising that mutations may arise in this domain in the context of escape from androgen ablation therapy. In practice, however, the number of reported mutations in the DBD in prostate cancer patients is relatively limited [15]. A

splicing variant that contains an additional 23 amino acids between the two zinc fingers has been found to remain cytosolic upon stimulation with DHT [16]. Among the point mutations that have been identified (T575A, K580R, A586V, A587S and C619Y), the positively-charged side-chain K580 that may interact directly with DNA is replaced by another positively-charged side-chain, while A586 and A587 lie on the back of the  $\alpha$ -helix expected to lie in the major groove, such that the mutations may disrupt the interaction. Both T575 and C619, which is not a zinc-coordinating residue, are located distal to the DNA-binding interface of the DBD and the C619Y mutant has been shown to be unable to bind DNA and to be transcriptionally inactive [17]. The site of the T575A mutation precedes the third Cys residue of the first zinc-binding site and is not expected to influence DNA-binding properties directly. It does, however, lead to a loss of specificity for AR-specific response elements. The mutation is also found coupled to the T877A mutation in the LBD, yielding an AR that is both “promiscuous” in its ligand binding and “unfaithful” in its DNA-binding properties [18].

We set out to characterise the effects of the T575A mutation on the biophysical properties of AR DBD, with the aim of offering an atomic level explanation of effects observed in tissues from prostate cancer patients. Our conclusions suggested further mutational studies, the results of which support the suggestion that the altered specificity of the androgen receptor, mutated at T575 in prostate cancer tissues, is associated with a change of the surface potential of AR DBD associated with a change in the protonation state of His570.

UNIVERSITE DE STRASBOURG



## Results

### 1. Wild-type and mutant structures are highly similar.

Full  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  resonance assignments were obtained for both wild-type and mutated AR DBD and structural restraints were derived from NOESY spectra,  $^1\text{H}$ - $^2\text{H}$  exchange experiments, TALOS software for prediction of backbone  $\Phi$  and  $\Psi$  angles, and measurement of residual dipolar couplings in gels. Analysis of the restraints allowed identification of elements of secondary structure (Figure 1B), which corresponded to those expected for nuclear receptor DBDs. Chemical shift differences between wild-type and mutant domains are concentrated around the site of mutation but are also observed for residues at the end of the second  $\alpha$ -helix (Figure 1B). The three-dimensional structures of AR DBD-WT and ARDBD-T575A are well-defined over much of the primary sequence (Figure 2A, 2B), and this is reflected in low values of the average pairwise r.m.s.d. values calculated using a 5-residue window (Figure 2C, 2D). In both structures, the N-terminal 7 residues were poorly defined due to a lack of restraints, as expected from the dynamic properties of these residues (see below). In contrast, the C-terminus of the domain is relatively tightly defined, and again this correlates with the observed dynamic properties. The first zinc-binding site appears more tightly defined by experimental restraints than the second site, in both wild-type and mutant domains. The tertiary structures of AR DBD-WT and AR DBD-T575A proved to be highly similar (Figure 3A, 3B) with an average pairwise backbone r.m.s.d. value between the two sets of structures of  $1.11 \pm 0.12$ , calculated using the backbone atoms of residues Cys559-Glu588 and Thr602-Thr625. Differences in orientation of the  $\alpha$ -helices are minor and the conformation around the site of mutation is very similar (Figure 3A, 3B). Changes in mean C $\alpha$ -C $\alpha$  distances are limited (Figure 3C) and involve only regions for which the standard deviations on the mean C $\alpha$ -C $\alpha$  distances are elevated. Both structures adopt a fold similar to that of other DNA-binding domains on nuclear receptors and that of rat AR DBD in a complex with a direct repeat of an AR-specific response element.

### 2. Dynamics differences are observed around H570.

$^{15}\text{N}$  relaxation measurements reveal a number of similarities in the dynamic properties of the wild-type and mutant domains (Figure 4). The base-line  $J(0)$  values are similar and reflect the monomeric nature of both AR DBD-WT and AR DBD-T575A domains, tumbling freely in solution. The  $J(0)$  values for AR DBD-WT are found, however, to be systematically higher than those of AR DBD-T575A. This correlates with systematically lower values of  $J(\omega_{\text{N}})$  and indicates a subtle change in the rotational diffusion of the domain upon mutation. Using the Lipari-Szabo approach to analyse relaxation measurements [19,20] with a fully anisotropic description of rotational diffusion, the major axis of the diffusion tensors of AR DBD-WT and AR DBD-T575A were found to match perfectly those of the inertia tensor calculated from the three-dimensional structures. The isotropic correlation time for AR DBD-WT was found to be slightly higher (9.4 ns) than that for AR DBD-T575A (9.1 ns).

Deviations in the relaxation rates and values of the spectral density,  $J(\omega)$ , at all three frequencies

are characteristic of ps-ns time-scale backbone motions, commonly observed in less structured portions of polypeptide sequences. Such deviations are noted for the N-terminal 7 residues and the C-terminal residue, in both wild-type and mutant domains. This confirms that the lack of structural definition determined above for these residues arises from dynamic properties rather than experimental short-comings. Increases in  $J(0)$  reflect rapid transverse relaxation due to conformational exchange on the  $\mu$ s-ms time-scale. Residues Thr602-Cys611 span the polypeptide sequence between the first and second pairs of zinc-coordinating cysteine residues of the second zinc-binding site. While it is difficult to determine the precise nature of the conformational exchange phenomenon affecting these residues, it is clearly common to both wild-type and mutant domains.

Differences in dynamic properties are apparent for a small number of residues close to the site of mutation. In AR DBD-T575A, Cys569 is characterised by increased motion on the ps-ns time-scale, while the increased  $J(0)$  value for His570 and Gly572 indicates  $\mu$ s-ms time-scale conformational exchange affecting these residues.

### 3. The protonation state of H570 is altered upon mutation.

To attempt to understand the origin of observed changes in backbone dynamics at residues Cys569, His570 and Gly572 upon modification of the side-chain at position 575, we examined the protonation state of the histidine side-chain in wild-type and mutant domains, under the same conditions. The chemical shifts of  $^{15}\text{N}$  nuclei in histidine side-chains are particularly sensitive probes of protonation state. The pattern of such shifts in long-range HSQC spectra show that at pH 6.5 His570 is doubly protonated, and therefore positively charged, in the wild-type domain, while the same residue is singly protonated (neutral) in the mutant domain (Figure 5). Clearly, the degree of protonation *in situ* will be sensitive to local conditions, particularly pH, but the difference in  $pK_a$  is expected to yield different degrees of protonation under physiological conditions.

Examination of the three-dimensional structures of AR DBD-WT and AR DBD-T575A reveals that the side-chains of His570 and Thr575 (or Ala575) are in close proximity (Figure 3A) and it seems reasonable, in the case of the wild-type domain, to suggest the presence of a hydrogen bond between the side-chains of the two residues. Mutation to alanine removes the possibility of hydrogen bond formation and this is reflected in an altered protonation state of His570. The absence of the hydrogen bond may then also lead to increased ps-ns time-scale motion for Cys569 and His570.

### 4. Hydrogen bonding strength is altered upon mutation

In well-structured domains, backbone amide hydrogen atoms are protected from hydrogen/deuterium exchange when they are inaccessible to solvent and/or are involved in hydrogen bonds. Typically, in well-structured domains, the cross-peaks of the most protected amide groups show no decrease in intensity over a period of hours or days. For both AR DBD-WT and AR DBD-T575A, no backbone amide hydrogen atoms exhibited this level of protection from exchange:

many had fully exchanged at the first time-point and the cross-peaks of the remainder were seen to decay during the time-course of the experiment (12-18 hours). The set of protected amide hydrogen atoms was identical for the wild-type and mutant domains indicating an identical set of hydrogen bonds, including those in the two  $\alpha$ -helices. The rates of exchange, however, were systematically faster for AR DBD-T575A (Figure 6), generally by a factor of 2-3, suggesting a general destabilisation of the domain upon mutation.

## 5. Impact of H570A mutation on AR transcriptional activity

To sustain our hypothesis, that the effect of mutation is to alter the charge of His570 through the loss of a stabilising interaction, we investigated whether replacing His570 in full-length AR by an uncharged residue orientates AR binding to non-specific responsive elements. Luciferase assays were performed in prostate cell line LNCaP cells by using with a specific and non-specific AR promoter reporter constructs (Figure 7). The higher transcriptional activity on the MMTV-luc non-specific reporter construct of the T575A AR compared to wild type AR has already been described [18]. The activity of H570A AR on the same construct was 1.5-fold higher to that one of the wild type AR, but was reduced to background levels on the PSA61 specific reporter construct, reflecting an inversion of specificity with respect to the wild-type protein.

## Discussion

In attempting to rationalise the clinical effects of a point mutation such as T575A from biophysical measurements, it is important to consider the context of the mutation. DNA-binding domains of nuclear receptors are very highly conserved across species and well-conserved across the set of receptors of a particular species (Figure 1A). This results in a conserved fold around two zinc ions and a common mode of interaction with DNA. The set of residues that are strictly conserved across the super-family of human nuclear receptors contains, not surprisingly, the eight coordinating cysteine residues (Figure 8A). The remainder of the strictly and very highly conserved residues are gathered rather centrally and include side-chains fulfilling a structural rôle, such as the pair of phenylalanine residues on the interior of the first  $\alpha$ -helix, and positively charged side-chains on the same  $\alpha$ -helix that interact in the major groove of the DNA (Figure 8B). Analysis of sequence alignments also reveals blocks of residues that are conserved only among the steroid receptors. These blocks are situated more peripherally than the residues conserved across all DNA-binding domains: some are seen to cluster around the zinc binding sites, while another set are grouped in the loop connecting  $\alpha$ 1 to the second zinc binding site (Figure 8C). With the possible exception of Tyr593, none of the residues in this latter group are expected to contribute directly to the interaction with DNA. A further set of residues are seen to be specific to AR, with amino acid types that differ from the other steroid receptors and that are not observed frequently at corresponding positions in other classes of receptor: Thr558, Ala573, Ala587, Lys590, Gln591,

Lys592, Ser597, Thr602, Phe606, Ser613 and Glu621. None of these sites contributes directly to interactions with DNA and most are situated on the opposite face of the domain to  $\alpha 1$  (Figure 8D).

It is rather difficult to rationalise AR specificity for DNA response elements in terms of a precise set of interactions. Since the blocks of residues specific to steroid receptors are associated with similarity of ligand, these may reflect evolutionary divergence of nuclear receptor classes rather than factors contributing to DNA recognition. The clustering of these blocks around the zinc binding sites, however may suggest a rôle in modulating the stability of the sites. The set of residues specific to AR are largely polar or charged (with the notable exception of Phe606) and may correspond to sites that define interactions with the ligand binding domain of AR or with other proteins. Some of these residues lie within the steroid receptor-specific blocks, suggesting AR-specific modulation of the rôle of these blocks. AR specificity therefore seems to be rather subtly defined: it can, however, be altered in cancerous cells by a single mutation in the DNA-binding domain.

The results of this study suggest several possible effects at the atomic level that may contribute to reduced specificity of AR for response elements. Firstly, the increased ps-ns time-scale motions, if quenched upon binding to DNA sequences, incur a small entropic penalty for the mutant domain which should raise the dissociation constant. Secondly, the reduction of charge on the histidine side-chain may have direct or indirect effects on binding to DNA. The importance of the protonation state of this residue was noted in the study of the effects of pH on the DNA-binding properties of the glucocorticoid receptor [21]. While His570 is not expected to make contacts with DNA that define specificity, the altered charge may reduce affinity in a non-specific manner. These considerations do not explain observed activities satisfactorily and suggest that the reduction in specificity is effected via a third party which could be either another portion of the AR sequence or another protein. A final possibility, however, is that the reduced overall stability of the fold (as reflected in hydrogen/deuterium exchange rates) may alter the populations of receptors competent for interaction with different response elements.

In a recent elegant study of the binding of the glucocorticoid receptor to DNA sequences, Miejsing et al. suggested an alternative viewpoint: “that DNA functions as a sequence-specific allosteric ligand of GR that tailors the activity of the receptor towards target genes”. A similar crystallographic study of AR would provide valuable insight into AR function and allow direct comparison of AR and GR response to DNA-binding. In this context, the analysis of the T575A mutant might shed further light on the effects of this mutation, although if the effects are as subtle as they appear to be from our data, structural data may not provide a full explanation. NMR spectroscopy may again contribute by allowing changes in dynamics and stability of the DNA binding domain to be monitored as a function of DNA sequence.

## Materials and methods

### Sample preparation

AR DBD-WT (AR<sub>555-627</sub>) and AR DBD-T575A were cloned into pETM<sub>30</sub> containing the GST tag. 50 [L of BL21 (DE3) *star* competent cells were transformed with plasmid DNA by heat-shock process. The proteins were expressed in *E.coli* BL21 (DE3) *star* cells. <sup>15</sup>N- and <sup>15</sup>N,<sup>13</sup>C-labelled proteins were produced in M9 minimal medium supplemented with 1.0 g.L<sup>-1</sup> of <sup>15</sup>NH<sub>4</sub>Cl and 2.0 g.L<sup>-1</sup> of <sup>13</sup>C glucose. Unlabelled proteins were expressed in LB medium. Single colonies of transformed BL21 (DE3) *star* cells were grown overnight in 200 mL LB medium at 37°C, with 50 μL.mL<sup>-1</sup> kanamycin. When the cells had grown to an OD<sub>600</sub> of 0.6, expression was induced by 1mM IPTG at 18°C and incubated overnight. The expressed cultures were harvested by centrifugation at 4000 rpm for 20 min at 4°C. Pellets were stored at -20°C.

Pellets were lysed in 40 mL binding buffer (50 mM Tris-HCl at pH 7.5, 150 mM NaCl, 10% glycerol, 1 mM DTT), sonicated and ultracentrifuged at 40000 rpm for 1 hour at 4°C. The fusion protein was incubated with resin for 2 hours at 4°C and washed with washing buffer (50 mM Tris-HCl at pH 7.5, 150 mM NaCl, 1 mM DTT). GST was cleaved from the fusion protein by addition of TEV protease and incubation at room temperature for 2 hours. The protein was eluted with the washing buffer and purified on a Superdex G75 (HiLoad 16/60) gel filtration column, equilibrated with buffer (50 mM phosphate at pH 6.5, 150 mM NaCl, 1 mM DTT). Fractions were concentrated to a volume of 1 ml using 5K amicon (Vivaspin) and transferred to 50mM phosphate pH 6.5 and concentrated further to 500 μL. Final concentrations of AR DBD-WT and AR DBD- T575A were about 600 μM and 1 mM for labelled and unlabelled samples, respectively.

### NMR spectroscopy

Most NMR spectra were recorded at 600 MHz (<sup>1</sup>H frequency) on a Bruker DRX600 spectrometer equipped with a triple-resonance cryoprobe with z-gradients. All spectra were recorded at 15°C, unless specifically stated otherwise. One-dimensional <sup>1</sup>H spectra were recorded with appropriate spectral widths for the sample, using a WATERGATE sequence [22] for water suppression. Additional homonuclear TOCSY and NOESY spectra were recorded at 800 MHz.

<sup>1</sup>H-<sup>15</sup>N HSQC were recorded with spectral widths set to avoid folding of <sup>15</sup>N resonances. Typically, 256 t<sub>1</sub> increments were acquired and quadrature was achieved using the method of States et al. [23]. Spectra were processed applying a 90°-shifted sine bell and zero-filling at least to the next power of two in each dimension prior to Fourier transformation. Spectra were processed using NMRPipe [24] and analysed in CARA [25] and the NEASY module of CARA. <sup>15</sup>N-edited TOCSY and NOESY were recorded with 50 increments in the indirect <sup>15</sup>N dimension and 320 increments in the indirect <sup>1</sup>H dimension. Spectra were processed as above with linear prediction in the indirect <sup>15</sup>N dimension.



A set of three-dimensional triple-resonance experiments (HN(CO)CA, HNCA, HN(CO)CACB, HNCACB, HNCO) were recorded to obtain backbone resonance assignments. Spectra were processed as above with linear prediction in the indirect  $^{13}\text{C}$  dimension. Predictions of backbone  $\Phi$  and  $\Psi$  angles were obtained from resonance assignments using the programme TALOS [26]. Residual dipolar couplings for backbone  $^1\text{H}$ - $^{15}\text{N}$  pairs were measured in polyacrylamide gels using IPAP pulse sequences [27].

Three-dimensional structure determination was performed using the semi-automatic ATNOS/CANDID procedure [28,29], with XPLOR-NIH [30,31] as molecular dynamics programme. Hydrogen bonds that could be unambiguously identified from hydrogen/deuterium exchange data and NOE patterns were introduced, together with constraints on dihedral angles predicted with a high degree of confidence by TALOS. The resulting structures were refined in XPLOR-NIH using a standard protocol (refine.inp), modified to start from a higher initial temperature (2000K), with a larger number of cooling steps (10000 steps) and a longer final energy minimisation step (1200 steps). The structures were then refined using parallhdg5.3 parameters and stereospecific assignments made where possible, before final refinement in explicit solvent [32,33]. Local and global pairwise r.m.s.d. values were calculated in XPLOR-NIH using in-house scripts. Figures were produced using Pymol [34].

$^{15}\text{N}$   $R_1$  and  $R_2$  relaxation rates were measured using pulse sequences based on those of Farrow et al. [35], with 8 scans per increment and 200 increments in the indirect dimension, with a delay of 2.2 s between scans. For  $^{15}\text{N}$   $R_1$  relaxation, intensities were extracted from a set of 11 spectra recorded with values of the longitudinal relaxation delay of 0, 150.5, 301.0, 451.4 (x2), 601.9, 752.4, 902.9, 1053, 1204 and 1354 ms, with  $180^\circ$  proton pulses every 2 ms to suppress cross-correlated relaxation [36]. For  $^{15}\text{N}$   $R_2$  relaxation, intensities were extracted from a set of 12 spectra recorded with values of the transverse relaxation delay of 0, 15.8, 31.7, 47.5 (x2), 63.4, 79.2, 95.0, 126.7, 158.4, 190.1 and 221.8 ms, with  $^{15}\text{N}$   $180^\circ$  pulses applied every 1.2 ms at a field strength of 4.2 kHz and  $^1\text{H}$   $180^\circ$  pulses applied every 2.4 ms to suppress cross-correlated relaxation. [36] Exponential decay rates were obtained from a non-linear least-squares two-parameter fit using a Levenberg-Marquardt algorithm implemented in Matlab (The Mathworks, Inc.).

$^1\text{H}$ - $^2\text{H}$  exchange rates of exchangeable hydrogen atoms were measured by diluting  $^{15}\text{N}$ -labelled samples 10-fold into  $^2\text{H}_2\text{O}$  and recording  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra over the following 14-18 hours. Intensities were fitted using a single exponential as for relaxation data. Long-range HSQC experiments were performed to correlate the side-chain  $^{15}\text{N}$  resonances of His570 with the non-exchangeable  $^1\text{H}$  resonances of the side-chain.

### Cell lines and cell cultures

The human androgen-sensitive prostate cancer cell line LNCaP clone FCG (European Collection of Cell Cultures, Salisbury, UK) (ECACC 89110211, passage number 08-11) was cultured in phenol

red-free RPMI 1640 medium supplemented with 10% fetal calf serum (FCS), 2 mM glutamine, 10 mM HEPES, 1 mM sodium pyruvate (Sigma-Aldrich), 100 U/ml penicillin and 100 µg/ml streptomycin (Invitrogen). Cells were cultured in a humidified atmosphere enriched in 5% CO<sub>2</sub> at 37°C.

## Plasmids

The pEGFP-C3 plasmid was from BD Clontech (Ozyme, France). The pEGFP-AR-WT construct was as described previously [9]. The pEGFP-AR-T575A expression plasmid was derived from pEGFP-AR-WT by replacing the wild type BstEII - BamHI fragment with the corresponding fragment from pAR-T575A [18]. The pEGFP-AR-H570A expression plasmid was constructed by site-directed mutagenesis performed on pEGFP-AR-WT. The MMTV-Luc reporter plasmid containing steroid hormone responsive elements was a gift from Prof. P. Chambon (IGBMC, Illkirch, France). The PSA-61-Luc reporter plasmid containing a 6-kb PSA promoter fragment with the three AREs (AREI, AREII, and AREIII) was from Dr. Trapman (Erasmus University, Rotterdam). The control luciferase reporter vector pGL4.70[hRLuc] was purchased from Promega (Promega, Charbonnières, France).

## Transfection and Luciferase reporter assay

Transient transfection was performed in LNCaP cells by using the JetPEI reagent (PolyPlus Transfection, Polyplus, France) according to instructions provided by the manufacturer. LNCaP cells were seeded in 96-well plates (10<sup>4</sup> cells/well) and transfected for 48 hours with 75 ng of the indicated pEGFP plasmid plus 150 ng of the reporter plasmid and 10 ng pGL4.70[hRLuc]. Twenty-four hours later, cells were switched to phenol red- and serum-free medium containing 100 nM DHT (Sigma-Aldrich, Saint-Quentin Fallavier, France) or vehicle. Luciferase activity was measured 48 hours later with the Dual-Glo luciferase reporter assay kit (Promega, Charbonnières, France). Transcriptional activity of the wild type AR was considered as control. Statistical analysis was performed with Graphpad Quickcalcs (Graphpad.com) using a paired t test from at least two independent experiments. A P value of less than 0.05 was considered significant

## Figure legends

Figure 1. A) Nuclear receptor DNA-binding domain sequences, grouped according to sequence conservation with respect to the sequence of AR DBD. An asterisk marks the site of mutation. Above the sequences is the change in accessible surface area per residue: values range from 0.0 to 0.35. Accessible Surface Areas (ASA) values were calculated for each residue as a fraction of the total surface area of each amino acid type, for one monomer in the crystal structure (1R4I.pdb) in the presence and absence of the double-stranded DNA. Below the sequences are indicated the  $\alpha$ -helices, inferred for AR DBD from chemical shifts and slowly exchanging <sup>1</sup>H<sub>N</sub> resonances and the set of cysteine residues that coordinate the two zinc ions. The colours of the zinc-binding sites and elements of secondary structure are retained in subsequent figures. B) Composite changes in chemical shift



upon mutation of T575 to alanine, calculated for  $^1\text{H}_\text{N}$  and  $^{15}\text{N}$  resonances (upper panel) and for all assigned  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  resonances (lower panel). The central portion shows the primary sequence of the exact construct of AR DBD studied here. The  $\alpha$ -helices and the sets of cysteine residues that coordinate the two zinc ions are indicated. An asterisk marks the site of mutation.

Figure 2. Bundles of 10 representative structures calculated for A) AR DBD-WT and B) AR DBD-T575A. The zinc-binding sites and elements of secondary structure are coloured as in Figure 1 and labelled. For clarity, residues preceding Gln6 are not shown. An asterisk marks the approximate site of mutation (see Figure 3A). A measure of local backbone definition is given by average pairwise r.m.s.d. values calculated over five-residue segments of the primary sequence for 20 calculated structures of C) AR DBD-WT and D) AR DBD-T575A. Residues mentioned in the text are labelled. An asterisk marks the site of mutation.

Figure 3. A) Two views of the superposition of representative structures of AR DBD-WT (grey) and AR DBD-T575A (deep red) with the zinc ions represented by spheres and the heavy atoms of the zinc-coordinating cysteine residues. In the right-hand view, the pair of structures are rotated slightly to show the residues at the site of mutation: the threonine and alanine residues at position 575 are shown as sticks. Elements of secondary structure are labelled. For clarity, residues preceding Gln6 are not shown. B) Variations in local conformation between AR DBD-WT and AR DBD-T575A are shown by average pairwise r.m.s.d. values calculated over five-residue segments of the primary sequence between two sets of 20 calculated structures. An asterisk marks the site of mutation. C) Changes in average C-C distances upon binding greater than  $2\text{\AA}$  are shown by blue (increased distance) and red (decreased distance) squares. An asterisk marks the site of mutation.

Figure 4. Spectral density values,  $J(\omega)$ , at frequencies A) 0, B)  $\omega_\text{N}$  and C)  $\omega_\text{H} + \omega_\text{N}$ , plotted against sequence.  $J(\omega)$  values were determined from relaxation measurements for backbone  $^{15}\text{N}$  nuclei in AR DBD-WT (black) and AR DBD-T575A (magenta). Light pink shading indicates residues affected by ps-ns time-scale motions in both domains; dark pink shading indicates residues affected only in one domain. Light blue shading indicates residues affected by  $\mu\text{s}$ -ms time-scale conformational exchange in both domains; dark blue shading indicates residues affected only in AR DBD-T575A. An asterisk marks the site of mutation.

Figure 5. Long-range  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra correlating  $\text{H}\epsilon 1$  and  $\text{H}^\text{TM}2$  resonances with  $\text{N}^\text{TM}1$  and  $\text{N}\epsilon 2$  resonances of His570 for A) AR DBD-WT and B) AR DBD-T575A. C) Titration of His570 monitored through the  $\text{H}\epsilon 1$  chemical shift in AR DBD-WT (black) and AR DBD-T575A (magenta).

Figure 6. A)  $^1\text{H}$ - $^2\text{H}$  exchange rates for AR DBD-WT (blue bars) and AR DBD-T575A (red bars), plotted against sequence. B) Ratio of  $^1\text{H}$ - $^2\text{H}$  exchange rates between AR DBD-T575A and AR DBD-WT. The zinc-binding sites and elements of secondary structure are shown schematically between the two panels. An asterisk marks the site of mutation.

Figure 7. Luciferase reporter assay for AR-WT, AR-T575A and AR-H570A. The H570A mutation orientates androgen receptor transactivation towards the non-specific responsive promoter. LNCaP cells were cotransfected with AR DBD-WT or the indicated mutant expression plasmid together with PSA61-Luc (A) or MMTV-luc (B) firefly luciferase construct and pGL4.70[hRLuc] (Renilla) for

transfection control. Cells were then treated with 100nM DHT or vehicle in serum-free medium for 48 hours. Luminescence (relative light units) was normalized to the Renilla luciferase expression from the pGL4.70[hRLuc]. P values of less than 0.05 were considered significant (\*\*\*).

Figure 8. Sets of conserved and non-conserved residues, shown as spheres, mapped onto a representative structure of AR DBD-WT shown as a grey ribbon and grey lines. Hydrogen atoms have been omitted for clarity. The sets of residues represented are as follows: A) zinc-coordinating cysteine residues (yellow), B) residues conserved across all human nuclear receptor DBDs (green), C) residues conserved across the subset of nuclear receptor DBDs from AR, mineral corticoid receptor, glucocorticoid receptor and progesterone receptor (magenta), D) residues for which AR differs from mineral corticoid receptor, glucocorticoid receptor and progesterone receptor (red). See Figure 1A for sequences.

## References

1. Evans, R. M. (1988) *Science* **240**(4854), 889-95
2. Robinson-Rechavi, M., Escriva Garcia, H., and Laudet, V. (2003) *J Cell Sci* **116**(Pt 4), 585-6
3. Lee, H. and Chang, C. (2003) *Cell Mol Life Sci* **60**(8), 1613-22
4. Culig, Z. (2003) *Urology* **62**(5 Suppl 1), 21-6
5. Taplin, M. and Balk, S. P. (2004) *J Cell Biochem* **91**(3), 483-90
6. Bergerat, J. and Céraline, J. (2009) *Hum Mutat* **30**(2), 145-57
7. Brooke, G. N. and Bevan, C. L. (2009) *Curr Genomics* **10**(1), 18-25
8. Céraline, J., Cruchant, M. D., Erdmann, E., Erbs, P., Kurtz, J., Duclos, B., Jacqmin, D., Chopin, D., and Bergerat, J. (2004) *Int J Cancer* **108**(1), 152-7
9. Lapouge, G., Erdmann, E., Marcias, G., Jagla, M., Monge, A., Kessler, P., Serra, S., Lang, H., Jacqmin, D., Bergerat, J., and Céraline, J. (2007) *Int J Cancer* **121**(6), 1238-44
10. Lapouge, G., Marcias, G., Erdmann, E., Kessler, P., Cruchant, M., Serra, S., Bergerat, J., and Céraline, J. (2008) *Adv Exp Med Biol* **617**, 529-34
11. Dehm, S. M., Schmidt, L. J., Heemers, H. V., Vessella, R. L., and Tindall, D. J. (2008) *Cancer Res* **68**(13), 5469-77
12. Hu, R., Dunn, T. A., Wei, S., Isharwal, S., Veltri, R. W., Humphreys, E., Han, M., Partin, A. W., Vessella, R. L., Isaacs, W. B., Bova, G. S., and Luo, J. (2009) *Cancer Res* **69**(1), 16-22
13. Guo, Z., Yang, X., Sun, F., Jiang, R., Linn, D. E., Chen, H., Chen, H., Kong, X., Melamed, J., Tepper, C. G., Kung, H., Brodie, A. M. H., Edwards, J., and Qiu, Y. (2009) *Cancer Res* **69**(6), 2305-13
14. Marcias, G., Erdmann, E., Lapouge, G., Siebert, C., Barthélémy, P., Duclos, B., Bergerat, J., Céraline, J., and Kurtz, J. (2010) *Hum Mutat* **31**(1), 74-80
15. Gottlieb, B., Beitel, L. K., Wu, J. H., and Trifiro, M. (2004) *Hum Mutat* **23**(6), 527-33
16. Jagla, M., Fève, M., Kessler, P., Lapouge, G., Erdmann, E., Serra, S., Bergerat, J., and Céraline, J. (2007) *Endocrinology* **148**(9), 4334-43

17. Nazareth, L. V., Stenoiien, D. L., Bingman, 3rd, W. E., James, A. J., Wu, C., Zhang, Y., Edwards, D. P., Mancini, M., Marcelli, M., Lamb, D. J., and Weigel, N. L. (1999) *Mol Endocrinol* **13**(12), 2065-75
18. Monge, A., Jagla, M., Lapouge, G., Sasorith, S., Cruchant, M., Wurtz, J., Jacqmin, D., Bergerat, J., and Céraline, J. (2006) *Cell Mol Life Sci* **63**(4), 487-97
19. Lipari, G. and Szabo, A. (1982) *J Am Chem Soc* **104**(17), 4546–4559
20. Lipari, G. and Szabo, A. (1982) *J Am Chem Soc* **104**(17), 4559–4570
21. Lundbäck, T., van Den Berg, S., and Härd, T. (2000) *Biochemistry* **39**(30), 8909-16
22. Piotto, M., Saudek, V., and Sklenár, V. (1992) *J Biomol NMR* **2**(6), 661-5
23. States, D. J., Haberkorn, R. A., and Ruben, D. J. (1982) *J Magn Reson* **48**(2), 286–292
24. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J Biomol NMR* **6**(3), 277–293
25. Keller, R. (2004). *The Computer Aided Resonance Assignment Tutorial*. CANTINA Verlag.
26. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J Biomol NMR* **13**(3), 289–302
27. Hall, J. B., Dayie, K. T., and Fushman, D. (2003) *J Biomol NMR* **26**(2), 181–186
28. Herrmann, T., Guntert, P., and Wuthrich, K. (2002) *J Mol Biol* **319**(1), 209–227
29. Herrmann, T., Guntert, P., and Wuthrich, K. (2002) *J Biomol NMR* **24**(3), 171–189
30. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) *J Magn Reson* **160**(1), 65-73
31. Schwieters, C. D., Kuszewski, J. J., and Clore, G. M. (2006) *Prog NMR Spectr* **48**(1), 47–62
32. Linge, J. P., Williams, M. A., Spronk, C. A. E. M., Bonvin, A. M. J. J., and Nilges, M. (2003) *Proteins* **50**(3), 496-506
33. Nabuurs, S. B., Nederveen, A. J., Vranken, W., Doreleijers, J. F., Bonvin, A. M. J. J., Vuister, G. W., Vriend, G., and Spronk, C. A. E. M. (2004) *Proteins* **55**(3), 483-6
34. DeLano, W. (2002). The PyMol Molecular Graphics System. Retrieved from <http://www.pymol.org>.
35. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry* **33**(19), 5984-6003
36. Kay, L. E., Nicholson, L. K., Delaglio, F., Bax, A., and Torchia, D. A. (1992) *J Magn Reson* **97**(2), 359–375

## Appendix

### 1. D<sub>2</sub>O exchange experiments

Deuterium exchange of HN protons is a tool which estimates the accessibility of labile protons to solvent in protein. The model of protein hydrogen exchange as a two-step process was proposed by Hvidt and Nielsen.[1] The rate of exchange is affected by the level of protection of HN protons from water. The exchange is slow when the HN proton is buried in the core of protein and/or is involved in a hydrogen bond. By changing the solvent from H<sub>2</sub>O to D<sub>2</sub>O, amide hydrogen exchange can be observed by the disappearance of signals in <sup>1</sup>H-<sup>15</sup>N HSQC spectra. The rate of decay of the

intensity of peaks observed in the HSQC spectrum can be used as an indicator to identify the secondary structure in proteins.

## 2. Methods

The water was removed from a sample of buffer, by freezing and using the spin vacuum apparatus. The same volume of 100% D<sub>2</sub>O was added to the buffer powder and agitated. The protein in its original buffer was diluted 5-fold into the new buffer. A series of <sup>1</sup>H-<sup>15</sup>N HSQC spectra was then recorded over time. The intensities of observed peaks in the <sup>1</sup>H-<sup>15</sup>N HSQC spectra were measured and plotted against time. Residues involved in hydrogen bonds can be identified from the exponential decay of intensities in these plots.

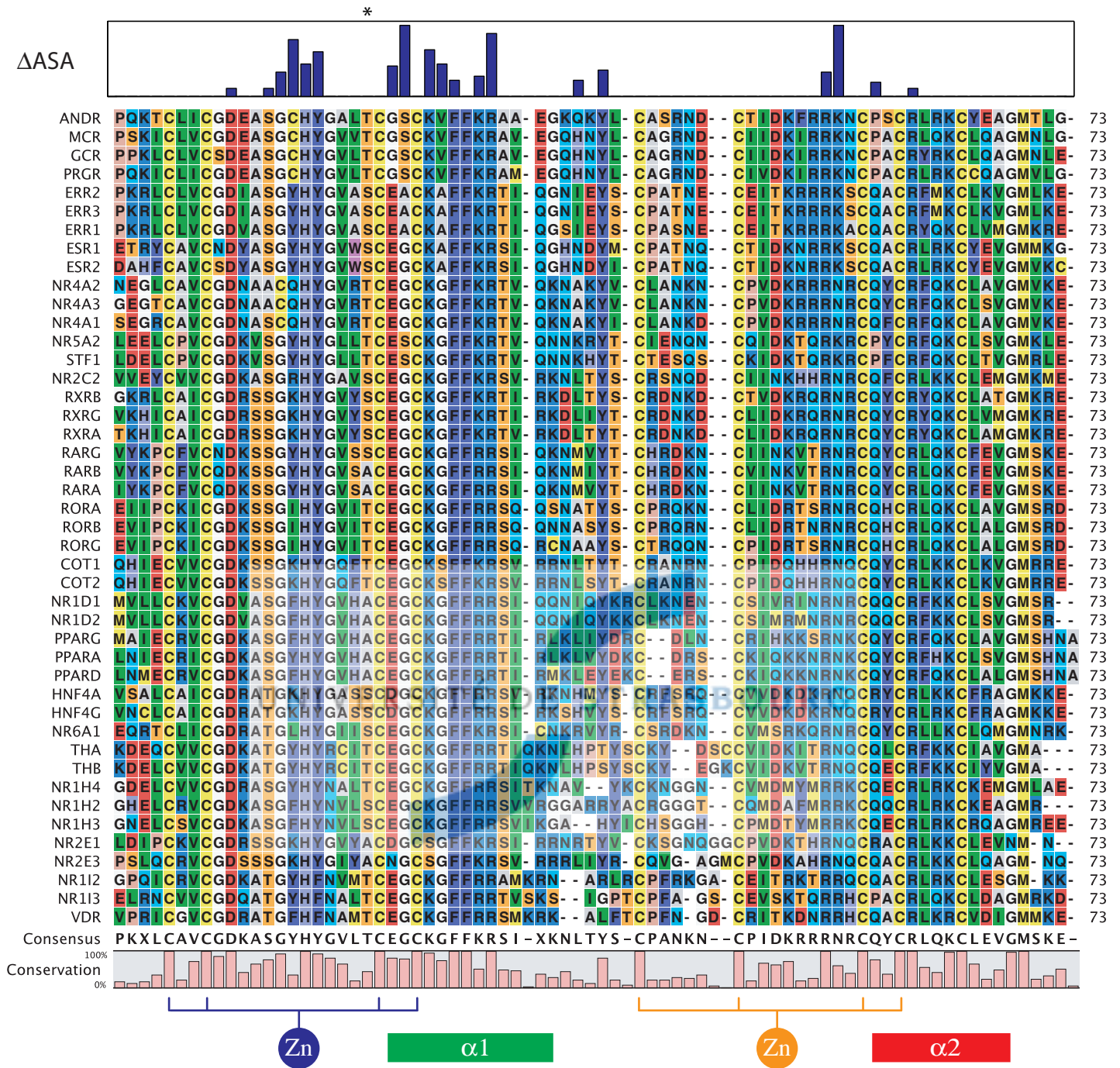
## Reference

1. Hvidt A, Nielsen SO: **Hydrogen exchange in proteins**. *Adv Protein Chem* 1966, **21**:287-386



# Figure 1

A)

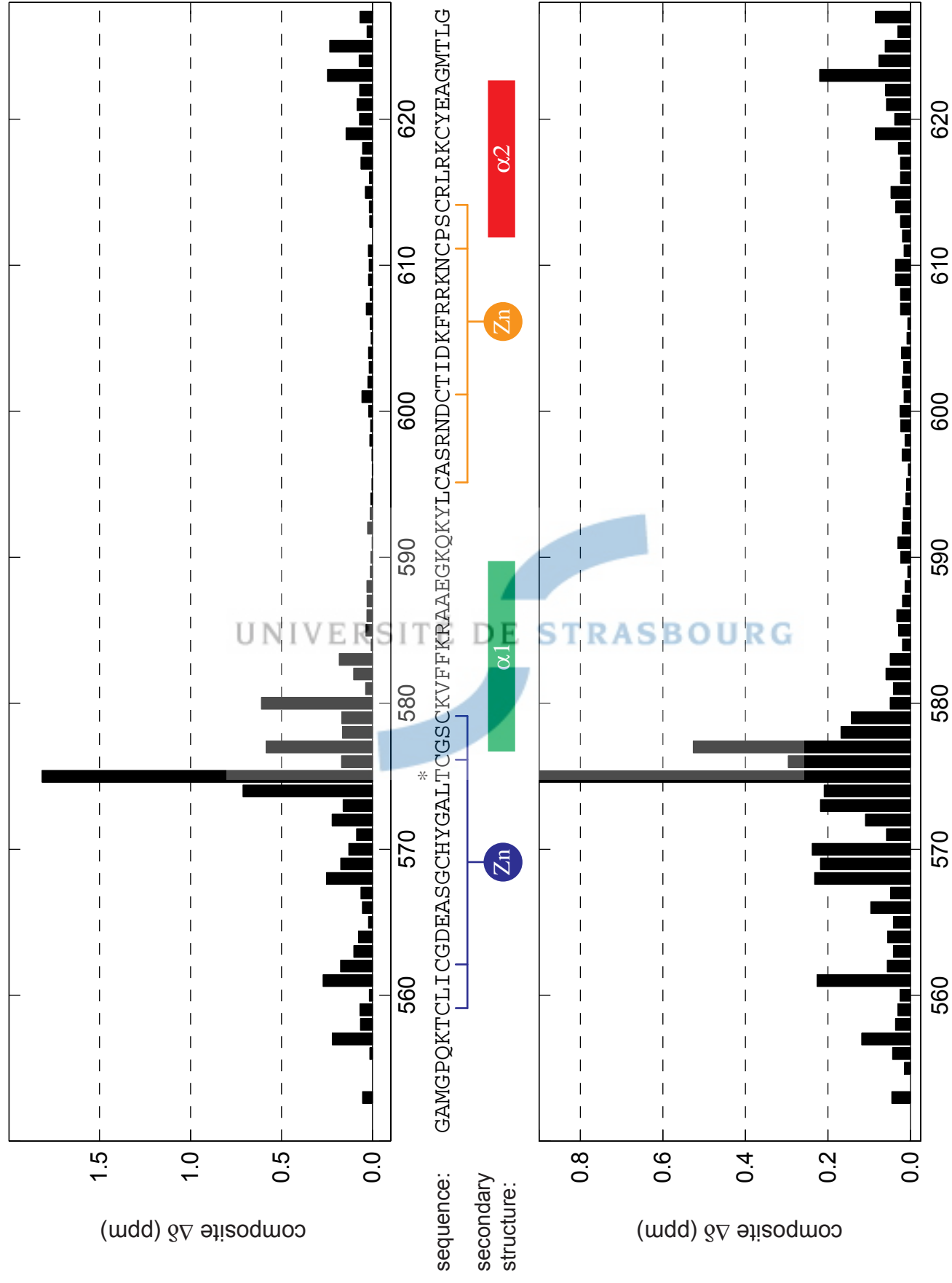


The logo of the University of Strasbourg, featuring two blue curved segments that form a stylized 'S' shape, positioned behind the text.

UNIVERSITÉ DE STRASBOURG

Figure 1

B)





The logo of the University of Strasbourg, featuring two blue curved segments that form a stylized 'S' shape, positioned behind the text.

UNIVERSITÉ DE STRASBOURG

**Table 1 NMR and refinement statistics for protein structures**

<b>ARDBD</b>	<b>Wild type</b>	<b>T575A mutant</b>
<b>NMR distance and dihedral constraints</b>		
Distance constraints		
Total NOE	993	935
Intra-residue	246	270
Inter-residue		
Sequential ( $ i - j  = 1$ )	274	254
Medium-range ( $ i - j  < 4$ )	156	125
Long-range ( $ i - j  > 5$ )	317	286
Protein-zinc restraints		
Hydrogen bonds	21	21
Total dihedral angle restraints		
$\phi$	36	39
$\psi$	36	39
<b>Structure statistics</b>		
Violations (mean and s.d.)		
Distance constraints (Å)	0.0366 ± 0.0017	0.0298 ± 0.0020
Dihedral angle constraints (°)	0.4882 ± 0.0919	0.5453 ± 0.1148
Max. dihedral angle violation (°)	4.195	4.339
Max. distance constraint violation (Å)	0.487	0.436
Deviations from idealized geometry		
Bond lengths (Å)	0.0146 ± 0.0006	0.0129 ± 0.0004
Bond angles (°)	1.8993 ± 0.0697	1.5316 ± 0.0563
Impropers (°)	1.8733 ± 0.0972	1.5082 ± 0.0756
Average pairwise r.m.s. deviation* (Å)		
Heavy	1.59 ± 0.25	1.62 ± 0.20
Backbone	0.28 ± 0.19	0.27 ± 0.17

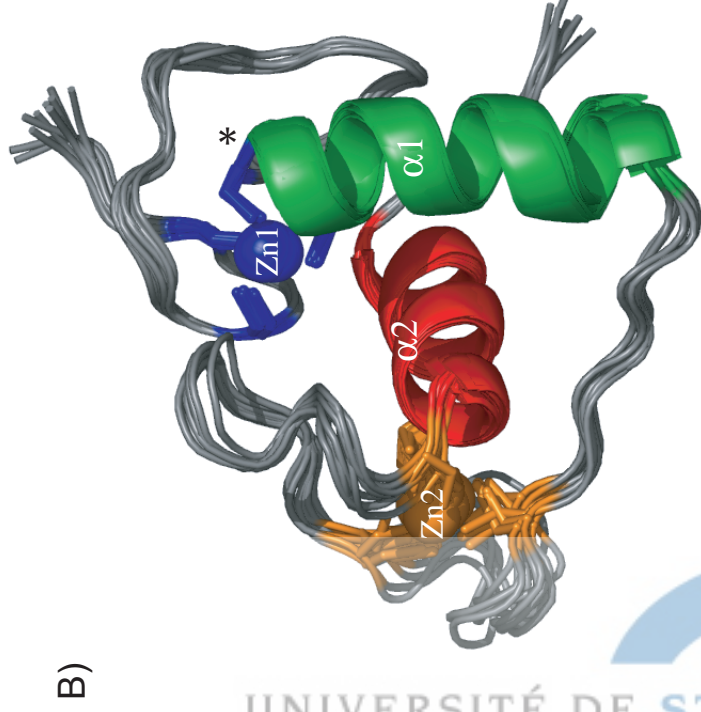
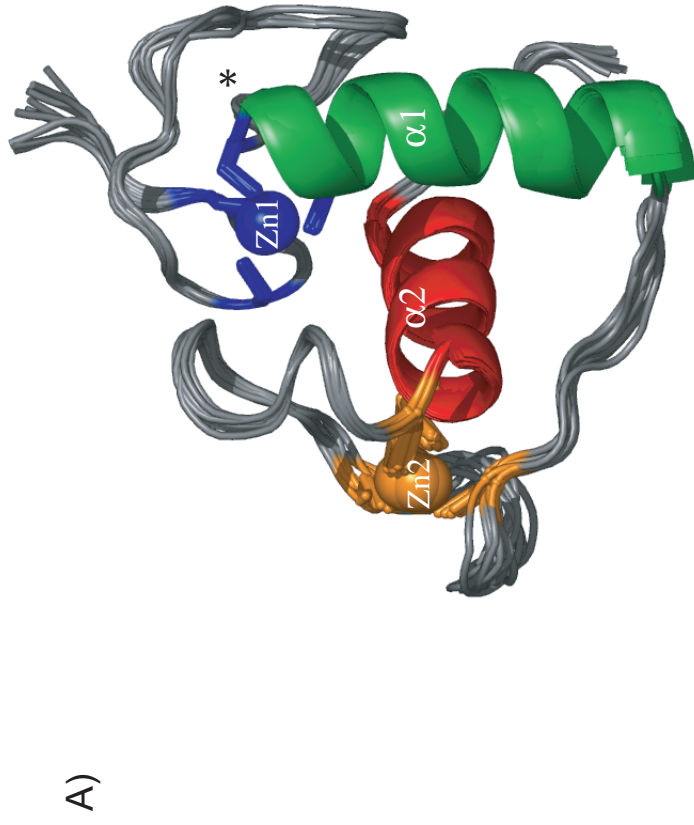
\*Pairwise r.m.s. deviation was calculated among 20 refined structures using the C<sub>α</sub> positions of residues 557-623 for ARDBD wild type and T575A mutant.

<b>Ramachandran plot (20 lowest-energy structures)</b>	<b>WT</b>	<b>T575A</b>
Residues in most favorable regions (%)	74.8 %	79.5 %
Residues in additional favorable regions (%)	19.6 %	17.3 %
Residues in generously favorable regions (%)	3.8 %	2.4 %
Residues in disallowed regions (%)	1.7 %	0.8 %

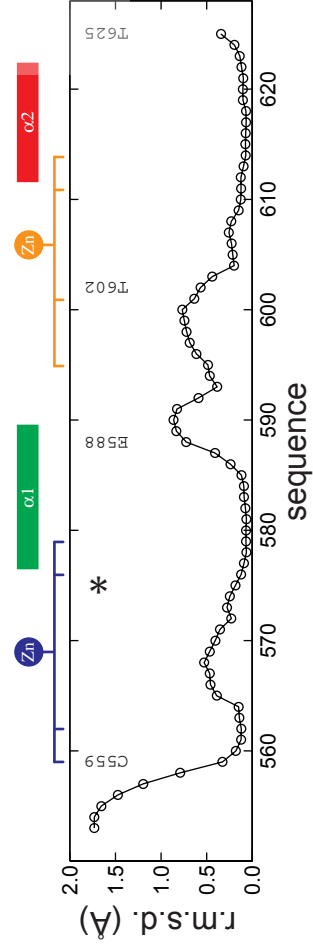
The logo of the University of Strasbourg, featuring two blue curved segments that form a stylized 'S' shape, positioned behind the text.

UNIVERSITÉ DE STRASBOURG

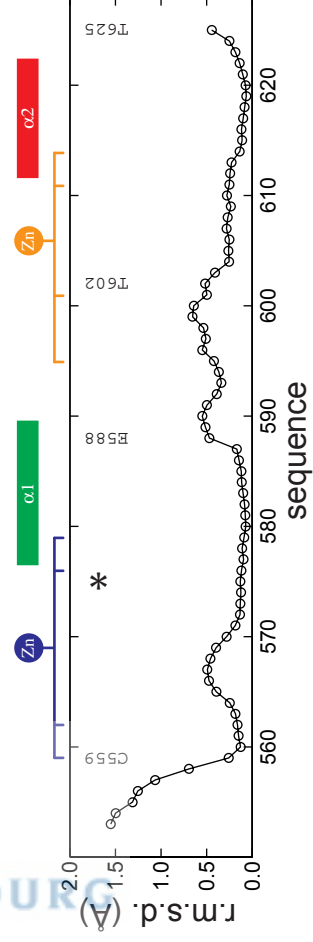
Figure 2



C)



D)

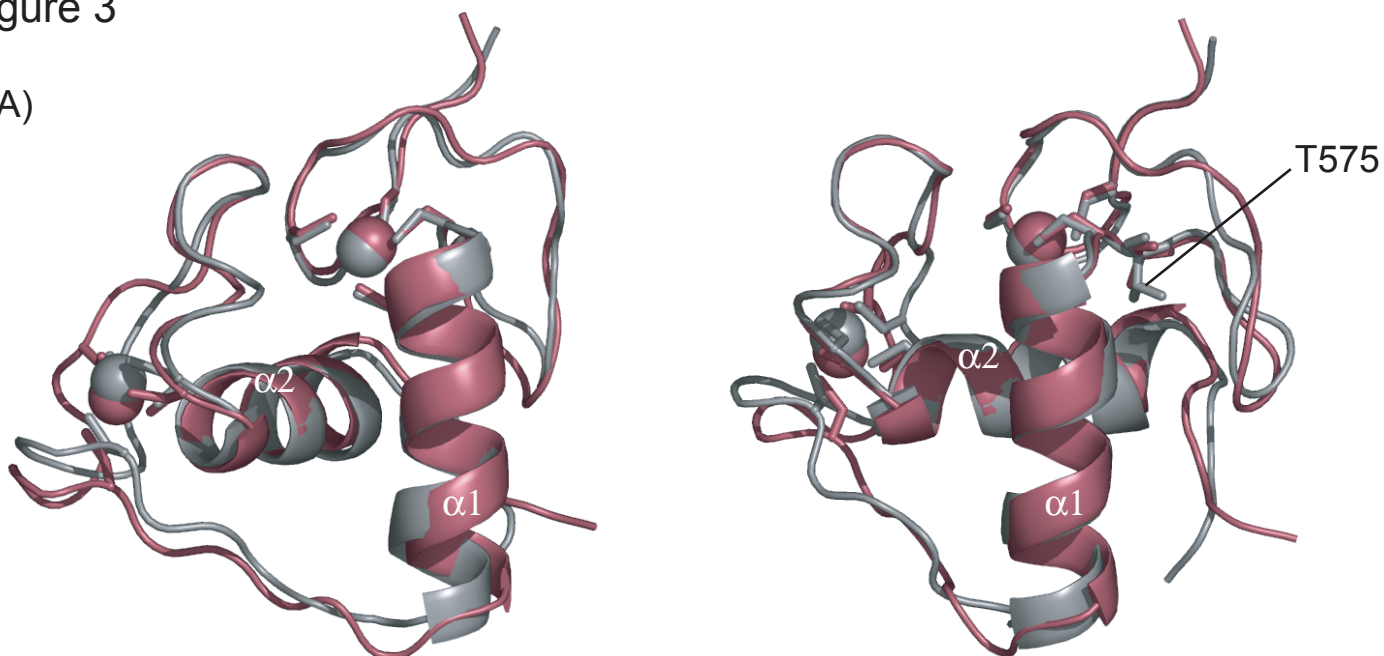


The logo of the University of Strasbourg, featuring two blue curved segments that form a stylized 'S' shape, positioned behind the text.

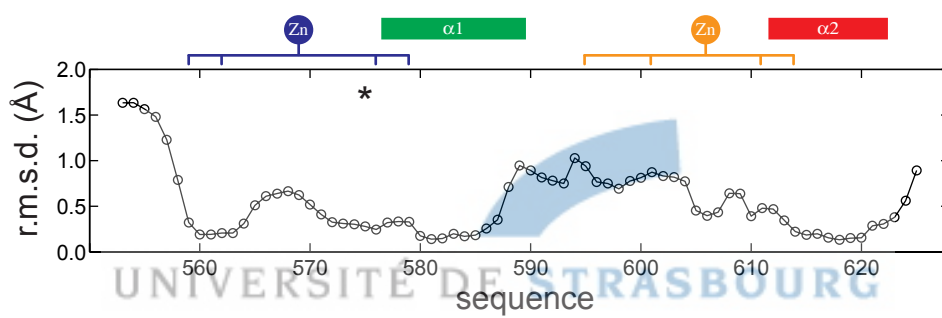
UNIVERSITÉ DE STRASBOURG

Figure 3

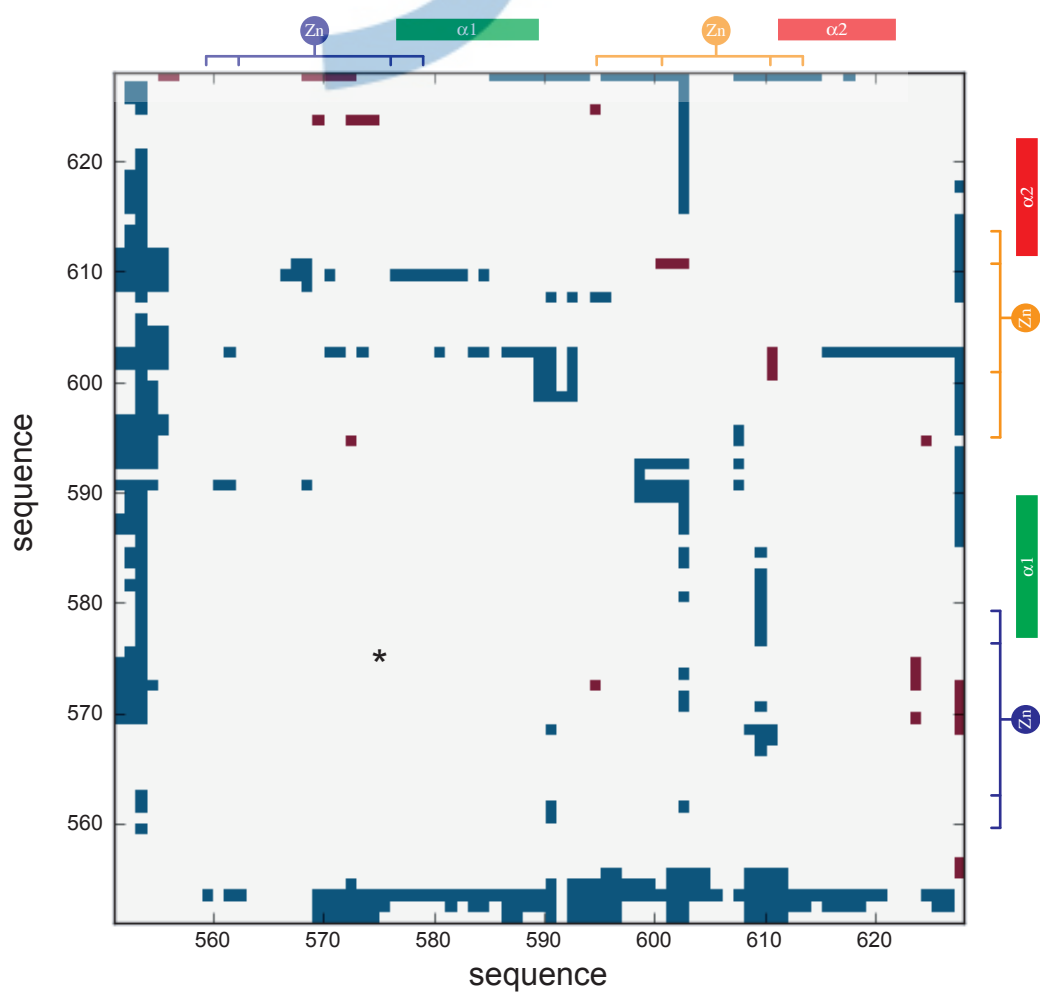
A)



B)



C)

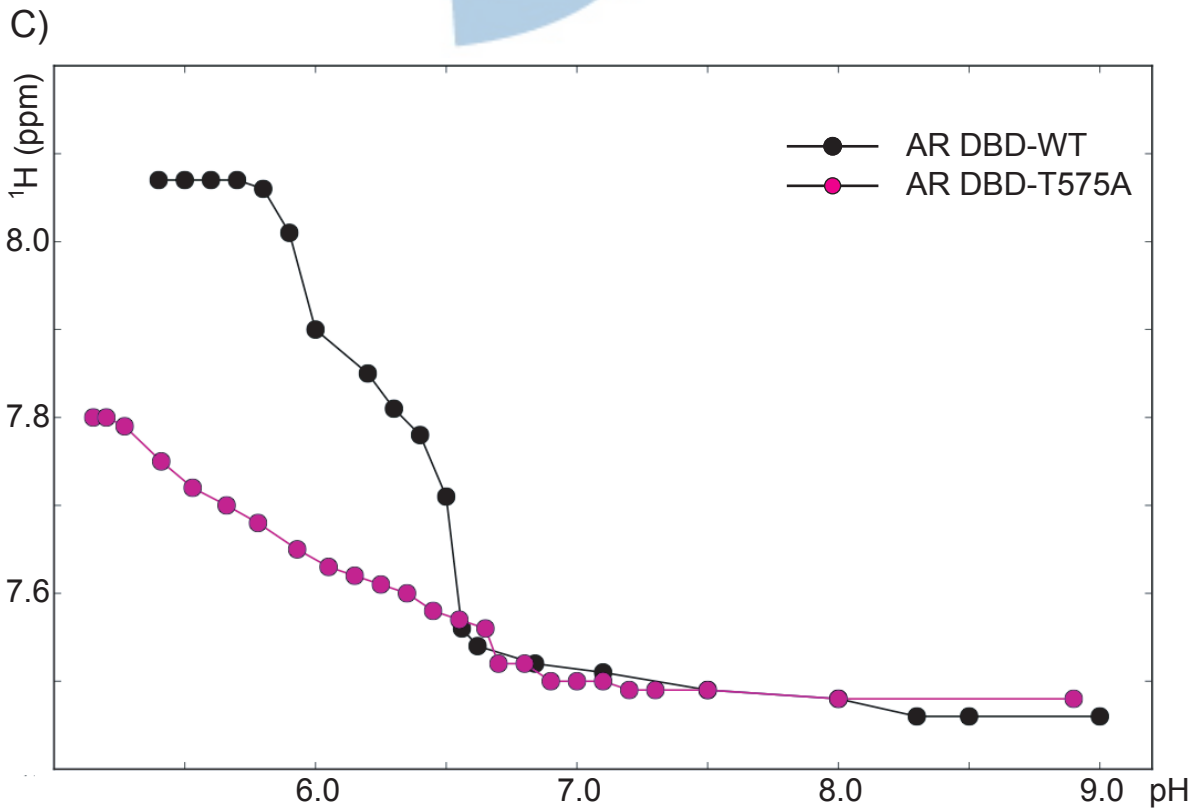
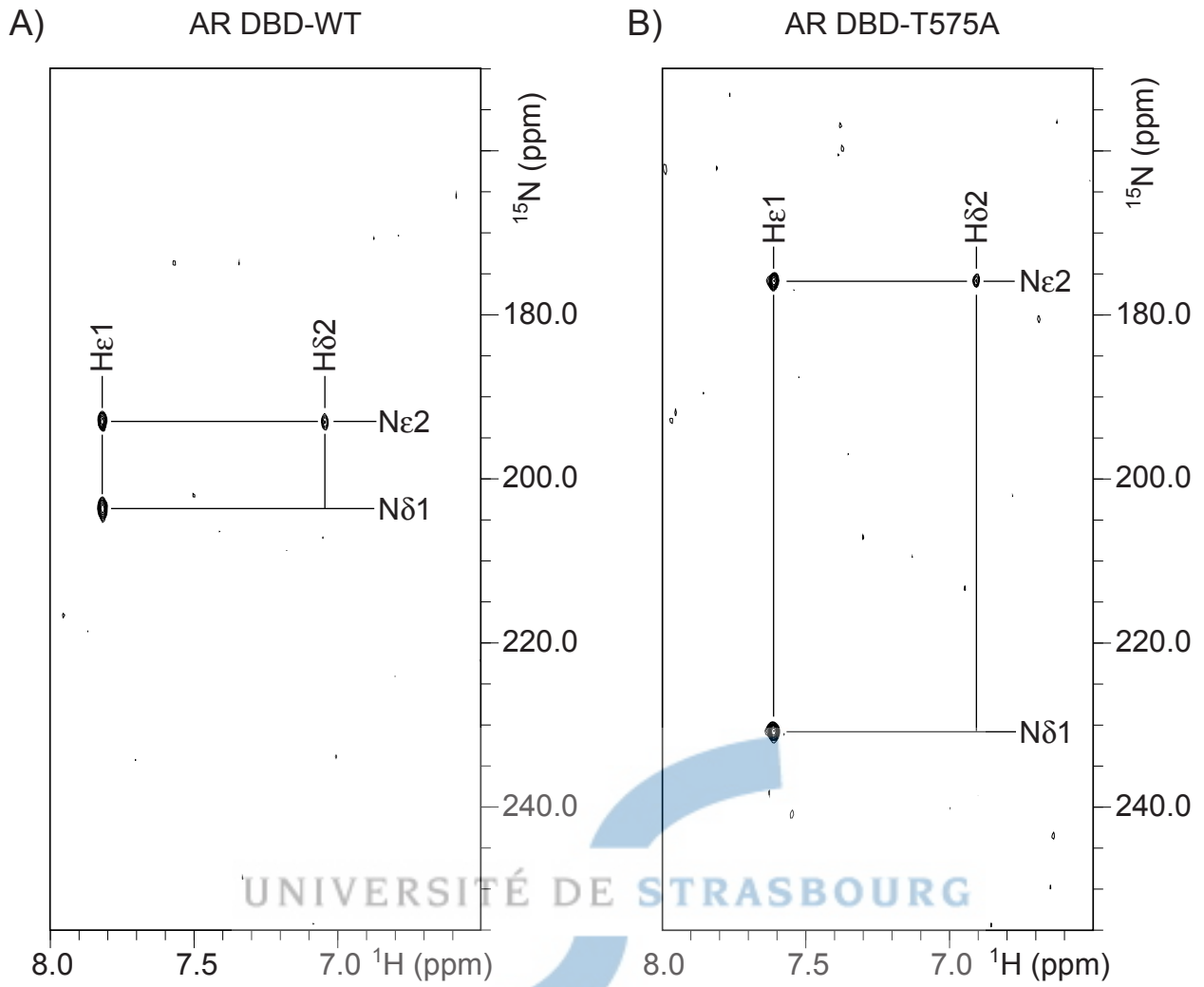


The logo of the University of Strasbourg, featuring two blue curved segments that form a stylized 'S' shape, positioned behind the text.

UNIVERSITÉ DE STRASBOURG



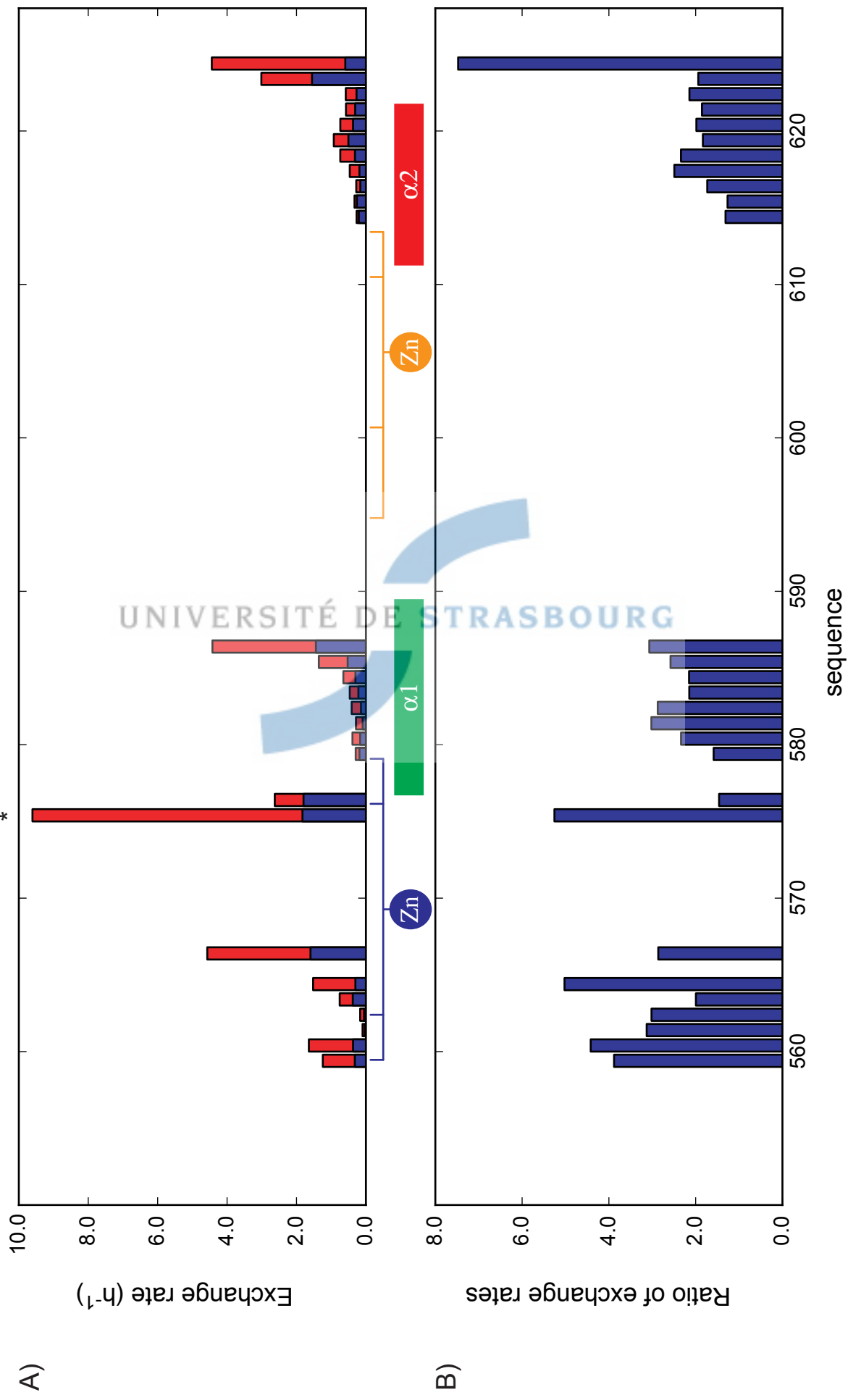
Figure 5



The logo of the University of Strasbourg, featuring two blue curved segments that form a stylized 'S' shape, positioned behind the text.

UNIVERSITÉ DE STRASBOURG

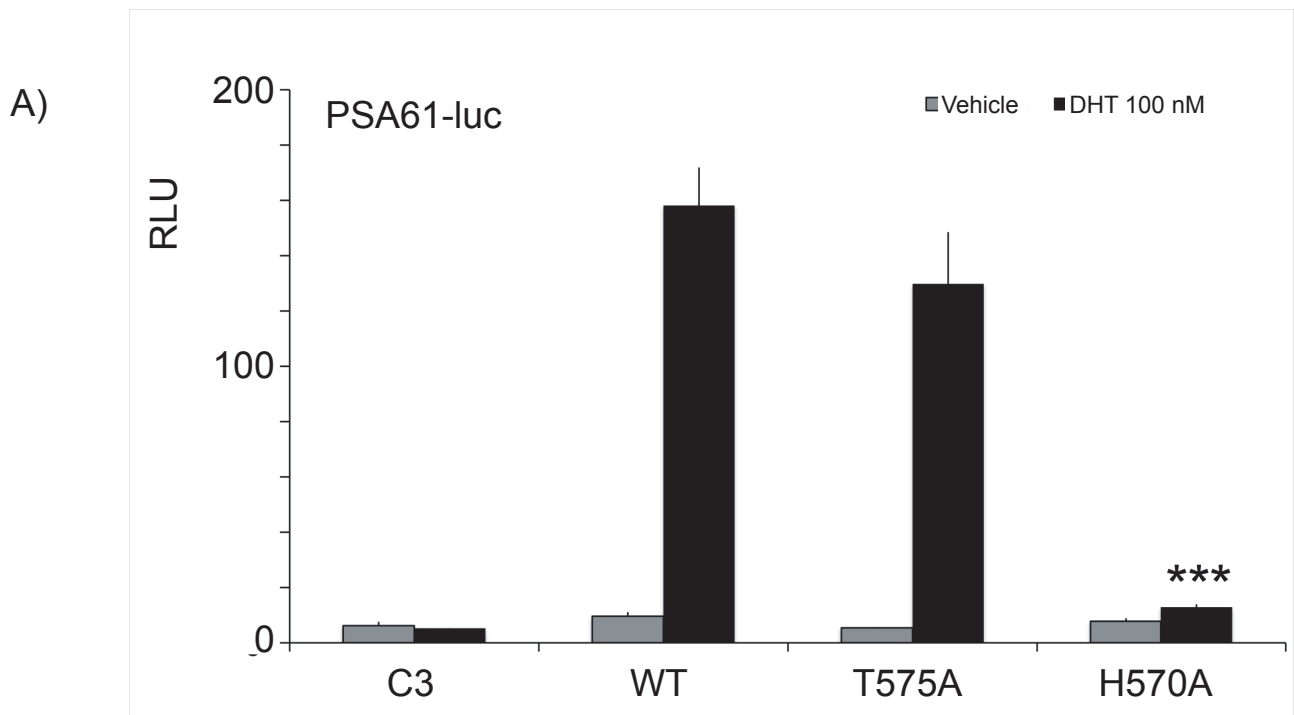
Figure 6



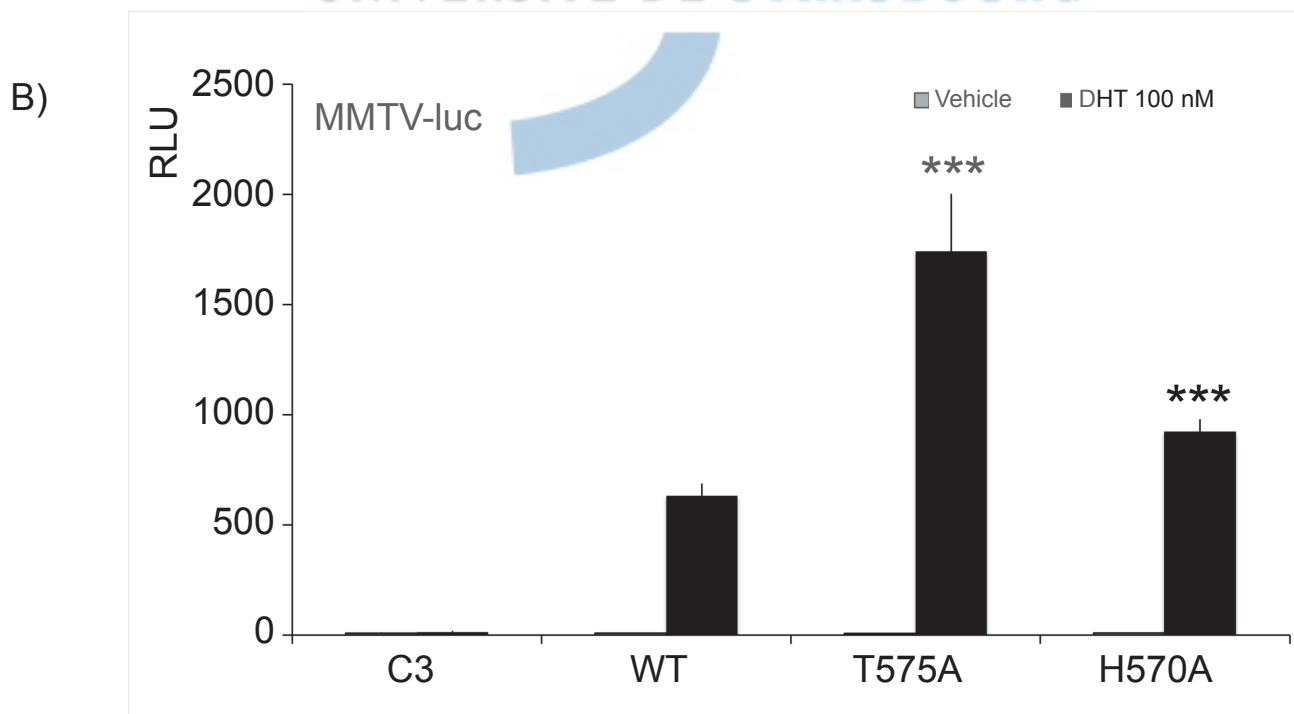
The logo of the University of Strasbourg, featuring two blue curved segments that form a stylized 'S' shape, positioned behind the text.

UNIVERSITÉ DE STRASBOURG

Figure 7



UNIVERSITÉ DE STRASBOURG

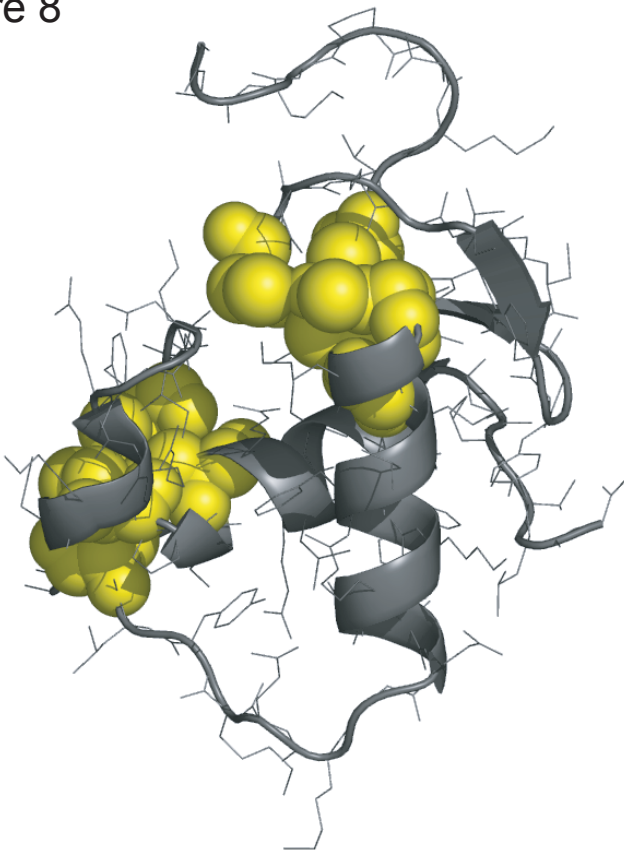


The logo of the University of Strasbourg, featuring two blue curved segments that form a stylized 'S' shape, positioned behind the text.

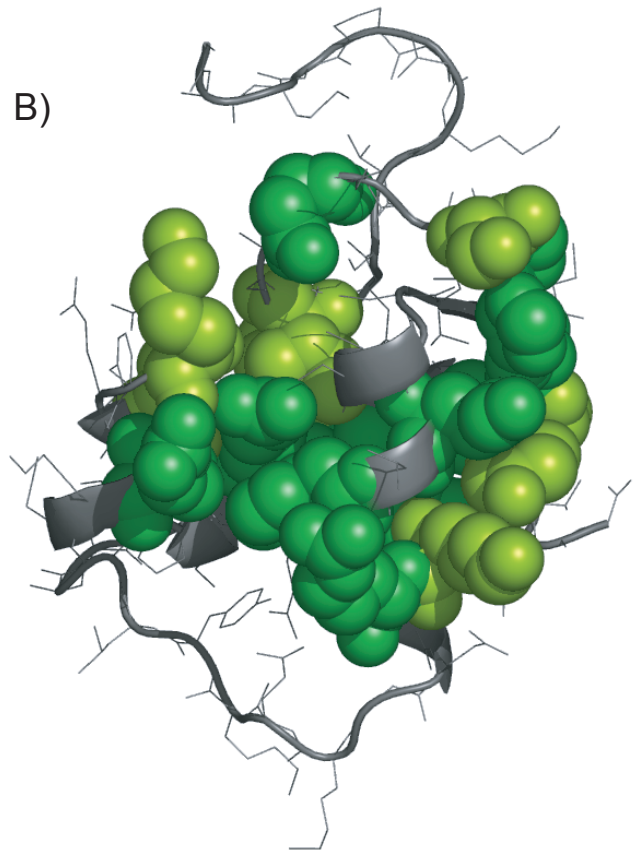
UNIVERSITÉ DE STRASBOURG

Figure 8

A)

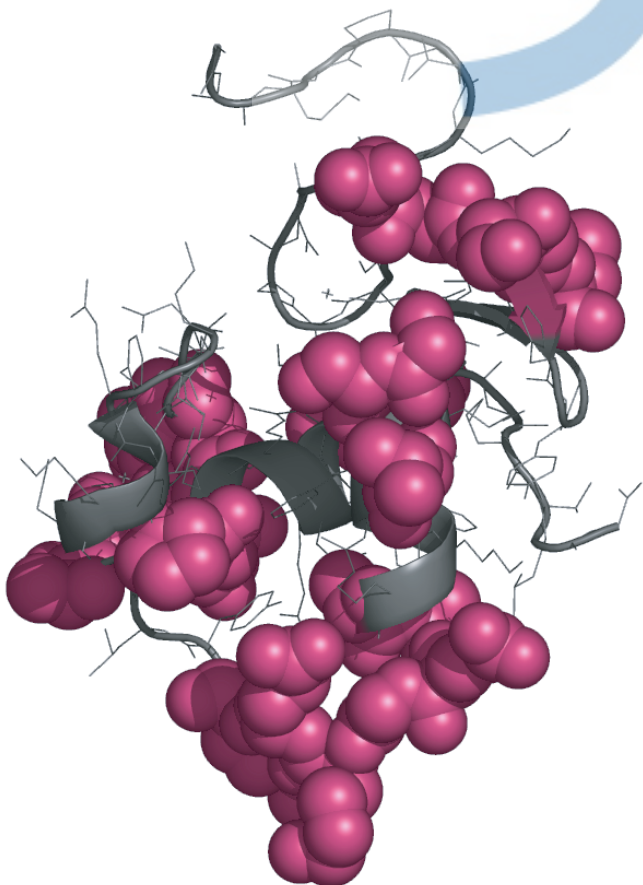


B)

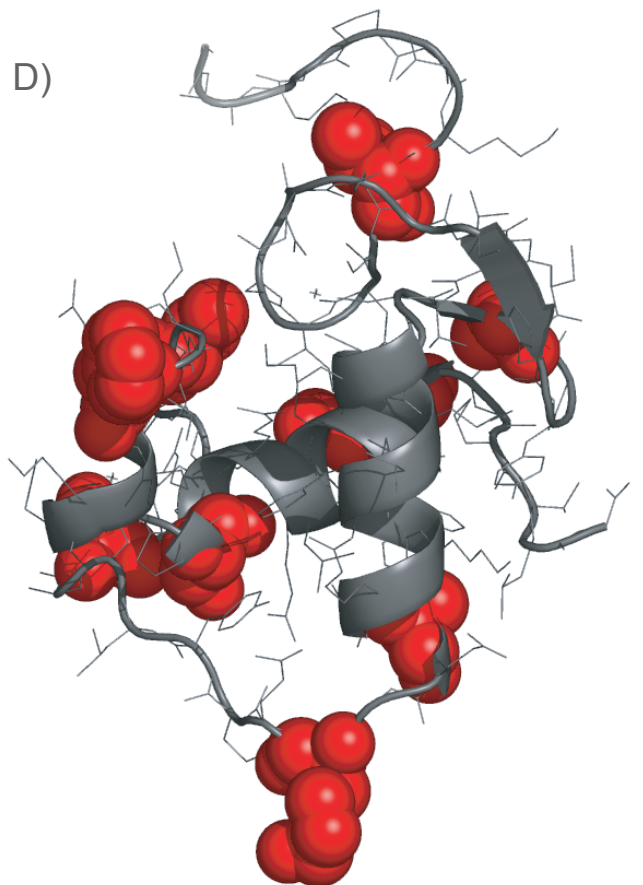


UNIVERSITÉ DE STRASBOURG

C)



D)





The logo of the University of Strasbourg, featuring two blue curved segments that form a stylized 'S' shape, positioned behind the text.

UNIVERSITÉ DE STRASBOURG

# Histone deubiquitination by SAGA is modulated by an atypical zinc finger domain of Ataxin-7

Jacques Bonnet<sup>1,3</sup>, Ying-Hui Wang<sup>2,3</sup>, R. Andrew Atkinson<sup>2</sup>, Jennifer Koffler<sup>1</sup>, Christophe Romier<sup>2</sup>,  
Ali Hamiche<sup>1</sup>, László Tora<sup>1</sup>, Didier Devys<sup>1,\*</sup> and Bruno Kieffer<sup>2,\*</sup>

UNIVERSITÉ DE STRASBOURG

<sup>1</sup>Department of Functional Genomics, <sup>2</sup>Department of Structural Biology and Genomics, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR 7104, INSERM U 964, Université de Strasbourg, BP 10142 - 67404 ILLKIRCH Cedex, France

<sup>3</sup>These authors contributed equally to this work

\*Corresponding authors: Kieffer, B. [kieffer@igbmc.u-strasbg.fr](mailto:kieffer@igbmc.u-strasbg.fr)

Devys, D. [devys@igbmc.u-strasbg.fr](mailto:devys@igbmc.u-strasbg.fr)

## Abstract

Co-activator complexes play key roles in the regulation of transcription by RNA polymerase II through modification of chromatin compaction. One such complex, SAGA (Spt-Ada-Gcn5 Acetyltransferase), harbors both a histone acetylation activity and a deubiquitination module. Two subunits of this module contain atypical zinc-finger domains, termed SCA7 domains that are found exclusively in proteins encoded by the ataxin-7 gene (*ATXN7*) family. We show that the SCA7 domain of hATXN7 and its yeast ortholog ySgf73 bind nucleosomes via histone H2A-H2B dimers, while that of hATXN7L3 does not. The solution structures of the SCA7 domains of hATXN7 and hATXN7L3 reveal a common, novel zinc-finger scaffold at the heart of two distinct folds, thus providing a molecular basis for the observed functional differences. Our findings define an additional mode of regulation of the deubiquitination activity of SAGA through the direct recognition of nucleosomes by the SCA7 domain of its hATXN7/ySgf73 subunit.



## Introduction

Transcription initiation is regulated by the diverse functions of multi-protein co-activator complexes that include histone modification and nucleosome remodeling activities, as well as interactions with gene-specific activators and general transcription factors <sup>1,2</sup>. SAGA (Spt–Ada–Gcn5 acetyltransferase) is one such co-activator complex that is highly conserved throughout evolution, with the histone acetyltransferase (HAT) GCN5 as catalytic subunit <sup>3,4</sup>. The structural organization of the SAGA complex revealed that its multiple activities are localized in distinct functional modules <sup>5</sup>. SAGA recruitment at specific loci is mediated by the hTRRAP/yTra1 subunit that interacts with DNA-bound activators. In addition, several other SAGA subunits contain well-characterized or putative chromatin-binding domains, such as bromodomains in GCN5 and in ySpt7 <sup>6,7</sup>, a Tudor domain in Sgf29 as well as a SANT and a SWIRM domain in ADA2b <sup>8</sup>. Interestingly, yeast Gcn5, Ada2 and Ada3 form a heterotrimer and both human ADA2b and ADA3 were shown to regulate the catalytic activity of hGCN5 <sup>9,10</sup>. It has thus been suggested that GCN5 activity is controlled by the chromatin-binding properties of SAGA subunits interacting with GCN5 and constituting a functional HAT module.

SAGA hosts a second enzymatic activity as it was shown to remove ubiquitin from monoubiquitinated histone H2B <sup>11,12</sup>. This activity is mediated by the ubiquitin protease hUSP22/yUbp8 that belongs to a distinct functional module of SAGA. This deubiquitination (DUB) module is formed by hUSP22/yUbp8, hATXN7L3/ySgf11, hENY2/ySus1 and hATXN7/ySgf73 (**Fig. 1a**) <sup>13-17</sup>. There is evidence that the catalytic activity of hUSP22/yUbp8 is regulated by interactions with the other subunits of the DUB module <sup>18,19</sup>. Furthermore, ySgf73 was shown to mediate the incorporation of the DUB module into SAGA <sup>19,20</sup>. hATXN7/ySgf73 contains two conserved zinc finger (ZnF) domains (**Fig. 1b**). The first N-terminal domain (ZnF-Sgf73) is a canonical ZnF. In yeast, the deletion of this domain leads to the dissociation of Ubp8, Sgf11 and Sus1 from SAGA <sup>19</sup>, indicating that it interacts with other subunits of the DUB module. The second ZnF, referred to hereafter as the SCA7 domain (InterPro: IPR013243), is highly conserved and found exclusively in ATXN7 homologues (**Fig. 1b**). The SCA7 domain of hATXN7 was shown to bind zinc but displays an unusual organization with a Cys-X<sub>9-10</sub>-Cys-X<sub>5</sub>-Cys-X<sub>2</sub>-His motif characterized by a long sequence insertion between the first two zinc-coordinating residues (**Fig. 1c**) <sup>21</sup>. Conservation of the SCA7 domain led to the identification of the ATXN7 gene family, consisting of four genes (*ATXN7*, *ATXN7L1*, *ATXN7L2* and *ATXN7L3*) in vertebrates and a single gene (*Sgf73*) in *S. cerevisiae* <sup>21</sup>. We have shown that hATXN7 and hATXN7L3 are two distinct components of the SAGA complex and are the orthologs of ySgf73 and ySgf11, respectively <sup>17</sup>. Indeed, the N-terminal ZnF domain (ZnF-Sgf11) of hATXN7L3 differs in sequence from that of hATXN7 (ZnF-Sgf73) and defines hATXN7L3 as the ySgf11 ortholog (**Fig. 1b**). Surprisingly, hATXN7L3 differs from ySgf11 in that the latter lacks the SCA7 domain found in the former (**Fig. 1b**).

Previously underestimated, the large range of interactions mediated by ZnF domains has been recognized only recently, due to the increasing availability of data on their structures and mechanisms

of interaction<sup>22</sup>. ZnF domains are most abundant in higher eukaryotes where they provide stable scaffolds to achieve specialized function. Here, we report a structure-function analysis of SCA7 ZnF domains, defined by a motif that is found solely in subunits of the SAGA complex. We identified the SCA7 domain of hATXN7/ySgf73, but not that of hATXN7L3, as a nucleosome-binding domain by direct interaction with the histone H2A-H2B dimer. We found that the SCA7 domain contains a novel zinc-binding fold that is reminiscent of the zinc-ribbon family. Furthermore, the SCA7 domains of hATXN7 and hATXN7L3 have distinct organizations of secondary structure elements, which is reflected in their different histone-binding properties.

## Results

### ATXN7 and ATXN7L3 contain distinct SCA7 domains

The SCA7 domain is found exclusively in members of the *ATXN7* gene family, which includes two distinct subunits of SAGA complexes: ATXN7 and ATXN7L3 orthologs (see Introduction and **Fig. 1b**). In order to examine the functional divergence of the SCA7 domains between ATXN7 and ATXN7L3 proteins, we performed multiple alignments of the corresponding sequences from different genomes. A set of invariant residues defines a consensus signature for this domain (**Fig. 1c**). These residues are located largely in the N-terminal part of the domain encompassing the putative zinc-coordinating residues. However, the sequences of SCA7 domains of ATXN7L3 orthologs show marked differences from those of ATXN7 orthologs. Indeed, we could identify sets of discriminating residues that were strictly conserved in hATXN7/ySgf73 orthologs but were not present in ATXN7L3 orthologs and vice-versa (**Fig. 1c**). The density of variant residues was enriched in the C-terminal part of the domain suggesting that these two sets of SCA7 domain might adopt a different structural organization.

The presence of a highly conserved SCA7 domain in all hATXN7 /ySgf73 orthologs suggests that it has an important function in SAGA structure and/or activity. During evolution, a SCA7 domain has been gained in ATXN7L3 in higher eukaryotes to fulfil a new yet uncharacterized, specialized function.

### SCA7 domains contain a novel type of zinc finger

The different patterns of sequence similarities of SCA7 domains from ATXN7 and ATXN7L3 suggested that these two domains might display distinct structural features. Based on sequence conservation analysis of the two SCA7 domains, <sup>13</sup>C,<sup>15</sup>N-labelled proteins encompassing residues Lys330 to His401 of ATXN7 and Gly197 to Gly277 of ATXN7L3, hereafter referred as ATXN7-SCA7 and ATXN7L3-SCA7 respectively, were produced and analysed using nuclear magnetic resonance (NMR) spectroscopy. Analysis of C<sub>α</sub> chemical shifts for the two proteins indicated major differences in the positions of folded and disordered regions with respect to the zinc-coordinating residues (**Fig. 2a,b**). Using the deviation from random chemical shifts values as an indicator of the folding state, it was clear that both ATXN7-SCA7 and ATXN7L3-SCA7 contained

unstructured regions encompassing the first and last ten residues of ATXN7-SCA7, and the last thirty-five residues of ATXN7L3-SCA7. These differences between the two domains were reflected in the relaxation properties of the amide  $^1\text{H}$  nuclei measured using a Het-SOFAST experiment (**Supplementary Fig. 1**).

Structure determination of the two proteins was performed using NOESY spectra recorded at 950 MHz, allowing efficient automated analysis by ATNOS/CANDID software (**Table 1**). Both sets of calculations resulted in a well-defined fold as shown by the superimposition of the lowest energy models of ATXN7-SCA7 and ATXN7L3-SCA7 (**Fig. 2c, d**). The larger number of NOEs assigned for ATXN7-SCA7 reflects the greater number of folded residues forming an extended core, as compared to ATXN7L3-SCA7.

SCA7 zinc fingers are defined by a conserved, atypical CCCH sequence motif with a large sequence insertion between the first and second cysteine residues that coordinate the zinc ion. In both ATXN7-SCA7 and ATXN7L3-SCA7, this extension corresponds to a protruding  $\beta$ -hairpin structure (**Fig. 3a, b**). The core of the zinc-binding site displays a conserved structure formed by two short adjacent loops located at the bottom of the  $\beta$ -hairpin. This particular three-dimensional organization of zinc-coordinating residues has not been observed previously and represents a novel fold of a zinc finger domain, which seems to be specific to SCA7 domains. The conserved histidine side-chain binds the zinc atom using its  $\text{N}_\delta$  atom as indicated by the protonation state of the imidazole ring, measured using a long-range  $^1\text{H}$ - $^{15}\text{N}$  HSQC. The zinc-binding scaffold contains two conserved hydrophobic residues that hook onto the remaining parts of the domain. In particular, the valine residue of the SCA7-specific Cys-Gly-Val motif (Val349 in ATXN7, Val211 in ATXN7L3) interacts with the  $\alpha$ -helix located immediately downstream from the zinc-binding site and contributes to its orientation (**Fig. 3a**). A second hook is provided by the Leu362 located at the tip of the first loop of the zinc coordination cage that anchors the second  $\alpha$ -helix in the ATXN7-SCA7 domain (**Fig. 3a**). Slight structural differences between the two zinc fingers are observed in the loop located at the extremity of  $\beta$ -hairpins. These differences reflect sequence variations found in the tip of the hairpin where the solvent-exposed Leu352 in ATXN7 is replaced by a histidine residue in ATXN7L3 (**Fig. 3a, b**).

### SCA7 domains of ATXN7 and ATXN7L3 differ in their helical structures

While the SCA7 domains of both ATXN7 and ATXN7L3 contain two  $\alpha$ -helices, these are not located at similar positions within the sequences. In ATXN7-SCA7, the two  $\alpha$ -helices (encompassing residues Leu369-Ala374 and Phe382-Thr393) are located downstream of the zinc-binding site and are separated by a loop containing a large number of positively-charged residues (**Fig. 1c**). In ATXN7L3-SCA7, the two  $\alpha$ -helices (encompassing residues Pro198-Thr207 and Asp231-Leu242) lie either side of the zinc-binding site leading to a different packing of the two helices (**Fig. 3c, d**).

In ATXN7-SCA7, the two helices have an almost perpendicular orientation, the  $\alpha 2$  helix being anchored to the zinc-binding site by the hydrophobic contact between Leu362 and Leu385 (**Fig. 3a**). This perpendicular orientation was confirmed by analysis of  $^{15}\text{N}$  heteronuclear relaxation data using an anisotropic tumbling model (data not shown). The closed conformation of the elbow formed by the

two helices is due to the tight packing of several conserved residues, namely Phe382 and Leu386 in helix  $\alpha 2$  and Arg372 in helix  $\alpha 1$  (**Fig. 3c**). In addition, a salt bridge is observed between Arg373 in helix  $\alpha 1$  and Asp383 in helix  $\alpha 2$ : this pair of residues is conserved in all ATXN7 sequences. Strikingly, the side-chain of Arg378 is buried inside the elbow and may probably contribute to the closed conformation via a cation- $\pi$  interaction with the side-chain of Phe382.

In ATXN7L3-SCA7, the helices  $\alpha 1$  and  $\alpha 2$  adopt an anti-parallel orientation defined by hydrophobic interactions involving residues Leu201 and Leu205 in helix  $\alpha 1$  and Val237, Phe241 and Leu242 in helix  $\alpha 2$  (**Fig. 3d**). Interestingly, sequence analysis of the *A. mellifera* ATXN7L3 ortholog suggests that the hydrophobic interaction between Leu201 and Leu242 is replaced by an electrostatic interaction between a lysine and a glutamate residue (**Fig. 1c**).

Analysis of the solution structures of ATXN7-SCA7 and ATXN7L3-SCA7 shows that sequence divergence between the two SCA7 domains translates into major structural rearrangement of helical structures around a conserved zinc-coordination scaffold.

### The SCA7 domain of hATXN7 and its yeast ortholog bind nucleosomes

We then asked whether these structural differences account for distinct functions of the SCA7 domains of ATXN7 and ATXN7L3. The deubiquitination activity of the hUSP22/yUbp8 subunit of SAGA is modulated by other subunits of the deubiquitination module and in particular by hATXN7/ySgf73<sup>19,20</sup>. By analogy with the HAT module that contains several chromatin-binding domains, we hypothesized that hATXN7/ySgf73 might recruit the deubiquitination module of SAGA to nucleosomes. In GST pull-down assays, nucleosomes were specifically retained by GST-ATXN7-SCA7 fusion protein but not by GST alone (**Fig. 4a**). Pre-incubation of the GST-ATXN7-SCA7 fusion protein with EDTA inhibited the interaction with nucleosomes, indicating that the structural determinants defined by the zinc-coordination sphere are required for this interaction (**Fig. 4a**). Upon addition of increasing amounts of mononucleosomes to a solution of <sup>15</sup>N-labelled ATXN7-SCA7, progressive loss of ATXN7-SCA7 resonances from the NMR spectrum was observed, indicating a binding event with kinetic parameters on the micro- to milli-second time-scale. From this experiment, we estimated an upper limit for the K<sub>d</sub> for this interaction of about 10  $\mu$ M, consistent with the detection of an interaction in GST-pull down assays using more stringent washing conditions (**Fig. 4b**).

Since the ATXN7-SCA7 domain contains a high number of positively charged residues, it might interact non-specifically with DNA. The addition of an oligonucleotide to a solution of <sup>15</sup>N-labelled ATXN7-SCA7 did not affect the <sup>1</sup>H-<sup>15</sup>N HSQC spectrum of the domain (data not shown). From this we conclude that the SCA7 domain does not bind non-specifically to DNA. The remaining possibility was that the domain binds histones. In GST-pull down assays, H2A-H2B dimers were specifically retained by the native GST-ATXN7-SCA7 fusion protein but not by GST alone nor by the fusion protein pre-treated with EDTA (**Fig. 4c**). To further determine whether this binding requires the N-terminal tail of histones, we performed similar experiments using histone octamers reconstituted from histones deleted of the N-terminal tails. The interaction was maintained in similar experiments



using histone octamers reconstituted from histones of which the N-terminal tails had been deleted, suggesting that ATXN7-SCA7 interacts with the histone core regions or the C-terminal extension of one of the histones (**Fig. 4d**).

As the SCA7 domain of ATXN7 is highly conserved through evolution, we then analyzed the nucleosome-binding properties of the domain from *S. cerevisiae* ySgf73. GST-ySgf73-SCA7 fusion protein retained nucleosomes specifically. As for GST-ATXN7-SCA7, this interaction was lost when the GST-ySgf73-SCA7 fusion protein was incubated with EDTA (**Fig. 4e**). In conclusion, we have identified a novel activity for the SCA7 domain of hATXN7/ySgf73, binding to nucleosomes through a direct interaction with the histone H2A-H2B dimer that is evolutionarily conserved.

### **H2B ubiquitination does not enhance hATXN7-SCA7 nucleosome-binding**

Since ATXN7 is a subunit of the deubiquitination module of SAGA that removes ubiquitin from monoubiquitinated H2B and binds histone H2A/H2B dimers through its SCA7 domain, we asked whether this binding is affected by monoubiquitination of histone H2B. We repeated in-vitro GST pull-down experiments using mononucleosomes prepared from HeLa cells and examined the amount of monoubiquitinated H2B retained by the SCA7 domain of hATXN7. Monoubiquitinated H2B was retained specifically by the SCA7 domain of hATXN7 in a ratio comparable to that observed in the input material, suggesting that the binding of ATXN7-SCA7 to nucleosomes is not enhanced by H2B ubiquitination (**Fig. 4f**). Indeed, a stronger interaction of this domain with the monoubiquitinated form of H2B would result in a high enrichment of monoubiquitinated H2B in the retained material as this modification is known to represent only 1 to 2 % of total histone H2B in HeLa cells.

### **Role of the C-terminal helix of ATXN7-SCA7 in nucleosome binding**

The structural differences between the SCA7 domains of ATXN7 and ATXN7L3 reflect sequence differences of these domains and suggest that they might have different functions. GST-pull down experiments using a GST-ATXN7L3-SCA7 fusion protein revealed that the SCA7 domain from ATXN7L3 could not retain either mononucleosomes or histones H2A-H2B dimers (**Fig. 4a-d**). These results indicate that the nucleosome-binding property of the SCA7 domain is conserved in hATXN7/ySgf73 orthologs but is absent in ATXN7L3.

The different organization of the two helices found in the SCA7 domains of ATXN7 and ATXN7L3 suggested that the nucleosome-binding properties of ATXN7 depends on sequence elements located outside the zinc-binding sequence of the SCA7 domain. To address this question, we generated and expressed mutant forms of the GST-ATXN7-SCA7 fusion protein, in which either the N-terminal or the C-terminal residues of the domain were deleted (ATXN7- $\Delta$ Nter and ATXN7- $\Delta$ Cter respectively) (**Fig. 5a-c**).

GST-pull down experiments using mononucleosomes revealed that the N-terminal deletion mutant of the ATXN7-SCA7 domain interacted with nucleosomes as efficiently as the full-length domain. In contrast, deletion of the C-terminal residues of the domain completely abolished the interaction (**Fig. 5b**). Superposition of the  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR spectrum of truncated ATXN7- $\Delta$ Cter on that of ATXN7-SCA7 domain revealed that the mutation disrupts  $\alpha$ 2 helix but leaves the rest of the structure

unaffected (**Fig. 5c**). This result, which emphasizes the role of the C-terminal helix of ATXN7-SCA7 in determining the nucleosome-binding property of ATXN7, was further reinforced by other mutagenesis experiments where the Gly-Arg-Arg motif, found specifically in ATXN7 homologs (position 377-379) in the elbow between the two helices, was mutated into the Tyr-Phe-Leu motif found in the ATXN7L3 domain. This mutation abolishes nucleosome binding while the mutation of the strictly conserved Arg360, located in the zinc finger fold and exposed to solvent, into an alanine did not affect nucleosome binding, suggesting that the core SCA7 zinc finger is not directly involved in the interaction (**Fig. 5b**). Together, these observations suggest that the orientation of the C-terminal helix within the SCA7 domain of ATXN7, which is strikingly different from that of ATXN7L3, contributes to the nucleosome binding properties of this domain.

## Discussion

We report here the solution structures of two SCA7 zinc finger domains belonging to the hATXN7 and hATXN7L3 subunits of the SAGA complex. This analysis reveals that these highly conserved domains display two distinct types of structural organization around a similar core zinc finger fold. Complementing these structural differences, our biochemical assays demonstrate that the SCA7 domain of hATXN7/ySgf73, but not that of hATXN7L3, binds nucleosomes, thus identifying a new chromatin-anchoring module within SAGA. Structure/function analysis of these SCA7 domains suggests that nucleosome binding is not defined by the zinc finger region. Instead, the structure of the zinc finger flanking regions, in particular the C-terminal  $\alpha$ -helix of ATXN7, appears to be a crucial determinant of nucleosome binding.

The SCA7 class of zinc finger is characterized by specific sequence features, defined by a Cys-Gly-Val motif and a large insertion between the first two zinc-coordinating residues. These sequence features translate into a specific three-dimensional fold that has no relatives in the PDB database. The lack of  $\alpha$ -helical secondary structure within the zinc coordination sphere, which is mostly organized in loop structures, is reminiscent of the zinc ribbon family. However, the position of the  $\beta$ -hairpin between the first two cysteine residues is specific to the SCA7 motif, adding therefore a new member to a rapidly expanding list of folds stabilized by the coordination of zinc<sup>23</sup>.

Originally discovered as DNA-binding motifs, the growing number of examples of ZnFs involved in protein-protein interactions is changing our vision of these small structural units<sup>24</sup>. It has been recognized recently that eukaryotic proteomes contains a higher proportion of zinc-binding proteins than archeal and bacterial proteomes, and this may be attributed to their role in the fine regulation of cellular processes<sup>22</sup>. The structures of the SCA7 domains of hATXN7 and hATXN7L3 provide striking new evidence of the diversity and plasticity of ZnFs as protein interaction mediators, characteristics that have only recently been revealed. For instance, the interaction between the TAZ1 domain of the transcriptional co-activator CBP and two distinct transcription factors, HIF-1 $\alpha$  and CITED2, involves different, albeit overlapping regions of the TAZ ZnF<sup>25,26</sup>. Another example was found in LIM ZnFs where the crystal structure of LMO4-ldb1-LID complex revealed that the nature

of the side-chain interactions between Ldb1-LID and the two LIM ZnFs of LMO4 differs between the two LIM domains<sup>27</sup>. The case of the SCA7 domain is particularly compelling because the interaction between ATXN7 and nucleosomes is mediated by residues located remotely from the zinc coordination sphere. The inability of ATXN7L3 to bind nucleosomes is associated with a difference in organization of secondary structure elements around an otherwise conserved zinc-binding scaffold. To our knowledge, this is the first example of a class of ZnF domains in which distinct interaction properties are associated with distinct folds.

Surprisingly, throughout the human proteome, SCA7 domains are exclusively found in ATXN7 and ATXN7L3, two distinct subunits of the SAGA deubiquitination module. According to our interaction studies<sup>17</sup>, the SCA7 domain of ATXN7 is in close proximity to the catalytic subunit, USP22. Here, we demonstrate that the binding of the SCA7 domain of ATXN7 to nucleosomes involves the core or the C-terminal end of the histone H2A and H2B dimer. Interestingly, this region is located on the lateral face of the nucleosome and contains the ubiquitinated Lys120 of H2B<sup>28,29</sup>. We propose that the SAGA deubiquitination module is recruited to the lateral face of the nucleosome through the interaction of the SCA7 domain of hATXN7/ySgf73 with the histones, thereby positioning hUSP22/yUbp8 in front of the C-terminal tail of histone H2B. It should be noted that this recruitment would provide an additional regulatory mechanism for hUSP22/yUbp8 independent of its catalytic activation which relies on its association with other SAGA subunits<sup>18,19</sup>.

For the USP family of deubiquitinating enzymes, the recognition of an ubiquitinated substrate appears to be mediated largely by the interaction of the catalytic domain with the ubiquitin moiety. Indeed, crystal structures from HAUSP and USP14 in complex with ubiquitin aldehyde revealed that ubiquitin alone interacts with the catalytic domain<sup>30,31</sup>. The level of specificity of each USP has not yet been established<sup>32</sup> but we propose that, in the case of the deubiquitination activity of SAGA, the SCA7 domain of ATXN7 contributes to the specificity of substrate recognition by targeting USP22 to the lateral surface of nucleosomes. If the SCA7 domain of ATXN7L3 also serves as a protein recognition module, it may play a role in specific targeting of non-histone proteins in higher eukaryotes. Indeed, while ubiquitinated H2B is the only substrate of USP22 that has been identified *in vivo*, studies in *Drosophila*<sup>33</sup> suggest that other ubiquitinated non-histone proteins could also be targeted by the deubiquitination machinery of SAGA.

ySgf73, and most likely its human ortholog hATXN7, is known to be the bridging subunit between SAGA and its deubiquitination module. *In vitro* pull-down assays suggested that the SCA7 domain of hATXN7 mediates association with other SAGA subunits and deletion mutants indicate that both the SCA7 domain and neighboring sequences play a role in the incorporation of ySgf73 in SAGA<sup>19,21</sup>. ySgf73, together with other subunits of the deubiquitination module, also plays a crucial role in tethering activated genes to the nuclear periphery, thereby coupling transcription and mRNA export<sup>19,34,35</sup>. Moreover, the deletion of *Sgf73* leads to the dissociation of the TREX2 mRNA export complex<sup>19,35</sup>. A central region of ySgf73 including the SCA7 domain was shown to be required for this function. However, it remains to be determined whether this effect acts through or independently of nucleosome binding to the SCA7 domain.

In summary, our structural analysis of the SCA7 ZnFs from ATXN7 and ATXN7L3 reveals two distinct types of structural organization. These structural differences are related to the specific ability of ATXN7, but not of ATXN7L3, to bind nucleosomes and suggest a speciation mechanism involving a structural re-organization of secondary structure elements around a conserved zinc-coordination motif. Our study supports a model in which the SCA7 ZnF of ATXN7 provides an additional regulatory mechanism of SAGA deubiquitination activity by targeting USP22 to chromatin.



## Methods

### Plasmids

Plasmids for the expression of GST-SCA7 fusion proteins (ATXN7 : 330-401, ATXN7- $\Delta$ Nter : 338-401, and ATXN7- $\Delta$ Cter : 330-388) were described previously<sup>21</sup>. The DNA coding for SCA7 domains of ySgf73 (211-283) and of hATXN7L3 (197-276) were amplified by PCR and inserted into pGEX-4T1 (GE Healthcare).

### Site-directed mutagenesis

Point mutants were generated using the QuikChange Lightning Site-Directed Mutagenesis Kit (Stratagene). The pGEX-SCA7 plasmid was used as the matrix for the mutagenesis. Mutagenic primers were designed using Stratagene's web-based QuikChange Primer Design Program available online at <http://www.stratagene.com/qcprimerdesign>.

### Expression and purification of SCA7 domains

Expression of all proteins was achieved using *E. coli* BL21 [DE3] cells (Novagen) in LB medium to obtain unlabelled samples or in <sup>15</sup>N, <sup>13</sup>C-labelled M9 minimal medium supplemented with 10 % v/v of <sup>15</sup>N or <sup>15</sup>N, <sup>13</sup>C SILANTES OD 2 to produce labelled proteins. Cells were grown at 37°C to an absorbance of 0.5 at 600 nm and the temperature was then lowered to 25°C. Growth was allowed to continue until cells reached an absorbance of 0.6-0.7 at 600 nm. Expression was induced by adding IPTG to a final concentration of 1 mM (Euromedex) and ZnSO<sub>4</sub> was also added to attain a concentration of 10  $\mu$ M of Zn<sup>2+</sup> in the culture. Cells were further grown overnight at 25°C, collected by low-speed centrifugation, re-suspended in lysis buffer (20 mM Tris-HCl pH 7.5, 250 mM NaCl and 2 mM DTT) and lysed by sonication. The soluble fraction recovered by high-speed centrifugation was mixed with glutathione sepharose beads (GE Healthcare). After 1 h incubation, the supernatant was removed and the resin washed extensively with lysis buffer. The resin was then re-suspended in 2 ml of lysis buffer and bovine thrombin (Sigma) was added overnight at 4°C, to cleave the GST tag. The supernatant was recovered and loaded on to a Superdex G75 (HiLoad 16/60) gel filtration column equilibrated with phosphate buffer (50 mM phosphate, pH 7.0, 200 mM NaCl, 2 mM DTT). The purified proteins were concentrated on 5K amicon concentrators (Vivaspin).

### NMR spectroscopy

NMR samples contained 0.2 mM of <sup>13</sup>C, <sup>15</sup>N-labelled or unlabelled proteins. Most NMR experiments were recorded at 298K on a Bruker DRX600 spectrometer. Backbone and side-chain assignment were obtained from 3D HN(CO)CA, HNCA, HN(CO)CACB, HNCACB, HNCO and HCCH-TOCSY spectra. For structure calculations, homonuclear 2D NOESY spectra were recorded with a mixing time of 150 ms on a Bruker 950 MHz spectrometer. The protonation state of the Zn-coordinating His367 of ATXN7 was set to N $\epsilon$ <sub>2</sub> based on the frequencies of the <sup>15</sup>N nuclei of the imidazole ring, measured using a long-range <sup>1</sup>H-<sup>15</sup>N HSQC spectrum. All spectra were processed

using NMRPipe<sup>36</sup> and analysed using CARA.

### Structure calculation

950 MHz 2D NOESY and 600 MHz 3D <sup>15</sup>N- and <sup>13</sup>C-edited NOESY spectra together with backbone  $\Phi$  and  $\Psi$  angles derived from resonance assignments using the programme TALOS<sup>37</sup>, were used as input data to the semi-automatic ATNOS/CANDID procedure<sup>38</sup>. Initial runs of calculations without zinc coordination constraints allowed unambiguous identification of residues involved in zinc ion binding. Floating chirality in structure calculations allowed stereospecific assignment of asymmetric centers. The final set of structures was refined in explicit water using the RECOORD protocol<sup>39</sup>. The final ensemble of 20 lowest energy structures contained no distance violations greater than 5Å, nor dihedral angle violations greater than 5°. The Ramachandran statistics are provided below.

	ATXN7	ATXN7L3
<b>Backbone dihedral angles statistics</b>		
Residues in most favorable regions (%)	72.4	76.0
Residues in additional favorable regions(%)	27.1	21.0
Residues in generously favorable regions (%)	0.6	2.0
Residues in disallowed regions (%)	0.0	1.1

Three-dimensional structures of ATXN7-SCA7 and ATXN7L3-SCA7 have been deposited in the Protein Data Bank under accession numbers 2kkk and 2kkt respectively.

### GST-pull down assay

Glutathione sepharose beads (GE Healthcare) equilibrated in binding buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 0,1% NP40, 10  $\mu$ M ZnCl<sub>2</sub>, 1 mM DTT and protease inhibitors), were incubated with bacterial lysates containing GST or GST-SCA7 domains of hATXN7, hATXN7L3 or ySgf73. Purified mononucleosomes, recombinant H2A-H2B dimers or recombinant tail-less histone octamers (in which the first 13 amino acids of H2A, the first 20 amino acids of H2B, the first 26 amino acids of H3, and the first 19 amino acids of H4 are removed) were added to the GST-coated sepharose beads and incubated overnight at 4°C in binding buffer. Beads were washed with binding buffer containing 250 mM NaCl. 30% of the bound proteins were analysed by SDS-PAGE followed by Coomassie blue R250 coloration and 30% were analysed by immunoblot analysis.

### Antibodies

The monoclonal anti-H2B antibody (5HH2-2A8) was obtained by immunization of mice with a branched peptide corresponding to residues 111-125 (VSEGTKAVTKYTSSK) of H2B with the 12 last residues of ubiquitin linked to Lys120. This antibody recognizes both forms of H2B (ubiquitinated or unmodified) with the same efficiency. An antibody against H3 (ab1791) was



obtained from Abcam. An antibody against ubiquitinated H2B (NRO3) was obtained from Médimabs.





## Acknowledgments

We thank Dominique Helmlinger for critical reading of this manuscript, Guillaume Lang for the gift of material and helpful discussion, Claude Ling and Matthieu Stierle for technical support and Eric Guittet for providing high-field NMR time. We also thank Marc Vitorino and Mustapha Oulad-Abdelghani for their involvement in early stages of this work. Research in the author's laboratories is supported by grants from the Fondation pour la Recherche Médicale (FRM), the Institut National du Cancer (INCA), Ligue contre le cancer, Structural Proteomics In Europe (SPINE-2) program and the Agence Nationale de la Recherche (ANR). Access to high-field NMR spectrometers was granted by the CNRS through the TGE RMN program.

## References

1. Martinez, E. Multi-protein complexes in eukaryotic gene transcription. *Plant Mol Biol* **50**, 925-47 (2002).
2. Naar, A.M., Lemon, B.D. & Tjian, R. Transcriptional coactivator complexes. *Annu Rev Biochem* **70**, 475-501 (2001).
3. Baker, S.P. & Grant, P.A. The SAGA continues: expanding the cellular role of a transcriptional co-activator complex. *Oncogene* **26**, 5329-40 (2007).
4. Nagy, Z. & Tora, L. Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. *Oncogene* **26**, 5341-57 (2007).
5. Wu, P.Y., Ruhlmann, C., Winston, F. & Schultz, P. Molecular architecture of the *S. cerevisiae* SAGA complex. *Mol Cell* **15**, 199-208 (2004).
6. Brownell, J.E. et al. Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* **84**, 843-51 (1996).
7. Gansheroff, L.J., Dollard, C., Tan, P. & Winston, F. The *Saccharomyces cerevisiae* SPT7 gene encodes a very acidic protein important for transcription in vivo. *Genetics* **139**, 523-36 (1995).
8. Lee, K.K. & Workman, J.L. Histone acetyltransferase complexes: one size doesn't fit all. *Nat Rev Mol Cell Biol* **8**, 284-95 (2007).
9. Balasubramanian, R., Pray-Grant, M.G., Selleck, W., Grant, P.A. & Tan, S. Role of the Ada2 and Ada3 transcriptional coactivators in histone acetylation. *J Biol Chem* **277**, 7989-95 (2002).
10. Gamper, A.M., Kim, J. & Roeder, R.G. The STAGA subunit ADA2b is an important regulator of human GCN5 catalysis. *Mol Cell Biol* **29**, 266-80 (2009).
11. Daniel, J.A. et al. Deubiquitination of histone H2B by a yeast acetyltransferase complex regulates transcription. *J Biol Chem* **279**, 1867-71 (2004).
12. Henry, K.W. et al. Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. *Genes Dev* **17**, 2648-63 (2003).
13. Ingvarsdottir, K. et al. H2B ubiquitin protease Ubp8 and Sgf11 constitute a discrete functional module within the *Saccharomyces cerevisiae* SAGA complex. *Mol Cell Biol* **25**, 1162-72 (2005).

14. Lee, K.K., Florens, L., Swanson, S.K., Washburn, M.P. & Workman, J.L. The deubiquitylation activity of Ubp8 is dependent upon Sgf11 and its association with the SAGA complex. *Mol Cell Biol* **25**, 1173-82 (2005).
15. Powell, D.W. et al. Cluster analysis of mass spectrometry data reveals a novel component of SAGA. *Mol Cell Biol* **24**, 7249-59 (2004).
16. Zhang, X.Y. et al. The putative cancer stem cell marker USP22 is a subunit of the human SAGA complex required for activated transcription and cell-cycle progression. *Mol Cell* **29**, 102-11 (2008).
17. Zhao, Y. et al. A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing. *Mol Cell* **29**, 92-101 (2008).
18. Bonnet, J., Romier, C., Tora, L. & Devys, D. Zinc-finger UBPs: regulators of deubiquitylation. *Trends Biochem Sci* **33**, 369-75 (2008).
19. Kohler, A., Schneider, M., Cabal, G.G., Nehrbass, U. & Hurt, E. Yeast Ataxin-7 links histone deubiquitination with gene gating and mRNA export. *Nat Cell Biol* **10**, 707-15 (2008).
20. Lee, K.K., Swanson, S.K., Florens, L., Washburn, M.P. & Workman, J.L. Yeast Sgf73/Ataxin-7 serves to anchor the deubiquitination module into both SAGA and Slik(SALSA) HAT complexes. *Epigenetics Chromatin* **2**, 2 (2009).
21. Helmlinger, D. et al. Ataxin-7 is a subunit of GCN5 histone acetyltransferase-containing complexes. *Hum Mol Genet* **13**, 1257-65 (2004).
22. Andreini, C., Banci, L., Bertini, I. & Rosato, A. Zinc through the three domains of life. *J Proteome Res* **5**, 3173-8 (2006).
23. Krishna, S.S., Majumdar, I. & Grishin, N.V. Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res* **31**, 532-50 (2003).
24. Gamsjaeger, R., Liew, C.K., Loughlin, F.E., Crossley, M. & Mackay, J.P. Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends Biochem Sci* **32**, 63-70 (2007).
25. Dames, S.A., Martinez-Yamout, M., De Guzman, R.N., Dyson, H.J. & Wright, P.E. Structural basis for Hif-1 alpha /CBP recognition in the cellular hypoxic response. in *P Natl Acad Sci Usa* Vol. 99 5271-6 (2002).
26. Freedman, S.J. et al. Structural basis for negative regulation of hypoxia-inducible factor-1alpha by CITED2. in *Nat Struct Biol* Vol. 10 504-12 (2003).
27. Deane, J.E. et al. Structural basis for the recognition of ldb1 by the N-terminal LIM domains of LMO2 and LMO4. *EMBO J* **22**, 2224-33 (2003).
28. Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. & Richmond, T.J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-60 (1997).
29. Zhang, Y. Transcriptional regulation by histone ubiquitination and deubiquitination. *Genes Dev* **17**, 2733-40 (2003).
30. Hu, M. et al. Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in complex with ubiquitin aldehyde. *Cell* **111**, 1041-54 (2002).

31. Hu, M. et al. Structure and mechanisms of the proteasome-associated deubiquitinating enzyme USP14. *EMBO J* **24**, 3747-56 (2005).
32. Nijman, S.M. et al. A genomic and functional inventory of deubiquitinating enzymes. *Cell* **123**, 773-86 (2005).
33. Poeck, B., Fischer, S., Gunning, D., Zipursky, S.L. & Salecker, I. Glial cells mediate target layer selection of retinal axons in the developing visual system of *Drosophila*. *Neuron* **29**, 99-113 (2001).
34. Kohler, A. et al. The mRNA export factor Sus1 is involved in Spt/Ada/Gcn5 acetyltransferase-mediated H2B deubiquitinylation through its interaction with Ubp8 and Sgf11. *Mol Biol Cell* **17**, 4228-36 (2006).
35. Pascual-Garcia, P. et al. Sus1 is recruited to coding regions and functions during transcription elongation in association with SAGA and TREX2. *Genes Dev* **22**, 2811-22 (2008).
36. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. in *J Biomol NMR* Vol. 6 277-93 (1995).
37. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. in *J Biomol NMR* Vol. 13 289-302 (1999).
38. Herrmann, T., Güntert, P. & Wüthrich, K. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. in *J Biomol NMR* Vol. 24 171-89 (2002).
39. Nederveen, A.J. et al. RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. in *Proteins* Vol. 59 662-72 (2005).
40. Schanda, P., Kupce, E. & Brutscher, B. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. in *J Biomol NMR* Vol. 33 199-211 (2005).

## Figure legends

### Figure 1 : Conservation of the SCA7 domain in ATXN7 and ATXN7L3 orthologs

(a) Schematic representation of the SAGA deubiquitination module (b) Domain organization of the SCA7 family. The highly conserved SCA7 domain (gray box) is shared by all hATXN7/ySgf73 orthologs and a subset of ATXN7L3 orthologs. Sequence conservation of the N-terminal ZnF domain distinguishes hATXN7L3/ySgf11 orthologs (ZnF-Sgf11 : white box) from hATXN7/ySgf73 orthologs (ZnF-Sgf73 : black box). (c) Multiple sequence alignment of the SCA7 domain from selected species reveals the sequence similarities and differences that allow the SCA7 domains from hATXN7/ySgf73 and ATXN7L3 orthologs to be distinguished. SwissProt/TrEmbl/RefSeq accession numbers are: ATXN7 : *Homo sapiens* (O15265.1), *Mus musculus* (Q8R4I1.1), *Xenopus laevis* (NP\_001085969.1), *Danio rerio* (XP\_001341475), *Apis mellifera* (translated from : NW\_001253266.1), *Homo sapiens* ATXN7L1 (Q9ULK2.2), *Homo sapiens* ATXN7L2 (Q5T6C5.1), Sgf73 : *Saccharomyces cerevisiae* (P53165.1), *Schizosaccharomyces pombe* (O94397.1), *Candida glabrata* (XP\_449594.1), *Ashbya gossypii* (NP\_985453.1), *Aspergillus nidulans* (XP\_680742.1), *Candida albicans* (Q59S68), *Neurospora crassa* (XP\_958273.1), *Cryptococcus neoformans* (XP\_572066), ATXN7L3 : *Homo sapiens* (Q14CW9.1), *Mus musculus* (A2AWT3.1), *Danio rerio* (A1L209.1), *Apis mellifera* (translated from : NW\_001253391.1). Strictly invariant residues, defining the SCA7 ZnF signature are indicated by an asterisk and zinc-chelating residues are colored in magenta. The color-coding scheme has been chosen to stress the conservation of biochemical properties within each group of SCA7 ZnF and to highlight sequence differences between the ATXN7 and ATXN7L3 domains with hydrophobic, positive and negative residues colored in green, red and blue respectively. Glycine residues are colored in yellow and proline are shown in inverted contrast. Secondary structure elements of the SCA7 ZnFs from ATXN7 and ATXN7L3 are drawn above and below the corresponding sequences.  $\alpha$ -helices are shown as rectangles and  $\beta$ -strands as arrows.

### Figure 2 : Solution structures of the SCA7 domains of human ATXN7 and ATXN7L3

(a,b) Deviation of  $C\alpha$  chemical shifts from random coil values and positions of secondary structure elements for (a) ATXN7-SCA7 and (b) ATXN7L3-SCA7 ZnFs. The positions of zinc-coordinating residues are indicated with orange triangles. (c,d) Stereoview of the  $C\alpha$  traces from the ten lowest energy structures of (c) ATXN7-SCA7 and (d) ATXN7L3-SCA7 ZnFs. Only the folded parts of the proteins are represented. The heavy atoms of the zinc-coordinating residues are shown in orange and the position of the zinc ion is indicated by a gray sphere. Helices  $\alpha_1$  and  $\alpha_2$  are coloured in blue and red respectively.

### Figure 3: Architecture of SCA7 zinc fingers

(a,b) Ribbon representation of (a) ATXN7-SCA7 and (b) ATXN7L3-SCA7 proteins showing the hydrophobic contacts that hook the two  $\alpha$ -helices onto the zinc finger. For clarity, the  $\beta$  secondary structures are represented as narrow tubes. (c,d) Conserved interactions define the relative

orientations of the two  $\alpha$ -helices within (c) ATXN7-SCA7 and (d) ATXNL3-SCA7 proteins. The side-chains of conserved hydrophobic, positive or negative residues are shown using green, blue or red sticks respectively. The figures were prepared using Pymol (DeLano Scientific LLC).

#### **Figure 4 : The SCA7 ZnF of hATXN7/ySgf73 binds nucleosomes**

(a-d) SCA7 domains from ATXN7 or ATXN7L3 fused to GST (GST-ATXN7 and GST-ATXN7L3) or GST alone were immobilized on glutathione-sepharose beads and incubated with mononucleosomes (a-b), H2A-H2B dimers (c) or tail-less histone octamers (d). The material retained on the column was analyzed by Coomassie blue staining or western blotting using anti-H2B or anti-H3 antibodies, as indicated. The interaction with nucleosomes was still detected after stringent washes at 350 mM salt concentration (b). (e) The SCA7 domains from both hATXN7 and its yeast ortholog ySgf73 retain mononucleosomes. All interactions were lost when the GST-ATXN7-SCA7 (or GST-Sgf73-SCA7) fusion protein was pre-incubated with EDTA. (f) The ATXN7-SCA7 domain does not bind monoubiquitinated H2B preferentially. GST-pull down experiments were analyzed using anti-H2B or anti-H2Bub antibodies as indicated.

#### **Figure 5 : Role of secondary structure elements of the SCA7 domain of ATXN7 in nucleosome binding**

(a) Schematic representation of ATXN7-SCA7 mutant domains generated in this study (b) GST-pull down experiments using mononucleosomes and mutant domains, shown in (a) were performed as in Figure 4 and analyzed by Coomassie blue staining and western blotting using anti-H2B and anti-H3 antibodies as indicated. (c) Superposition of HSQC spectra from ATXN7-SCA7 (black) and the mutant ATXN7- $\Delta$ Cter (red). Correlations with significant shifts upon mutation are labelled. Zinc-coordinating residues are boxed.

#### **Supplementary Figure 1 : Dynamical behaviour and structural packing of ATXN7-SCA7 and ATXN7L3-SCA7**

Profiles of peak intensity ratios measured with and without perturbation of aliphatic protons for ATXN7-SCA7 (in blue) and ATXN7L3-SCA7 (in red) domains under fast pulsing conditions using the SOFAST experiment<sup>40</sup>. The sequences of both proteins were aligned using the similarities shown in Figure 1. Secondary structure elements of ATXN7-SCA7 and ATXN7L3-SCA7 are shown in blue and red, respectively. A value of 0.8 indicates a low proton density and flexible regions whereas lower values are associated with high proton density and restricted motions. The aliphatic protons were perturbed using a 4 ppm wide selective 180° pulse centered at 2 ppm prior to the amide proton selective HMQC sequence. The mixing time was set to 200 ms.

**Table 1 : NMR and refinement statistics for protein structures**

	ATXN7	ATXN7L3
NMR distance and dihedral constraints		
Distance constraints		
Total NOE	1589	970
Intra-residue	387	289
Inter-residue		
Sequential ( $ i - j  = 1$ )	474	383
Medium-range ( $ i - j  < 4$ )	376	167
Long-range ( $ i - j  > 5$ )	352	131
Protein-zinc restraints	10	10
Hydrogen bonds	0	0
Total dihedral angle restraints		
$\phi$	29	23
$\psi$	27	23
Structure statistics		
Violations (mean and s.d.)		
Distance constraints (Å)	0.0664 $\pm$ 0.0034	0.0725 $\pm$ 0.0040
Dihedral angle constraints (°)	0.1706 $\pm$ 0.2998	0.1293 $\pm$ 0.1373
Max. dihedral angle violation (°)	0.488	0.356
Max. distance constraint violation (Å)	4.935	2.441
Deviations from idealized geometry		
Bond lengths (Å)	0.0177 $\pm$ 0.0009	0.0147 $\pm$ 0.0007
Bond angles (°)	2.0232 $\pm$ 0.0400	1.8581 $\pm$ 0.0471
Impropers (°)	1.9413 $\pm$ 0.0777	1.8466 $\pm$ 0.0865
Average pairwise r.m.s. deviation* (Å)		
Heavy	1.17 $\pm$ 0.16	1.44 $\pm$ 0.17
Backbone	0.60 $\pm$ 0.19	0.54 $\pm$ 0.12

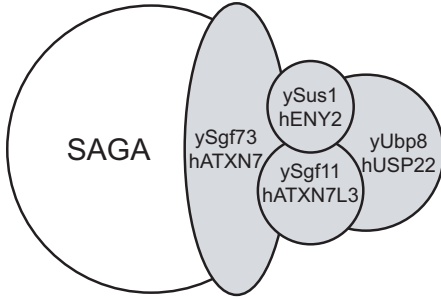
\* Pairwise r.m.s. deviation was calculated among 20 refined structures using the C $_{\alpha}$  positions of residues 342-394 for ATXN7-SCA7 and residues 199-211 and 220-244 for ATXN7L3-SCA7.

The logo of the University of Strasbourg, consisting of two blue curved segments that form a stylized 'S' shape, positioned behind the text.

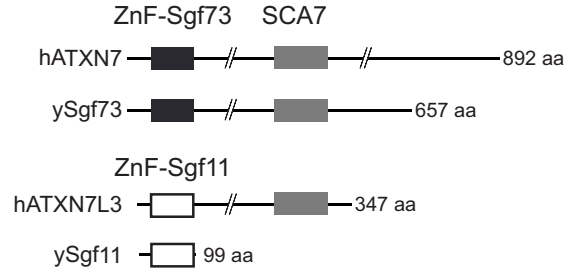
UNIVERSITÉ DE STRASBOURG



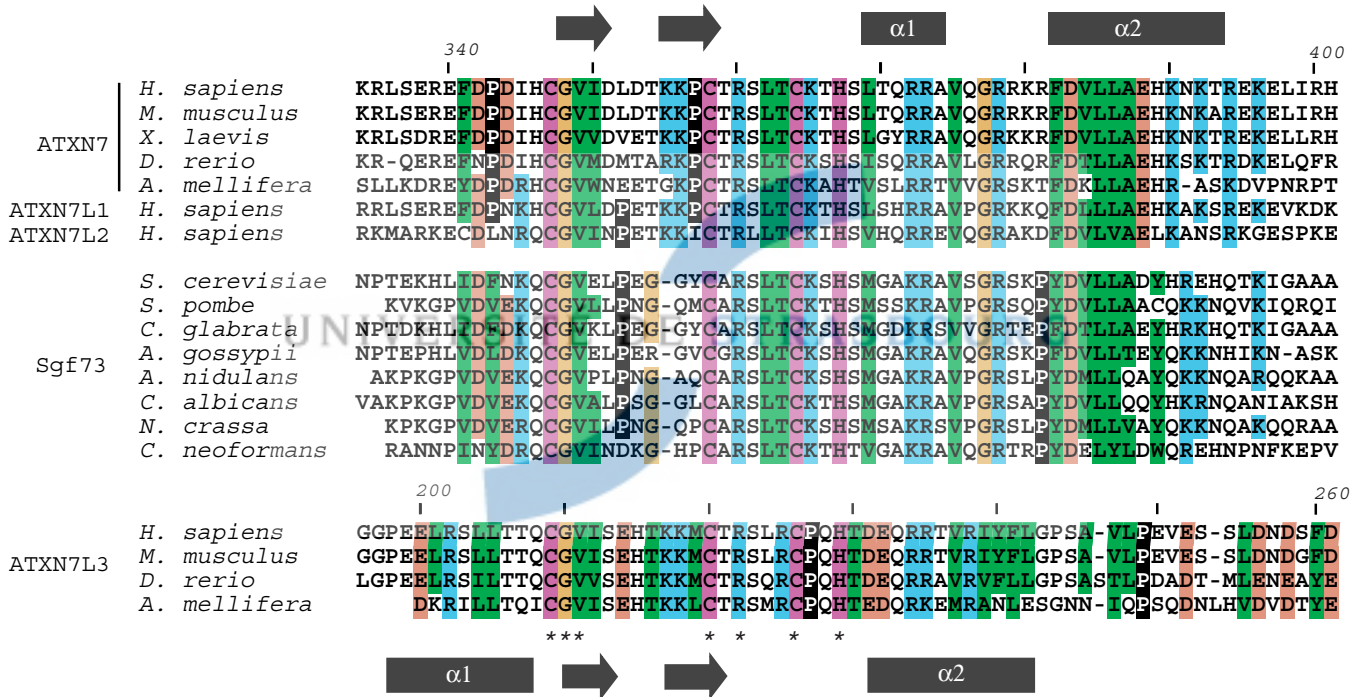
**a**



**b**

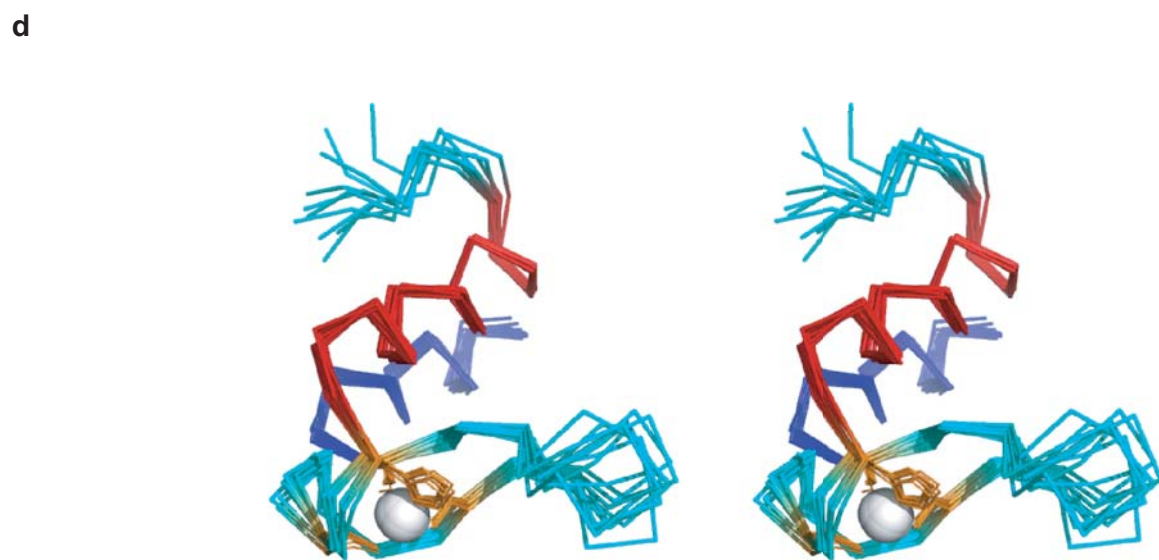
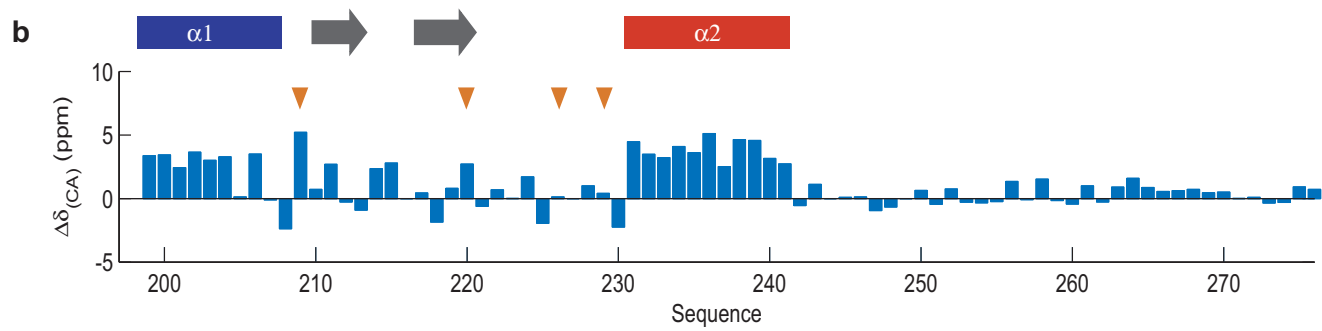
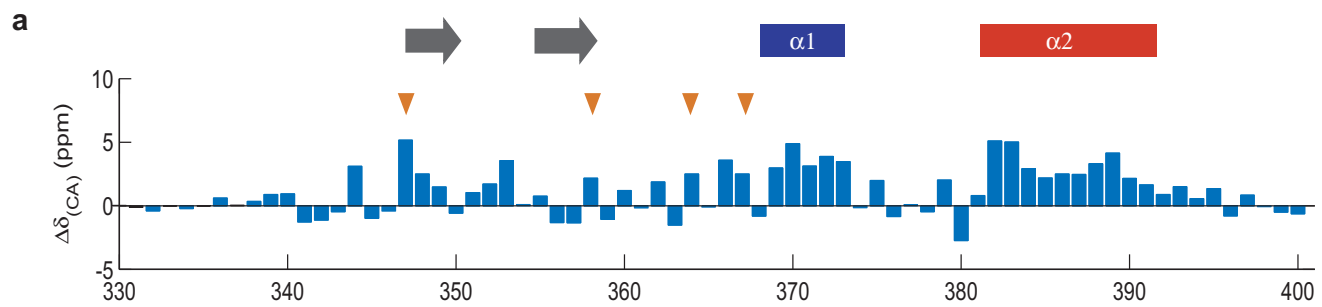


**c**



The logo of the University of Strasbourg, featuring two blue curved segments that form a stylized 'S' shape, positioned behind the text.

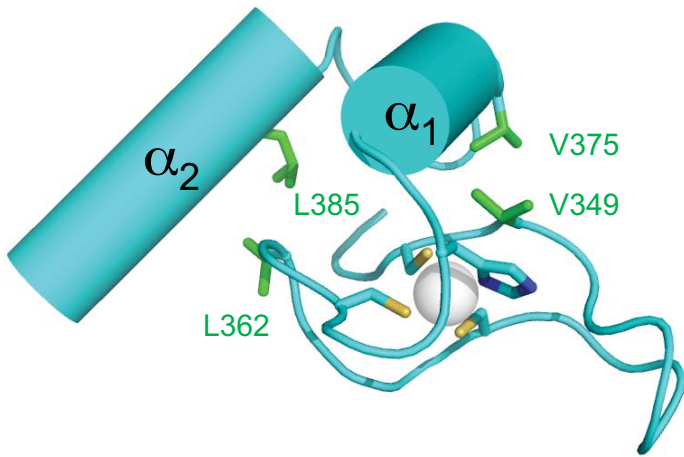
UNIVERSITÉ DE STRASBOURG



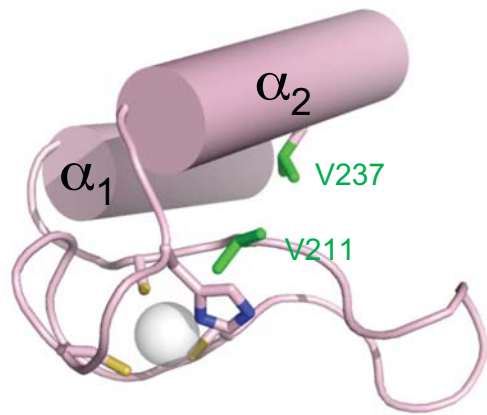
The logo of the University of Strasbourg, featuring two blue curved segments that form a stylized 'S' shape, positioned behind the text.

UNIVERSITÉ DE STRASBOURG

a

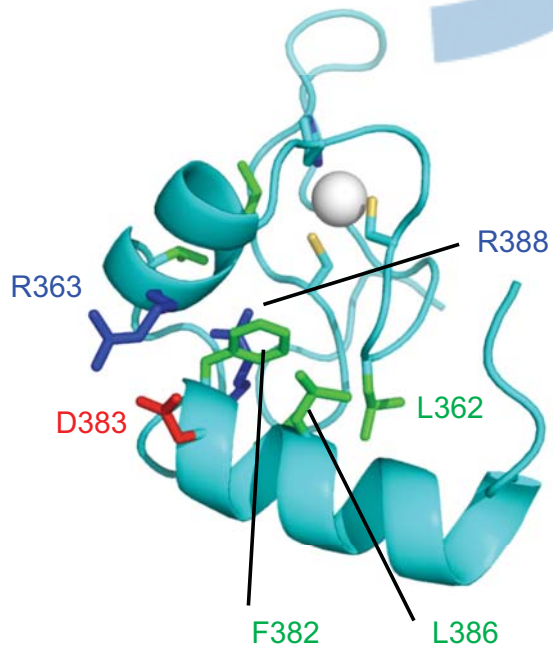


b

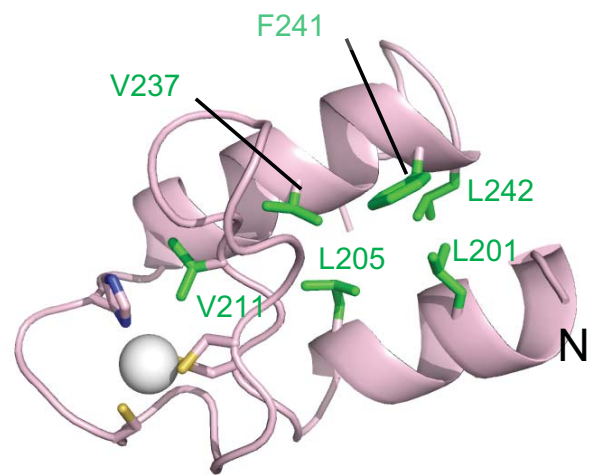


UNIVERSITÉ DE STRASBOURG

c

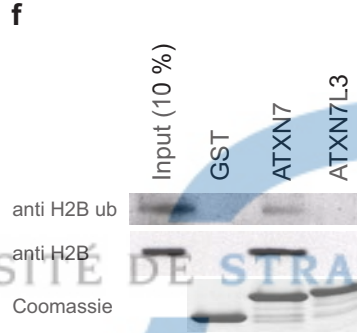
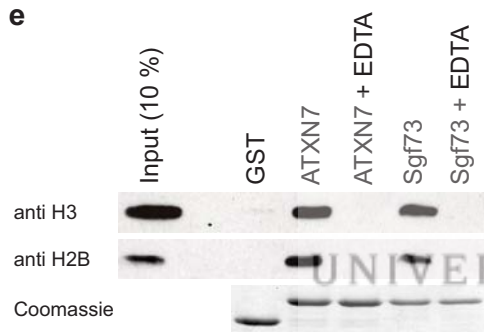
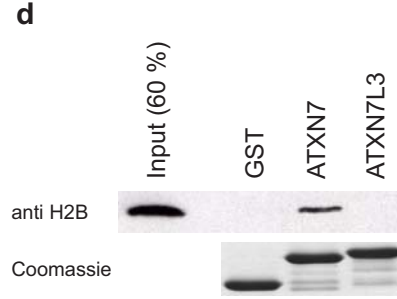
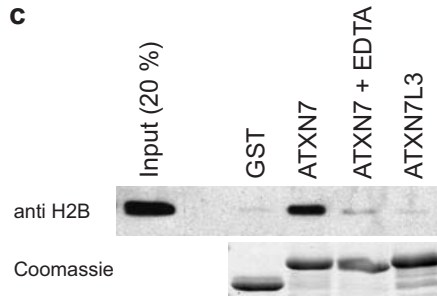
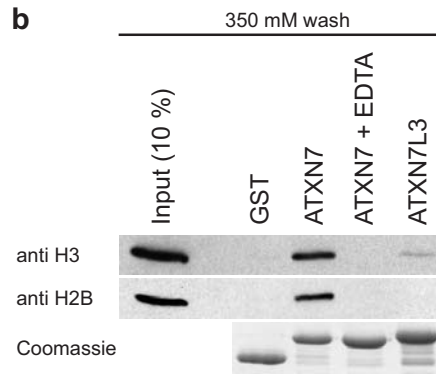
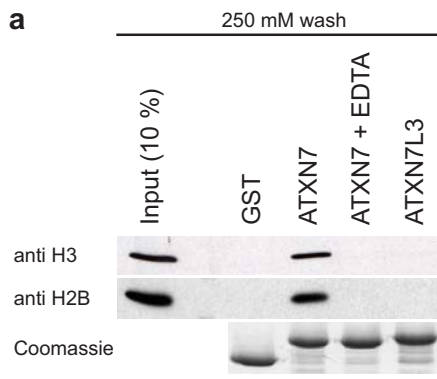


d



The logo of the University of Strasbourg, featuring two blue curved segments that form a stylized 'S' shape, positioned behind the text.

UNIVERSITÉ DE STRASBOURG

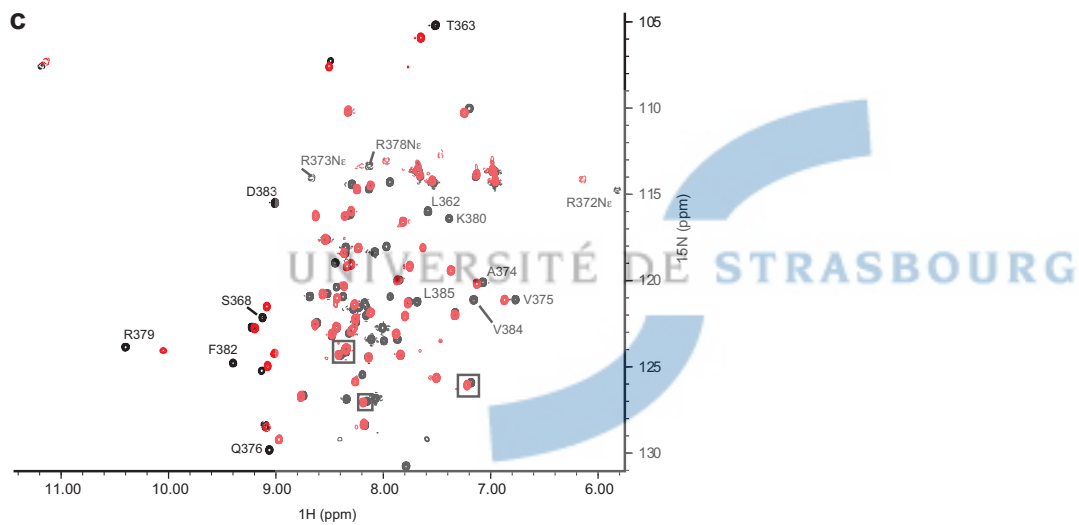
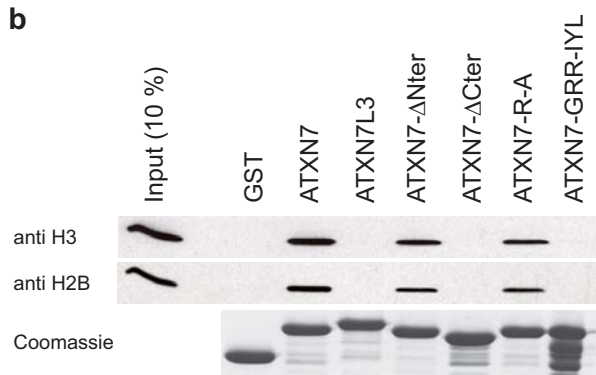
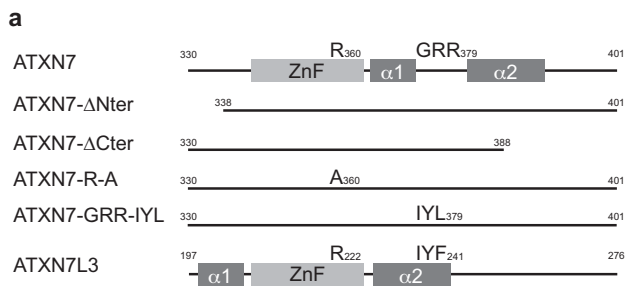


UNIVERSITÉ DE STRASBOURG



The logo of the University of Strasbourg, featuring two blue curved segments that form a stylized 'S' shape, positioned behind the text.

UNIVERSITÉ DE STRASBOURG



The logo of the University of Strasbourg, featuring two blue curved segments that form a stylized 'S' shape, positioned behind the text.

UNIVERSITÉ DE STRASBOURG

### Conclusion and Perspectives

During this thesis, I have studied two families of zinc finger proteins, the androgen receptor DNA binding domain and the SCA7 domain in ATXN7 and ATXN7 L3 that differ in both their fold and recognition properties. These studies have provided me with an illustration of how versatile and rich this family of protein is. Similar experimental approaches were applied to these two families and proved to be highly instructive as to the subtle structural and dynamic features that underline their functions.

Among the range of experiments employed, the measurement of metal exchange kinetics provided particularly interesting results, revealing significant differences in metal binding properties between AR-DBD and SCA7 domains. The high sensitivity of the kinetic exchange rates allows subtle differences in the molecular environment to be observed. The two metal binding sites in AR-DBD, indeed, reveal quite different exchange rates, despite the nature of coordination spheres composed by four cysteine residues, being similar. The T575A mutation of AR-DBD results in changes in exchange rates again emphasizing the sensitivity of this factor. These findings allow us to suggest possible mechanisms of effects of metal substitution on protein functions. Cadmium is considered a toxic metal but the mechanisms of its toxicity were rarely reported. It was only recently that the effects of cadmium on DNA repair mechanisms and apoptosis have been described in several reports.[1] The dominant mechanism allows the toxicity of cadmium to be explained by the creation of an oxygen species associated with an oxidative stress.[2] Our work on AR-DBD suggests that these mechanisms may be more complicated, involving the modulation of gene regulation via alternation of protein-DNA recognition. This hypothesis deserves closer attention and the androgen receptor may provide an ideal model for further investigation.

Our work on AR-DBD has focused on the description of two domains, differing in a single amino acid and the discovery of changes in charge property of the conserved histidine provided some clues about possible molecular mechanisms of altered DNA recognition. However, this work needs to be completed through studies of the interaction of the two AR-DBDs with specific and non-specific DNA response elements. We investigated successfully the interactions of AR-DBD with its half-site DNA response element but were unable to discover optimal conditions with full-site response elements due to line broadening. The observed line broadening may result from multiple exchanges between the DNA and the DBD domains and this prompted us to reconsider the boundaries of the DBD domain by addition of the C-terminal extension (CTE) to the DBD domain. In some type of nuclear receptors, the CTE was shown to enhance DNA binding activity.[3, 4] Unfortunately, this work could not be

completed due to biochemical difficulties and a lack of time. In addition, the covalent dimerization of AR-DBD on its full-site may be necessary to simplify the system by restraining the number of possible states. Such methods have proven to be powerful in the study of binding of the Lac repressor headpiece to DNA.[5] [Kalodimos 2005 Science]. Other approaches have been considered to address the question of the geometry of binding and provide evidence that the head-to-head geometry found in the crystal structure is the major complex type in solution.[6] The use of stable nitroxide spin labels could provide an interesting direction for future studies, since the distance between two electronic spins could be obtained with good precision using EPR double resonance techniques such as DEER experiments. The labeling is usually performed by the chemical modification of cysteine residues, which, in the case of zinc finger domains, may prove challenging.

Similar perspectives were offered by our pioneering work on SCA7 zinc finger domains that led to discovery of a new fold and a novel member of the growing family of histone interacting domains. This work leaves many questions unanswered, in particular, about the mechanism of interactions with histones. Preliminary data seems to designate the domain's flanking regions as being directly involved in the interaction. Indeed, it was possible to observe magnetization transfer from the H2A.H2B histone dimer to the free ataxin7 SCA7 domain. This work should be pursued, since this interaction proves to be highly dependent on the state of the histone preparation. The interaction experiment could be performed using  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled histones to map the interaction site on the histone, and possibly to obtain information allowing a model for interactions of the SCA7 domain with nucleosomes to be proposed. The ATXN7 and ATXN7 L3 proteins provide a very interesting system whereby different structural domains mediate different interactions within a large complex. The role of the linker sequences between the structured domains would be of particular interest to be studied using NMR methods. A first glimpse of what might be gained from such studies was provided by analysis of the dynamic behavior of the disordered C-terminal tail of the ATXN7 L3 SCA7 domain. Despite a complete lack of data on the biological functions for this domain, its architecture strongly suggests a possible role in molecular interaction. Meanwhile, this construct provides a very interesting model to test several approaches to describe the behavior of a disordered tail attached to a rigid frame. This knowledge gained on this system will be useful to undertake dynamics-function studies on systems where the biological role of a disordered region is more established, such as for the AB domain of nuclear receptors.

Chemically, zinc is considered an uninteresting element with its complete d-shell and single oxidation state [7], but these features render it a powerful element biochemically [8]. In enzymes, its highly concentrated charge, Lewis acidity and lack of ligand field

stabilization energy are exploited to catalyze chemical reactions. In other proteins, zinc serves to stabilize folds of protein domains involved in inter-molecular interactions. In biochemical terms, the bland chemistry of zinc is an advantage, with variation in function being achieved in these non-enzymatic proteins through the variety of polypeptide sequences that can achieve the twin requirements of selective zinc binding and folding to define a given function. Here, we have seen how a single amino acid change can dramatically alter function through a subtle change in properties, and how differences in function can be achieved for a common zinc-binding scaffold by different architecture of secondary structure elements. It remains unclear to what extent the occurrence of exchange of zinc from zinc-binding proteins is relevant to function: what fractions of zinc finger domains are zinc-bound in the cytoplasm? and in the nucleus? how does local pH affect these populations? are relative stabilities of different zinc finger domains important in defining an order in the competition for zinc ions in vivo? is cadmium binding to zinc finger proteins relevant to cadmium toxicity or carcinogenesis? I hope that the work in this thesis may form the basis for further experiments to address these important questions.

### References

1. Chatterjee S, Kundu S, Sengupta S, Bhattacharyya A: **Divergence to apoptosis from ROS induced cell cycle arrest: effect of cadmium.** *Mutat Res* 2009, **663**(1-2):22-31.
2. Valko M, Morris H, Cronin MT: **Metals, toxicity and oxidative stress.** *Curr Med Chem* 2005, **12**(10):1161-1208.
3. Hsu MH, Palmer CN, Song W, Griffin KJ, Johnson EF: **A carboxyl-terminal extension of the zinc finger domain contributes to the specificity and polarity of peroxisome proliferator-activated receptor DNA binding.** *J Biol Chem* 1998, **273**(43):27988-27997.
4. Hsieh JC, Whitfield GK, Oza AK, Dang HT, Price JN, Galligan MA, Jurutka PW, Thompson PD, Haussler CA, Haussler MR: **Characterization of unique DNA-binding and transcriptional-activation functions in the carboxyl-terminal extension of the zinc finger region in the human vitamin D receptor.** *Biochemistry* 1999, **38**(49):16347-16358.
5. Kalodimos CG, Biris N, Bonvin AM, Levandoski MM, Guennuegues M, Boelens R, Kaptein R: **Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes.** *Science* 2004, **305**(5682):386-389.
6. Shaffer PL, Jivan A, Dollins DE, Claessens F, Gewirth DT: **Structural basis of androgen receptor binding to selective androgen response elements.** *Proc Natl Acad Sci U S A* 2004, **101**(14):4758-4763.

7. Cotton FA, Wilkinson,G.: **Advanced inorganic chemistry**. 1980.
8. Frausto da Silva JJR, Williams,R.J.P.: **The biological chemistry of the elements: The inorganic chemistry of life**. 1991.





## Appendix I. NMR spectroscopy and data analysis

### 1. Spectral acquisition for backbone assignments

NMR experiments for AR-DBD wild type, T575A mutant, ATXN7 and ATXN7 L3 were as follows:

Protein	AR-DBD WT/T575A mutant	ATXN7/ATXN7 L3
Acquisition temperature	303K	298K
Sample	$^1\text{H}$ , $^{15}\text{N}$ - and $^{13}\text{C}$ -, $^{15}\text{N}$ -labeled proteins	As for AR-DBD
Spectra for backbone assignments	HN(CO)CA, HNCA, HN(CO)CACB, HNCACB and HNCO (Acquired on a Bruker DRX 600 MHz NMR spectrometer).	As for AR-DBD
Spectra for side-chain assignments	1. 2D $^1\text{H}$ - $^1\text{H}$ NOESY with a mixing time of 200 ms (Acquired on a Bruker Avance 800 MHz NMR spectrometer). 2. 3D $^{15}\text{N}$ -edited NOESY with a mixing time of 150 ms (Acquired on a Bruker DRX 600 MHz NMR spectrometer).	1. HCCH-TOCSY (Acquired on a Bruker DRX 600 MHz NMR spectrometer). 2. 2D $^1\text{H}$ - $^1\text{H}$ NOESY with a mixing time of 200 ms (Acquired on a Bruker Biospin 950 MHz NMR spectrometer).

## 2. Structural Calculation

### 2.1 Initial calculation using ATNOS-CANDID

Three dimensional structures were calculated using ATNOS-CANDID software.[2] ATNOS is a novel software which uses automated NOESY peak picking and NOE signal identification in homonuclear 2D and heteronuclear-resolved 3D spectra during structure determination by NMR. It combines with CANDID and CYANA or X-PLOR to allow automated protein structure determination in iterative cycles. The input data for ATNOS consists of the amino acid sequence, the chemical shift tables, dihedral angle constraints, zinc coordination constraints and 2D homonuclear or 3D heteronuclear NOESY spectra. The advantage of ATNOS-CANDID is that it adapts the chemical shift values to the individual NOESY spectra. This process can account for the deviations in calibration of the acquired spectra. Typically, seven cycles are run during ATNOS-CANDID calculations. In the first cycle, ATNOS only select peaks picked automatically from the spectra and validated according to the chemical shift tables. All information acquired from the first cycle is transferred to the second cycle.

An intermediate three dimensional structure is built by combining the peaks selected by ATNOS and NOE assignments validated by CANDID. The precision and accuracy of intermediate structures is improved from cycle to cycle. The final structures are obtained from cycle 7.

### 2.1.1 Input data for AR-DBD wild type and T575A mutant

The input data for AR-DBD wild type, T575A mutant, ATXN7 and ATXN7 L3 lists as follows:

Protein structure	AR-DBD	ATXN7/ATXN7 L3
Prot file	only <sup>1</sup> H chemical shift	
Spectra	2D <sup>1</sup> H- <sup>1</sup> H NOESY spectra with a mixing time of 200ms (acquired at 800 MHz)	For <b>ATXN7</b> : 2D <sup>1</sup> H- <sup>1</sup> H NOESY spectra with a mixing time of 200ms (acquired at 950 MHz) 3D <sup>15</sup> N-edoted NOESY with a mixing time of 150ms (acquired at 600MHz) For <b>ATXN7 L3</b> : 2D <sup>1</sup> H- <sup>1</sup> H NOESY spectra with a mixing time of 200ms (acquired at 950 MHz)
Sequence file	included	included
Zinc coordination constraints	Described in Table 1	Described in Table 2
H-bonds	Described in Table 3	not included
Patch for protonation state of side-chain in histidine	For <b>WT</b> : patch HISD reference=NIL=( resid 20 ) For <b>T575A</b> mutant no patch	For <b>ATXN7</b> : patch HISE reference=NIL=( resid 17 ) patch HISE reference=NIL=( resid 38 ) For <b>ATXN7 L3</b> : patch HISE reference=NIL=( resid -1 ) patch HISE reference=NIL=( resid 19 ) patch HISE reference=NIL=( resid 33 )

Table 1 X-PLOR input file of zinc coordination constraints for AR-DBD WT/T575A mutant

assign ( segid "zn" and resid 1 and name zn ) ( resid 9 and name sg ) 2.4 0.2 0.2
assign ( segid "zn" and resid 1 and name zn ) ( resid 12 and name sg ) 2.4 0.2 0.2
assign ( segid "zn" and resid 1 and name zn ) ( resid 26 and name sg ) 2.4 0.2 0.2
assign ( segid "zn" and resid 1 and name zn ) ( resid 29 and name sg ) 2.4 0.2 0.2

Table 2 X-PLOR input file of zinc coordination constraints for ATXN7/ATXN7 L3

---

```

assign ( segid "zn" and resid 1 and name zn ) ( resid 18 and name sg ) 2.4
0.2 0.2
assign ( segid "zn" and resid 1 and name zn ) ( resid 29 and name sg ) 2.4
0.2 0.2
assign ( segid "zn" and resid 1 and name zn ) ( resid 35 and name sg ) 2.4
0.2 0.2
assign ( segid "zn" and resid 1 and name zn ) ( resid 38 and name nd1 ) 2.2
0.2 0.2

```

---

Table 3 Input file of H-bonds for AR-DBD WT/T575A mutant

---

```

assign ( resid 9 and name HN ) ( resid 14 and name O ) 1.80 0.00 0.50
assign ( resid 9 and name N ) ( resid 14 and name O ) 2.80 0.00 0.50
assign ( resid 10 and name HN ) ( resid 24 and name O ) 1.80 0.00 0.50
assign ( resid 10 and name N ) ( resid 24 and name O ) 2.80 0.00 0.50
assign ( resid 16 and name HN ) ( resid 7 and name O ) 1.80 0.00 0.50
assign ( resid 16 and name N ) ( resid 7 and name O ) 2.80 0.00 0.50
assign ( resid 25 and name HN ) ( resid 18 and name O ) 1.80 0.00 0.50
assign ( resid 25 and name N ) ( resid 18 and name O ) 2.80 0.00 0.50
assign ( resid 30 and name HN ) ( resid 26 and name O ) 1.80 0.00 0.50
assign ( resid 30 and name N ) ( resid 26 and name O ) 2.80 0.00 0.50
assign ( resid 31 and name HN ) ( resid 27 and name O ) 1.80 0.00 0.50
assign ( resid 31 and name N ) ( resid 27 and name O ) 2.80 0.00 0.50
assign ( resid 32 and name HN ) ( resid 28 and name O ) 1.80 0.00 0.50
assign ( resid 32 and name N ) ( resid 28 and name O ) 2.80 0.00 0.50
assign ( resid 33 and name HN ) ( resid 29 and name O ) 1.80 0.00 0.50
assign ( resid 33 and name N ) ( resid 29 and name O ) 2.80 0.00 0.50
assign ( resid 34 and name HN ) ( resid 30 and name O ) 1.80 0.00 0.50
assign ( resid 34 and name N ) ( resid 30 and name O ) 2.80 0.00 0.50
assign ( resid 35 and name HN ) ( resid 31 and name O ) 1.80 0.00 0.50
assign ( resid 35 and name N ) ( resid 31 and name O ) 2.80 0.00 0.50
assign ( resid 36 and name HN ) ( resid 32 and name O ) 1.80 0.00 0.50
assign ( resid 36 and name N ) ( resid 32 and name O ) 2.80 0.00 0.50
assign ( resid 65 and name HN ) ( resid 62 and name O ) 1.80 0.00 0.50
assign ( resid 65 and name N ) ( resid 62 and name O ) 2.80 0.00 0.50
assign ( resid 66 and name HN ) ( resid 62 and name O ) 1.80 0.00 0.50
assign ( resid 66 and name N ) ( resid 62 and name O ) 2.80 0.00 0.50
assign ( resid 67 and name HN ) ( resid 63 and name O ) 1.80 0.00 0.50
assign ( resid 67 and name N ) ( resid 63 and name O ) 2.80 0.00 0.50
assign ( resid 68 and name HN ) ( resid 64 and name O ) 1.80 0.00 0.50
assign ( resid 68 and name N ) ( resid 64 and name O ) 2.80 0.00 0.50
assign ( resid 69 and name HN ) ( resid 65 and name O ) 1.80 0.00 0.50
assign ( resid 69 and name N ) ( resid 65 and name O ) 2.80 0.00 0.50
assign ( resid 70 and name HN ) ( resid 66 and name O ) 1.80 0.00 0.50
assign ( resid 70 and name N ) ( resid 66 and name O ) 2.80 0.00 0.50
assign ( resid 71 and name HN ) ( resid 67 and name O ) 1.80 0.00 0.50
assign ( resid 71 and name N ) ( resid 67 and name O ) 2.80 0.00 0.50
assign ( resid 72 and name HN ) ((resid 68 and name O) or (resid 69 and name O))
1.80 0.00 0.50
assign ( resid 72 and name N ) ((resid 68 and name O) or (resid 69 and name O))
2.80 0.00 0.50
assign ( resid 73 and name HN ) ((resid 69 and name O) or (resid 70 and name O))
1.80 0.00 0.50
assign ( resid 73 and name N ) ((resid 69 and name O) or (resid 70 and name O))
2.80 0.00 0.50
assign ( resid 74 and name HN ) ( resid 69 and name O ) 1.80 0.00 0.50
assign ( resid 74 and name N ) ( resid 69 and name O ) 2.80 0.00 0.50

```

---

### 2.1.2 Generate the input data using X-PLOR scripts

The backbone and side chain assignments and amino acid sequence were generated in CARA [3] and dihedral angle restraints were obtained from TALOS.[4] The protein structure files (PSF) were generated by a standard X-PLOR script, “generate.inp”. In this script, the protonation states of histidine side chain are set. In AR-DBD wild type, the unique histidine is doubly protonated while the T575A mutant is protonated on the N $\epsilon$  atom. For ATXN7 and ATXN7 L3, all the histidines are protonated on the N $\epsilon$  atom. In addition, residues involved in the zinc coordination need to be modified (“patched”) in the script. Each cycle of the ATNOS calculation produces peak lists which are picked and assigned by CANDID. The “Cycle\_AtnosCandid.out” provides a detailed description of the run. The “Cycle\_AtnosCandid.upl” file contains the NOE constraints produced by CANDID. Although cycle7 is the final ATNOS calculation, it is essential to refine the structure further using X-PLOR-NIH.[5]

## 2.2 Structure refinement using X-PLOR

### 2.2.1 X-PLOR version transfer

The three-dimensional structure obtained from cycle7 of ATNOS/CANDID requires refinement.[2, 5] This was achieved here using the strategy described below:

In the first step, the structures resulting from cycle7 of the ATNOS/CANDID calculations were refined using the same parameter set, parallhdg.pro. This set of structure files were then converted for refinement using the parallhdg5.3.pro parameter set that provides a better geometric description of polypeptides, in particular for side-chain rotamers. In this approach, the distance restraints in the file Cycle7\_AtnosCandid.upl were treated with a softer potential than the distance restraints imposed to ensure zinc coordination.

### 2.2.2 Stereo configuration refinement

Since pairs of resonances of stereochemically related nuclei, such as the  $^1\text{H}$  nuclei and  $\text{CH}_2$  groups, were not stereospecifically assigned, they were allowed to “float” in all calculations up to this point. The set of structures were therefore analyzed using XPLOR scripts to examine the configuration of such pairs and establish whether the experimental data permitted stereoassignment of the resonances. The restraint file was update manually to account for such stereoassignments or where they remained ambiguous. With “floating” assignment no longer allowed, the structures were refined again.

### 2.3 Water refinement

The set of structures were then copied for water refinement. This required particular attention to the zinc coordination sphere. Simple distance restraints are unsuitable and a more explicit definition of the coordination environment is required. The protocol was used as provided with no further modification. Water refinement leads to an improved Ramachandran plot.

The 64 structures refined after modification of stereoassignments, were copied to the water refinement dataset. When generating the new protein structure file containing the covalent geometry definitions and atom information, it is important to “patch” the zinc coordinating atoms correctly. The water refinement was initiated using PYTHON programming interface. The final 20 lowest-energy structures were selected after water refinement. The agreement of backbone dihedral angles for each residue in Ramachandran plot was checked by the PROCHECK program [6].



The logo of the University of Strasbourg, featuring two blue curved segments that form a stylized 'S' shape, positioned behind the text.

UNIVERSITÉ DE STRASBOURG

## Appendix II: DNA interaction of AR-DBD wild type and T575A mutant

### 1. Introduction

Steroid hormones induce their effect on the expression of target genes by activating their cognate steroid receptor. The androgen receptor (AR) is member of the super-family of nuclear receptors (NRs) (Table 1). The architecture of AR contains two highly conserved domains, the ligand binding domain and the DNA binding domain (Fig.1). The ligand-binding domain (LBD) mediates ligand binding and dimerization and contains the ligand-dependent transactivation function.[19] In addition, gene expression mediated by nuclear receptor can **also regulate through** binding to DNA response elements associated with target genes. AR recognizes and binds to the androgen response element (ARE) site through two zinc fingers located in the DNA-binding domain. The half-site of ARE can be arranged as either inverted repeats or direct repeats. Previous studies showed that the nuclear receptors except for the estrogen receptor recognize a partial palindromic repeat of 5'-TGTTCT-3' as a homodimer.[20] The ARE in the probasin promoter (PB) recognizes AR specifically and residues in the C-terminal extension of AR-DBD bind to PB-ARE-2 with high affinity.[21] However, the response elements within the *sc* (human secretory component) and *slp* (mouse sex-limited protein) enhancers display specificity for the androgen receptor with low affinity.[22] These significant differences allow us to study molecular mechanism of DNA interaction of AR using DNA as a probe. In our study, we started the DNA titration of AR-DBD using a half-site GRE (TAT-GRE II) (glucocorticoid receptor response element) which is a non-specific response element for GR (glucocorticoid receptor)[22] and compared the binding affinity between WT and T575A mutant domains.



	Short name	type	Full name	Ligand
Class I	TR	$\alpha, \beta$	Throid hormone receptor	Throid
	RAR	$\alpha, \beta, \gamma$	Retinoic acid receptor	Retinoic acid
	VDR		Vitamin D <sub>3</sub> receptor	VitaminD Dihydroxy-vitamin D <sub>3</sub>
	PPAR	$\alpha, \beta, \gamma$	Peroxisome proliferators-activator receptor	Prostaglandine
Class II	RXR <sub>s</sub>	$\alpha, \beta, \gamma$	Retinoid X receptor	cis Retinoic acid
	TR2	$\alpha, \beta$		
Class III	GR		Glucocorticoid receptor	Glucocorticoid
	AR		Androgen receptor	Androgen
	PR		Progesterone receptor	Progesterone
	ER	$\alpha, \beta$	Estrogen receptor	Estrogen
	MR		Mineralocorticoid receptor	Mineralocorticoid
Class IV	NGFI-B	$\alpha, \beta, \gamma$	Nerve growth factor induced clone B	
Class V	SF-1/FTZ-F1	$\alpha, \beta$	Steroidogenic factor-1	Oxysteroid
Class VI	GCNF		Germ cell nuclear factor	
Class O	SHP		Small heterodimer partner	
	DAX-1		Dosage-sensitive sex	

Table 1 Classification of nuclear receptors.

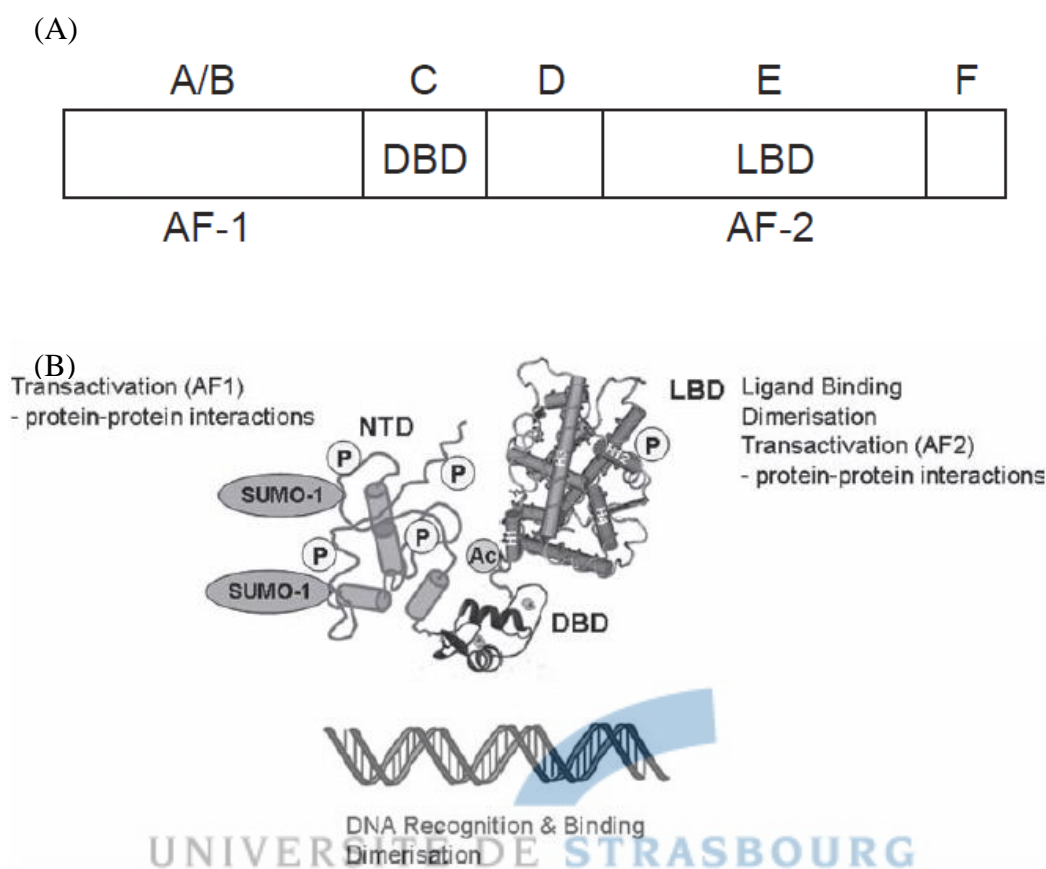


Fig.1 (A) Schematic representation of domains of androgen receptor. AF: activation function; DBD: DNA-binding domain; LBD: ligand-binding domain. (B) Structure of nuclear receptor complex. The N-terminal domain is a ligand-dependent domain that regulates transactivation. The DBD is located in the central region involved in DNA interactions. LBD has ligand binding activity and contains AF-2 domain which mediate protein-protein interaction. Between DBD and LBD is the hinge region which acts as an adaptor to connect DBD and LBD.

## 2. Methods

A sample of  $^{15}\text{N}$ -labeled AR-DBD was titrated by addition of the half-site of TAT-GRE II (sequence: 5'-GCTGTACAGC-3') at ratios of 4:1, 2:1, 1:1 and 1:2. A series of 1D spectrum of GRE and  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra was recorded at 303K. The cross-peaks in  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra were picked and chemical shift differences against each residue were plotted.

## 3. Results

Fig.2 shows 1D spectra of the half-site of TAT-GRE II acquired at 283K, 293K and 303K. Signals of four imino groups are located at 12~14 ppm.

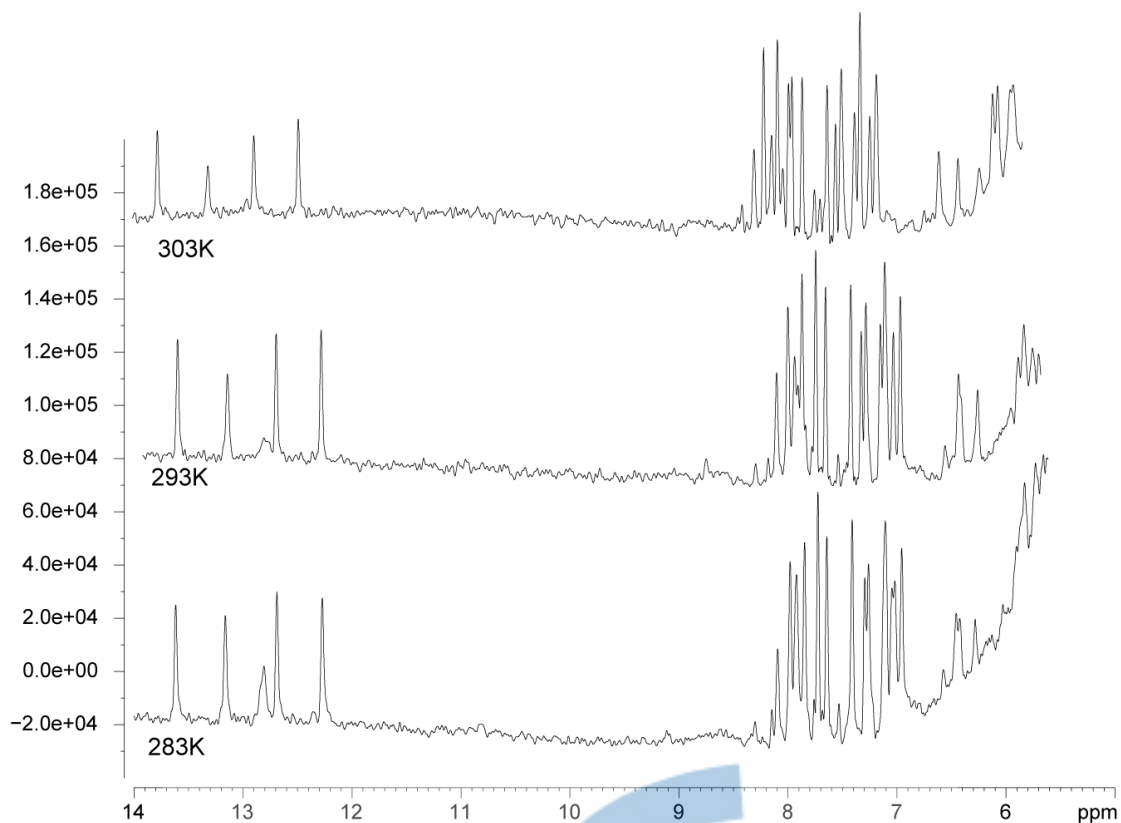


Fig.2 1D spectra of half-site TAT-GRE II recorded at 283K, 293K and 303K.

Fig.3 shows the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of AR-DBD WT upon titration with the half-site of TAT-GRE II. In both WT and T575A mutant domains, residue C579 shows the largest shift in the spectrum and several residues in close proximity to the first zinc binding site including I561, H570 and G577 have large chemical shift changes, indicating involvement in DNA interaction. (Fig.4) Patterns of chemical shift changes between WT and T575A mutant domains are similar except for F606 that reveals larger chemical shift changes in WT than T575A mutant. These results differ from the binding pocket of AR-DBD bound to ADR3 [23] although the structure of AR-DBD bound to the half-site of TAT-GRE II has not yet been determined.

#### 4. Conclusions and perspectives

We investigated the DNA interaction of AR-DBD using TAT-GRE II as a probe. The results showed residues in proximity of the first zinc binding site may be involved in DNA interaction. However, this result does not provide sufficient information for DNA recognition by AR and structure of AR-DBD bound to TAT-GRE II needs to be determined.

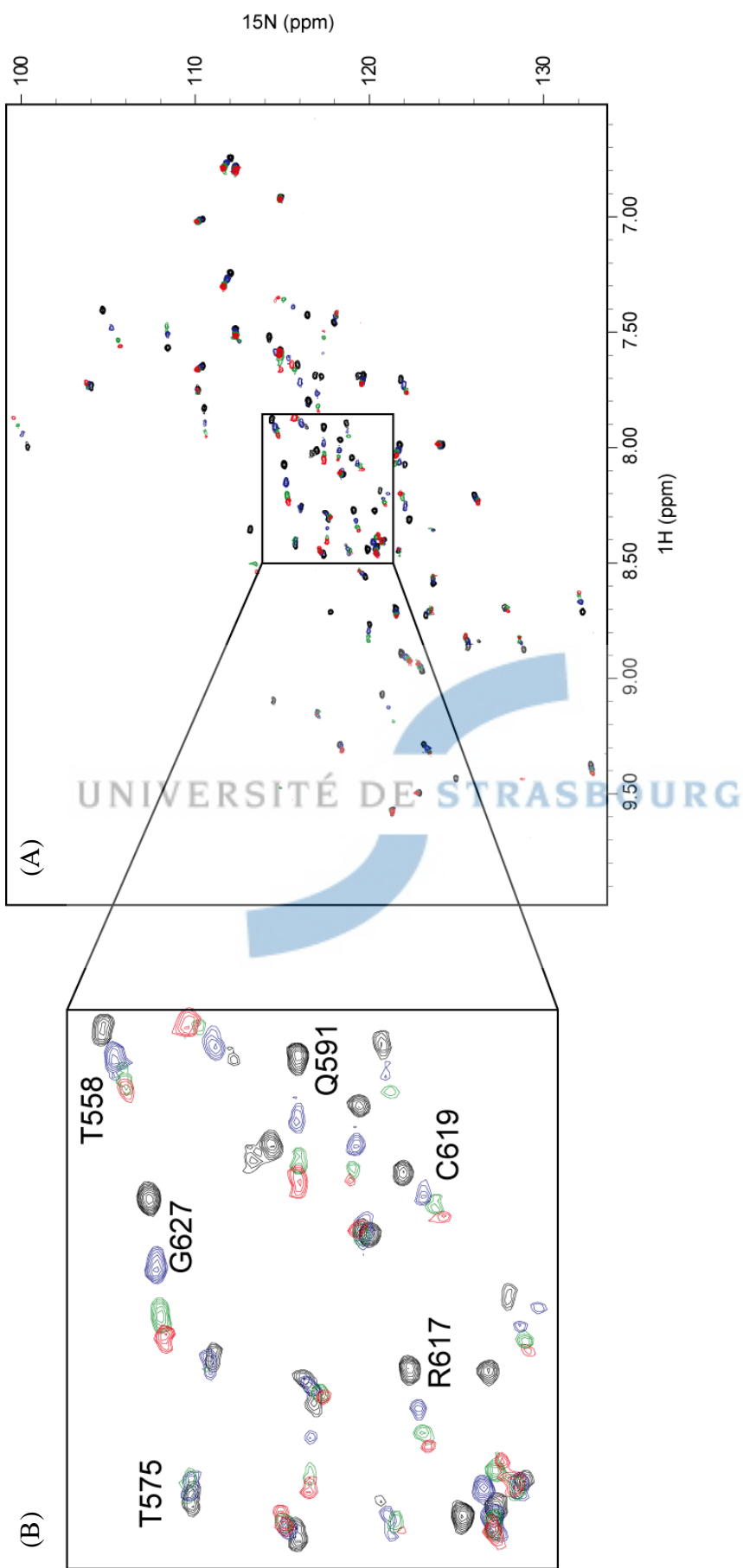


Fig.3 (A) Superimposed  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of AR-DBD WT bound to the half-site of TAT-GRE II. (B) The central region of the spectrum shown in A. AR-DBD WT was titrated with half-site of TAT-GRE II at the ratios of 2:1 (blue), 1:1 (green), 1:2 (red).

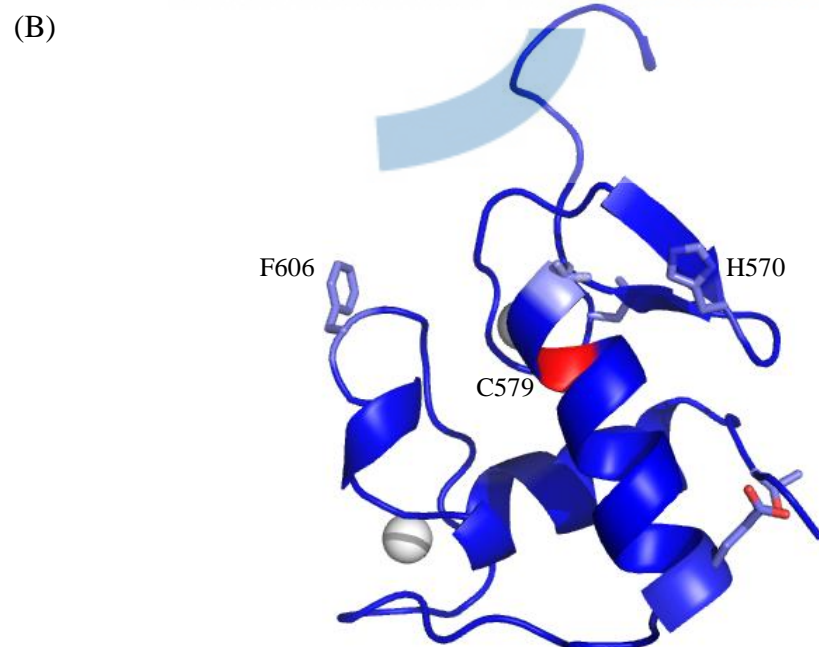
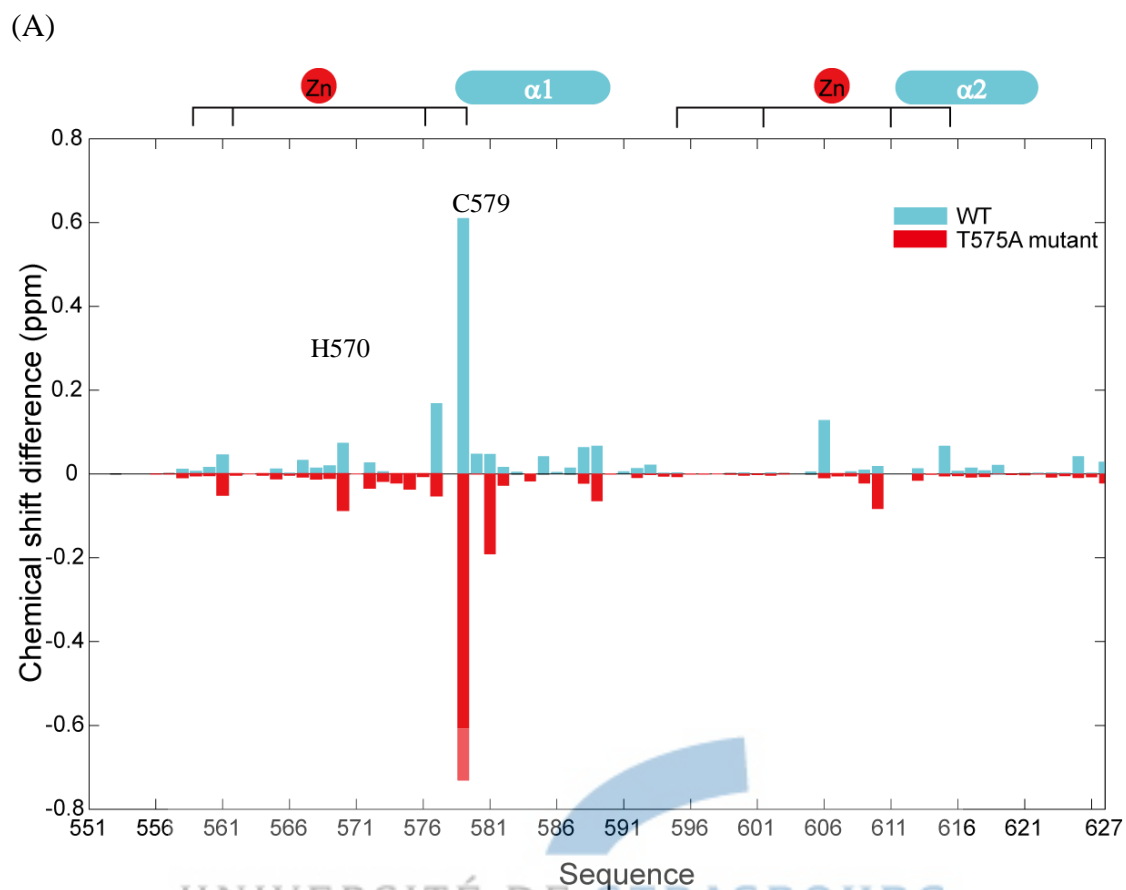


Fig.4 (A) The chemical shift difference of AR-DBD WT and T575A mutant upon fully titration. (B) Ribbon representation of AR-DBD WT. Several residues in close proximity to the first zinc binding site including L560, H570, G577 and C579 show larger chemical shift changes in WT upon binding to DNA.

## Appendix III: Investigation of the dynamic behavior of the C-terminal extension region (CTE) of AR-DBD

### 1. Introduction

The DNA binding domain (DBD) of nuclear receptors plays an important role in DNA interaction and thus promotes transcriptional activity. The C-terminal extension (CTE) region in close proximity to the DBD was previously shown to enhance DNA binding activity.[24] For AR, previous studies showed that AR-DBD together with its CTE binds to PB-ARE2, an AR-specific response element with high affinity.[25] CTE is a 30 amino acid segment that is not conserved across the class of nuclear receptors (Fig.1).[26] The CTE in nuclear receptors was thought to be an unstructured segment but recent studies have shown that part of the C-terminal extension of ecdysone receptor formed an  $\alpha$ -helix and contacts the minor groove of the DNA binding surface.[27] Hence, investigation of the role of the CTE in androgen receptor could prove to be of interest and could help to explain the mechanisms of DNA interaction of this receptor.

The logo of the University of Strasbourg, featuring two blue curved segments forming a stylized 'S' shape.

UNIVERSITÉ DE STRASBOURG

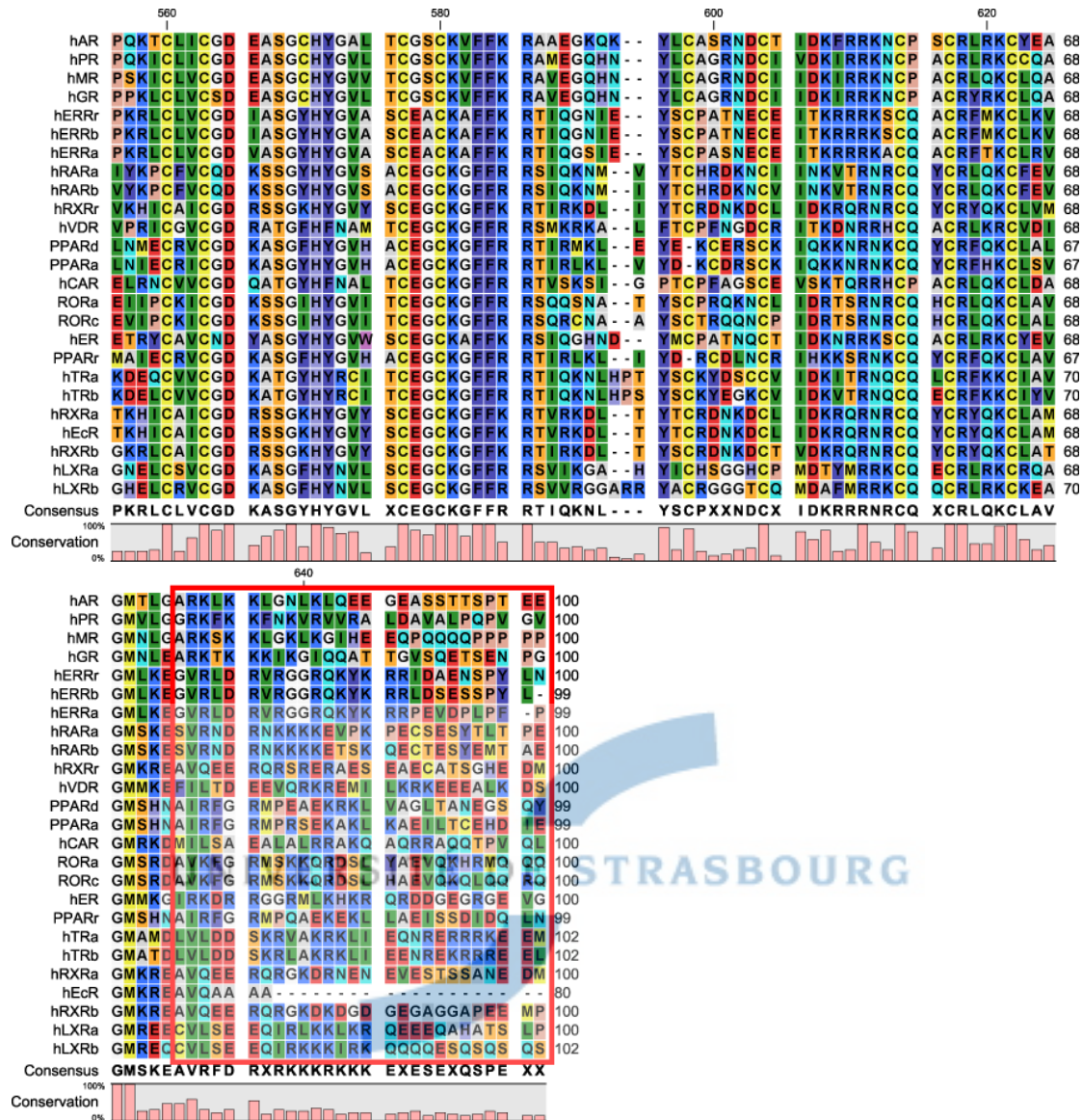


Fig.1 Sequence alignments of the C-terminal extension (the red box region) in human nuclear receptors.

## 2. Results and discussion

### 2.1 Comparison of $^1\text{H}$ - $^{15}\text{N}$ HSQC spectra of AR-DBD and AR-DBD-CTE

To initiate this study, a sample of  $^{15}\text{N}$ -labeled AR-DBD-CTE wild type was prepared and a  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectrum was recorded first to check the protein fold. Fig.2 shows the superposition of  $^1\text{H}$ - $^{15}\text{N}$ -HSQC of AR-DBD and AR-DBD-CTE acquired at 303K. It can be seen that the cross-peaks in the spectrum of AR-DBD-CTE are well-dispersed, indicating a folded protein. Additional peaks were observed in the spectrum of AR-DBD-CTE, reflecting the presence of the additional H-N groups of each residue in the C-terminal extension region. The patterns of the cross-peaks, with the exception of these additional peaks, are similar to those in AR-DBD, indicating a structural similarity in the core region between AR-DBD and AR-DBD-CTE and thus facilitating



further structure analysis.

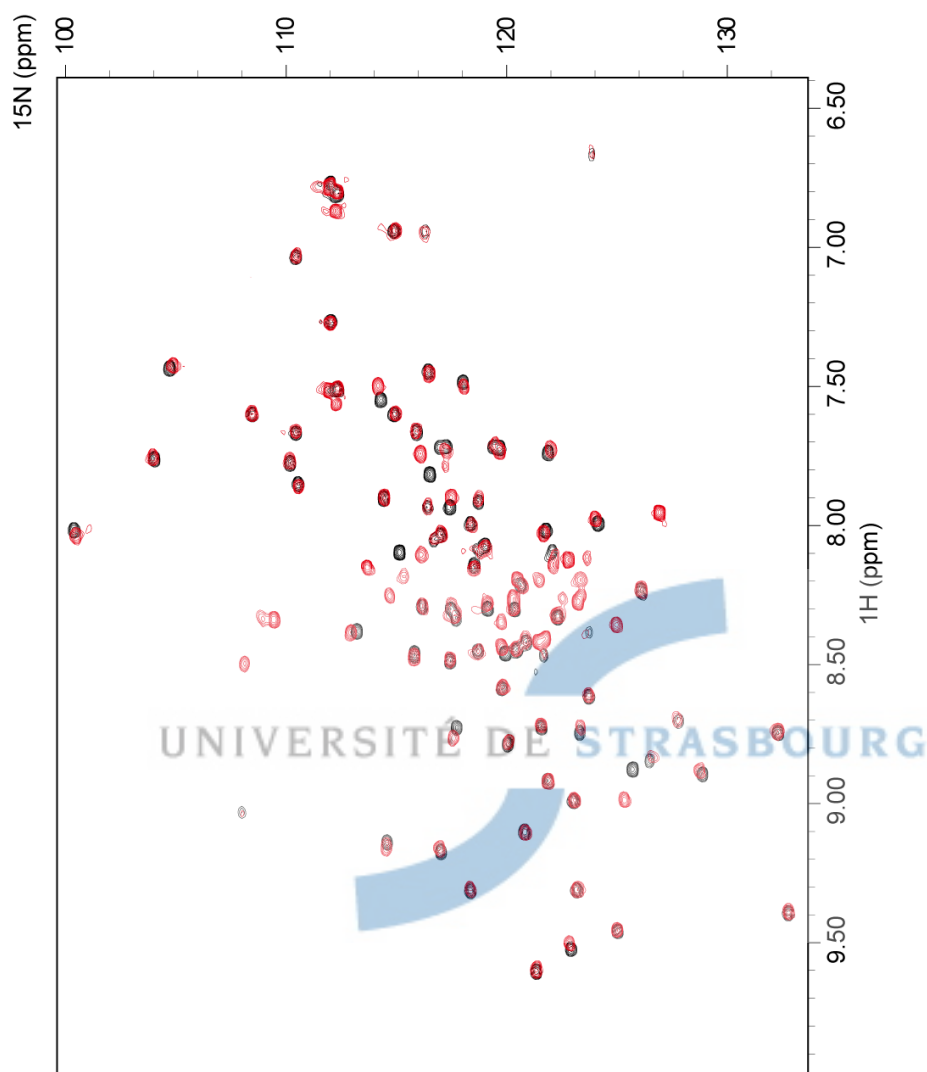


Fig.2 Superposition of  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of AR-DBD (black) and AR-DBD-CTE (red).

## 2.2 Investigation of the dynamic behavior in AR-DBD-CTE

We started to probe the dynamic behavior of CTE by recording  $^1\text{H}$ - $^{15}\text{N}$  NOEs. Fig.3 shows the profile of  $^1\text{H}$ - $^{15}\text{N}$  NOEs of AR-DBD-CTE. Although the backbone resonances for the CTE were not assigned in a residue-specific manner, the reduced  $^1\text{H}$ - $^{15}\text{N}$  NOEs of CTE in the shaded region of Fig.2 reveal fast motions on the ps-ns time-scale. Comparing the patterns in the core region between AR-DBD and AR-DBD-CTE, high mobility was observed for some residues of AR-DBD-CTE, suggesting higher structural flexibility of the core region of AR-DBD-CTE. This difference in dynamic behavior may indicate distinct conformations of AR-DBD and

AR-DBD-CTE when binding to DNA.

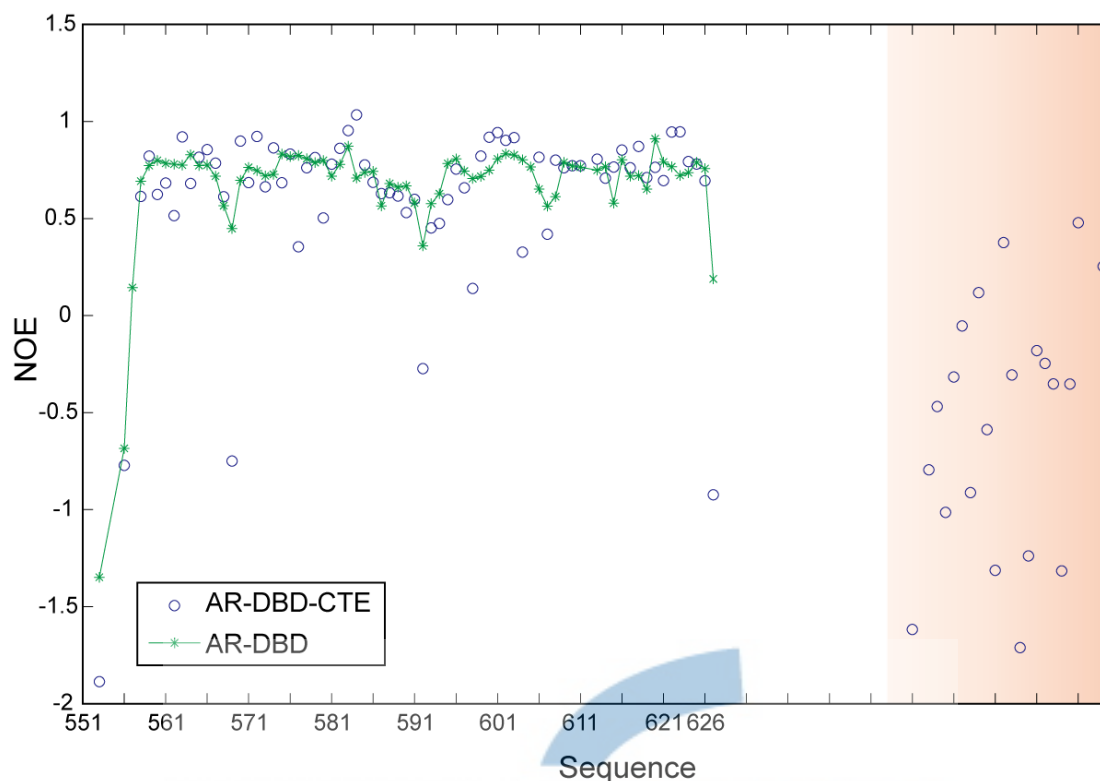


Fig.3  $^1\text{H}$ - $^{15}\text{N}$  NOEs of AR-DBD and AR-DBD-CTE. Values for residues of the CTE are shown in the shaded region.

### 2.3 Conclusions and perspectives

Our initial studies of AR-DBD-CTE provide useful information for further investigations such as structure determination, dynamics measurement and DNA interactions. The CTE in AR-DBD-CTE showed ps-ns time-scale motions and the core region was also observed to exhibit higher mobility for several residues. These differences in dynamic behavior between AR-DBD and AR-DBD-CTE can provide clues about the mechanism of DNA recognition. In addition, high mobility of CTE may allow it to contact DNA binding surface.

## Appendix IV: Structural role of ATXN7 in histone 2A·2B

### heterodimer binding

It has been shown that ATXN7 binds to histone 2A·2B (H2A·H2B) in GST-pull down experiments. To investigate the structural role of ATXN7 binding to H2A·H2B, a series of titration experiments was recorded using NMR. The sample of  $^{15}\text{N}$ -labeled ATXN7 was mixed with H2A·H2B at ratios of 4:1 and 1:1. Fig.1 represents the 1D spectrum of ATXN7 with H2A·H2B. Upon mixing ATXN7 with H2A·H2B in equimolar amounts, a broad peak appears beside a H2A·H2B resonance, indicating ATXN7 binding to H2A·H2B. Fig.2 represents the superimposition of HSQC spectra for ATXN7 with H2A·H2B. It is clear that most peaks are unaffected. This may be because the correlation time of H2A·H2B is much slower than that of ATXN7. In order to detect the precise signal, it is possible to measure the magnetization transfer from H2A·H2B to ATXN7. The signal of the methyl group in Fig.1 can be used a probe to detect the phenomenon of magnetization transfer. The methyl group of H2A·H2B was irradiated using a selective pulse. Due to the slow correlation time of H2A·H2B, magnetization is able to diffuse within the histone first and then within the complex. The transfer of magnetization in the complex can be detected in the free state of ATXN7 through changes of intensities in HSQC spectra. Fig.3 gives the ratios of intensities for free ATXN7 after irradiating the signal of H2A·H2B or of the complex. It is clear that the N- and C-terminal region are perturbed, suggesting involvement in histone interaction. According to the dynamic behavior of ATXN7, the N-terminal tail exhibits fast motions on the ps-ns time scale, showing that this part is flexible and unstructured free ATXN7 and thus is able to interact with histone.

### Methods

2 different samples were prepared, one containing 210  $\mu\text{l}$  of ATXN7 ( $\sim 60 \mu\text{M}$ ) and 40  $\mu\text{l}$  of histone (1ng/ $\mu\text{l}$ ) (ATXN7: histone=4:1) the other contains 60  $\mu\text{l}$  of ATXN7 and 160  $\mu\text{l}$  of histone (ATXN7: histone=1:1). The buffer used is 50mM phosphate pH 7.1, 200mM NaCl and 1mM DTT. 1D spectra of histone only and the complex were acquired and a series of  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra were recorded with these two samples.

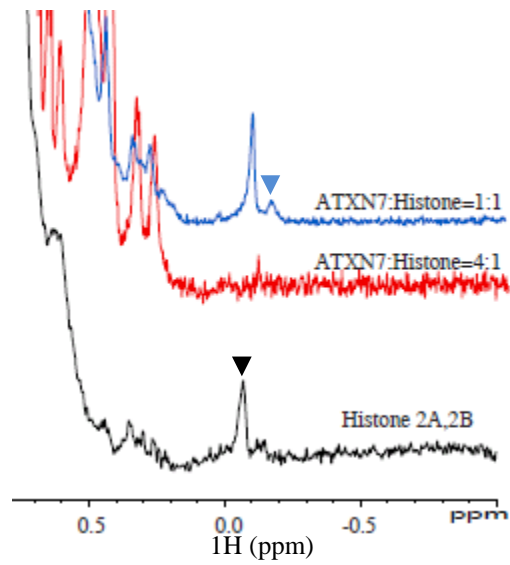


Fig.1 1D spectrum of ATXN7 with H2A ·H2B dimer. The ratio of ATXN7 to H2A ·H2B dimer are 4: 1 and 1: 1. The black triangle indicates the methyl group of H2A ·H2B while the blue triangle represents the signal that appears upon ATXN7 binding to H2A ·H2B.



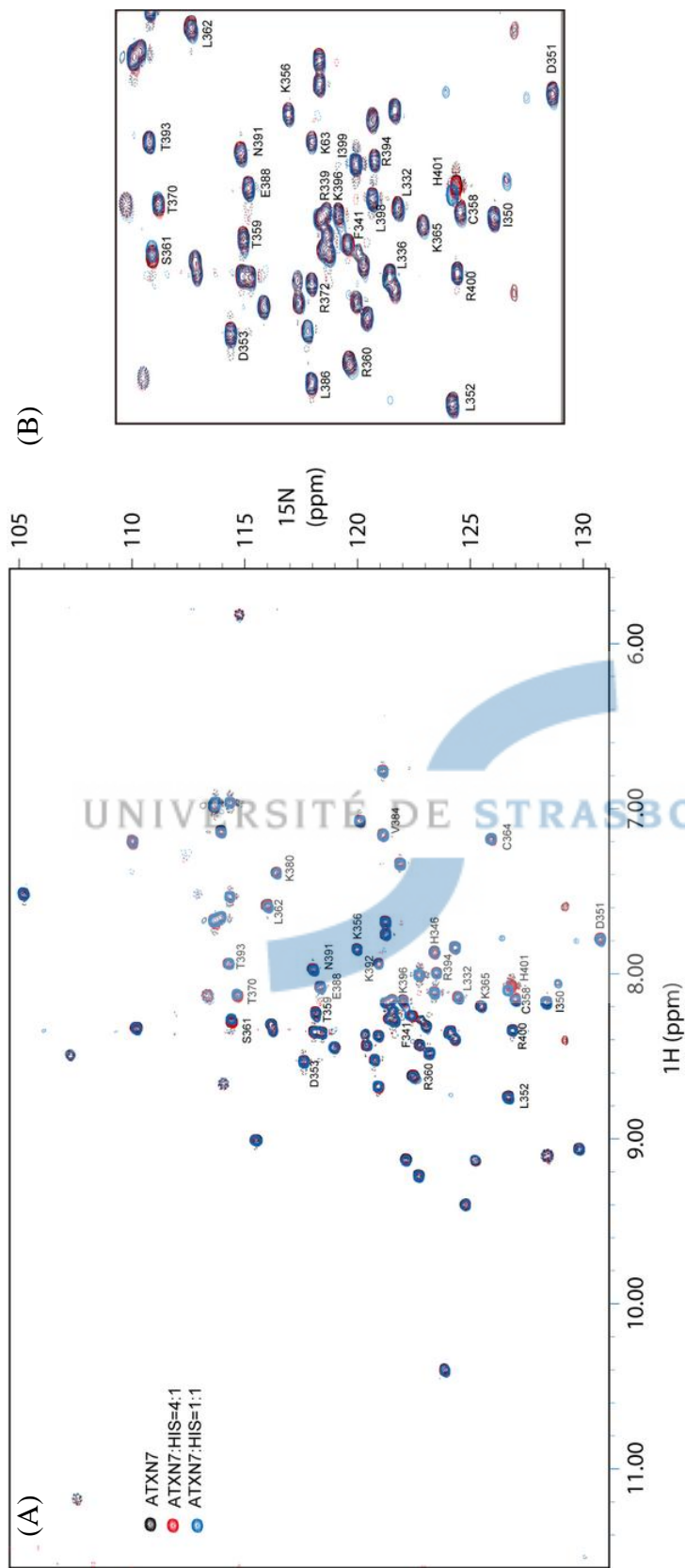


Fig.2 (A) Superimposed of  $^1\text{H}$ - $^{15}\text{N}$  HSQC of ATXN7 bound to histone H2A H2B dimer. (B) The central region of the spectrum shown in A.

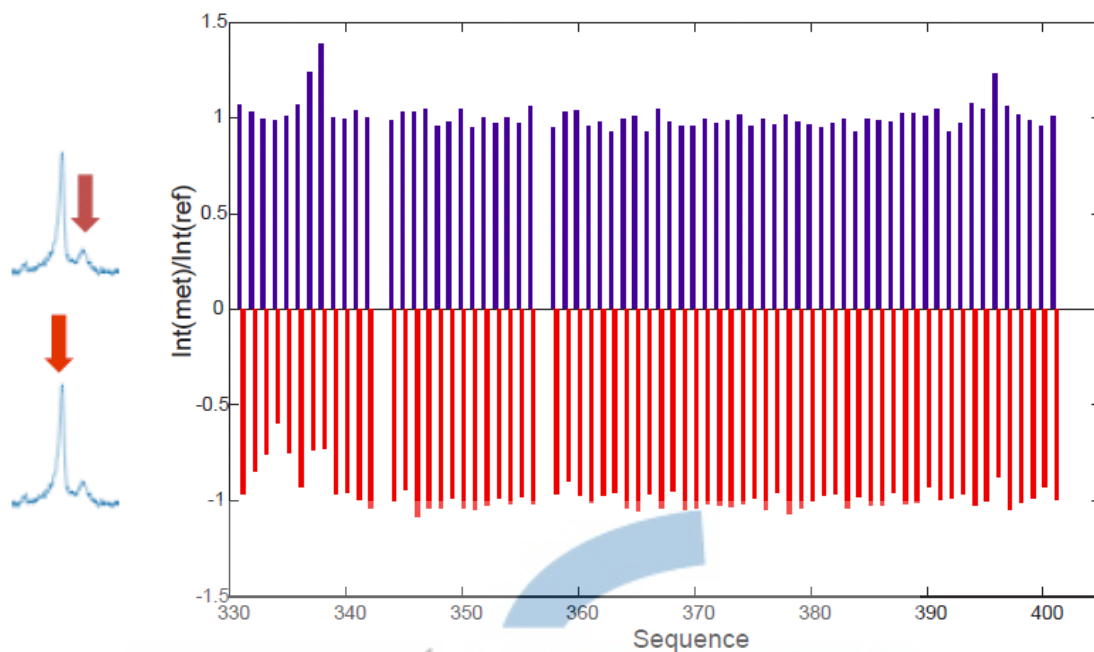


Fig.3 Scheme of intensity ratios of ATXN7 binding to H2A-H2B dimer upon irradiation of signals (shown on the left) at the frequency of free H2A-H2B dimer and complex. Irradiation at different frequencies causes different perturbing effects on signals. Detection of the change of intensity in the free state of ATXN7 allows identification of the regions involved in histone interaction.

## Appendix V: ATXN7 lacks DNA binding activity

A GST-pull down assay showed that ATXN7 lacks DNA binding activity. To confirm this result,  $^{15}\text{N}$ -labeled ATXN7 was titrated with DNA oligonucleotide and a series of the HSQC spectra with various dosages was recorded. Fig.1 shows the superimposition of HSQC spectra of ATXN7 with different ratios of DNA oligonucleotides. All the spectra reflect the folded conformation of ATXN7, however, no peak shifts were observed in the spectra, indicating no perturbation by DNA. DNA binding activity of ATXN7 could be neither observed in GST-pull down assay nor NMR spectroscopy.

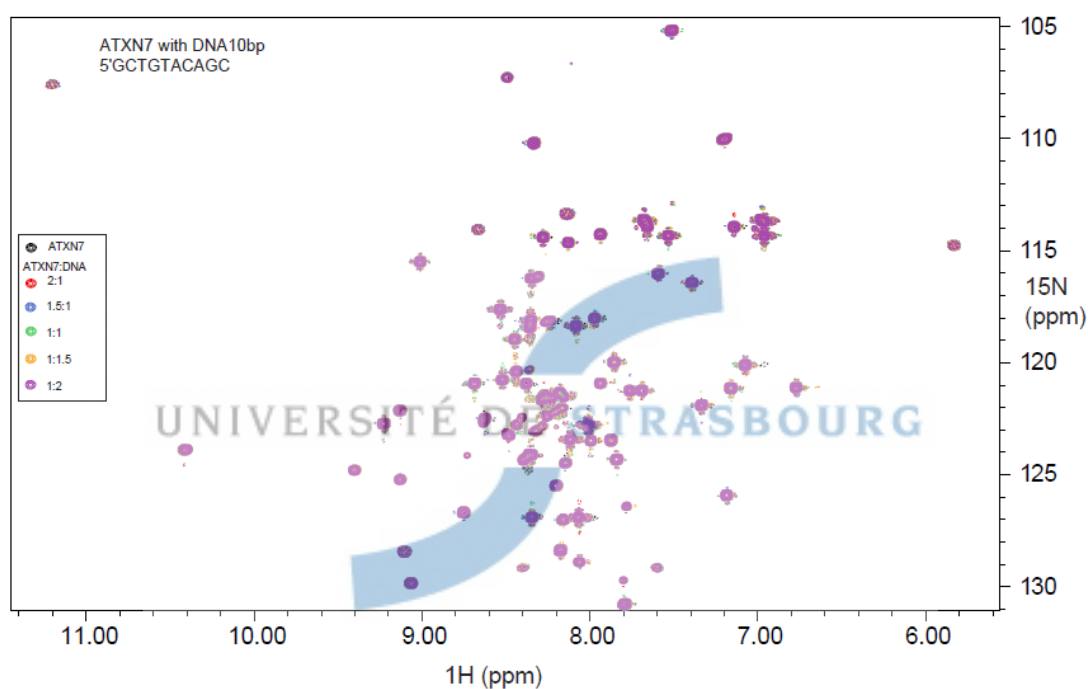


Fig.1 Superimposition of HSQC spectra of ATXN7 mixing with various ratios of DNA oligonucleotides. The DNA sequence is 5'GCTGTACAGC which is an inverted repeat. It is clear that no peak moves in the spectra, indicating a lack of DNA binding activity of ATXN7.



The logo of the University of Strasbourg, featuring two blue curved segments that form a stylized 'S' shape, positioned behind the text.

UNIVERSITÉ DE STRASBOURG

## References

1. Hvidt A, Nielsen SO: **Hydrogen exchange in proteins.** *Adv Protein Chem* 1966, **21**:287-386.
2. Herrmann T, Guntert P, Wuthrich K: **Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS.** *J Biomol NMR* 2002, **24**(3):171-189.
3. Keller R, Damberger, F., Bettendorff, P.: <http://cara.nmr-software.org/portal/>. 2004.
4. Cornilescu G, Delaglio F, Bax A: **Protein backbone angle restraints from searching a database for chemical shift and sequence homology.** *J Biomol NMR* 1999, **13**(3):289-302.
5. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM: **The Xplor-NIH NMR molecular structure determination package.** *J Magn Reson* 2003, **160**(1):65-73.
6. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM: **AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR.** *J Biomol NMR* 1996, **8**(4):477-486.
7. Guntert P: **Automated structure determination from NMR spectra.** *Eur Biophys J* 2009, **38**(2):129-143.
8. Johnson BA: **Using NMRView to visualize and analyze the NMR spectra of macromolecules.** *Methods Mol Biol* 2004, **278**:313-352.
9. Koradi R, Billeter M, Engeli M, Guntert P, Wuthrich K: **Automated peak picking and peak integration in macromolecular NMR spectra using AUTOPSY.** *J Magn Reson* 1998, **135**(2):288-297.
10. O'Connell JF, Pryor KD, Grant SK, Leiting B: **A high quality nuclear magnetic resonance solution structure of peptide deformylase from Escherichia coli: application of an automated assignment strategy using GARANT.** *J Biomol NMR* 1999, **13**(4):311-324.
11. Nilges M, Macias MJ, O'Donoghue SI, Oschkinat H: **Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin.** *J Mol Biol* 1997, **269**(3):408-422.
12. Guntert P: **Automated NMR structure calculation with CYANA.** *Methods Mol Biol* 2004, **278**:353-378.
13. Herrmann T, Guntert P, Wuthrich K: **Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA.** *J Mol Biol* 2002, **319**(1):209-227.

14. Prestegard JH, Mayer KL, Valafar H, Benison GC: **Determination of protein backbone structures from residual dipolar couplings.** *Methods Enzymol* 2005, **394**:175-209.
15. Cavalli A, Salvatella X, Dobson CM, Vendruscolo M: **Protein structure determination from NMR chemical shifts.** *Proc Natl Acad Sci U S A* 2007, **104**(23):9615-9620.
16. Atkinson RA, Saudek V: **The direct determination of protein structure by NMR without assignment.** *FEBS Lett* 2002, **510**(1-2):1-4.
17. Tjandra N, Bax A: **Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium.** *Science* 1997, **278**(5340):1111-1114.
18. Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu G, Eletsky A, Wu Y, Singarapu KK, Lemak A *et al*: **Consistent blind protein structure generation from NMR chemical shift data.** *Proc Natl Acad Sci U S A* 2008, **105**(12):4685-4690.
19. MacLean HE, Warne GL, Zajac JD: **Localization of functional domains in the androgen receptor.** *J Steroid Biochem Mol Biol* 1997, **62**(4):233-242.
20. Truss M, Beato M: **Steroid hormone receptors: interaction with deoxyribonucleic acid and transcription factors.** *Endocr Rev* 1993, **14**(4):459-479.
21. Schoenmakers E, Verrijdt G, Peeters B, Verhoeven G, Rombauts W, Claessens F: **Differences in DNA binding characteristics of the androgen and glucocorticoid receptors can determine hormone-specific responses.** *J Biol Chem* 2000, **275**(16):12290-12297.
22. Verrijdt G, Schoenmakers E, Haelens A, Peeters B, Verhoeven G, Rombauts W, Claessens F: **Change of specificity mutations in androgen-selective enhancers. Evidence for a role of differential DNA binding by the androgen receptor.** *J Biol Chem* 2000, **275**(16):12298-12305.
23. Shaffer PL, Jivan A, Dollins DE, Claessens F, Gewirth DT: **Structural basis of androgen receptor binding to selective androgen response elements.** *Proc Natl Acad Sci U S A* 2004, **101**(14):4758-4763.
24. Hsieh JC, Whitfield GK, Oza AK, Dang HT, Price JN, Galligan MA, Jurutka PW, Thompson PD, Haussler CA, Haussler MR: **Characterization of unique DNA-binding and transcriptional-activation functions in the carboxyl-terminal extension of the zinc finger region in the human vitamin D receptor.** *Biochemistry* 1999, **38**(49):16347-16358.
25. Haelens A, Verrijdt G, Callewaert L, Christiaens V, Schauwaers K, Peeters B, Rombauts W, Claessens F: **DNA recognition by the androgen receptor: evidence for an alternative DNA-dependent dimerization, and an active role of sequences**

## References

- flanking the response element on transactivation.** *Biochem J* 2003, **369**(Pt 1):141-151.
26. Khorasanizadeh S, Rastinejad F: **Nuclear-receptor interactions on DNA-response elements.** *Trends Biochem Sci* 2001, **26**(6):384-390.
27. Jakob M, Kolodziejczyk R, Orłowski M, Krzywda S, Kowalska A, Dutko-Gwozdz J, Gwozdz T, Kochman M, Jaskolski M, Ozyhar A: **Novel DNA-binding element within the C-terminal extension of the nuclear receptor DNA-binding domain.** *Nucleic Acids Res* 2007, **35**(8):2705-2718.



The logo of the University of Strasbourg, featuring two blue curved segments that form a stylized 'S' shape, positioned behind the text.

UNIVERSITÉ DE STRASBOURG

# Curriculum Vitae

Ying-Hui WANG

Born September 2<sup>nd</sup> 1978 in Kaohsiung, Taiwan



## Education

- 2005-2010      PhD. Molecular and Cell Biology  
Dept. of Structural Biology and Genomics  
Institute of Genetics and Molecular and Cellular Biology, University of Strasbourg  
Thesis: Molecular interaction of zinc finger domain: Study of androgen receptor  
DNA binding domain and SCA7 domain of Ataxin 7 by NMR
- 2001-2003      MS. Science  
Dept. of Life Science, National Tsing Hua University (Hsinchu, Taiwan)  
Thesis: Taiwan Cobra Toxin induced intracellular calcium increase and its biological  
implication.
- 1996-2000      BS. Science  
Dept. of Chemistry, National Tsing Hua University (Hsinchu, Taiwan)
- 1990-1996      St. Dominic high school (Kaohsiung, Taiwan)

## Working Experience

- 2003-2004      Research assistant  
National Genotyping Center  
Inst. Biomedical Science, Academia Sinica, Taipei, Taiwan
- 2004-2005      Research assistant  
Dr. Tai-Huang Huang's lab  
Inst. Biomedical Science, Academia Sinica, Taipei, Taiwan  
Project: Solution structure and dynamic analysis of small ubiquitin-like modifier 2  
(SUMO-2). PDB code: 2awt

## Publication

1. J. Bonnet, **Y.-H. Wang**, R. A. Atkinson, J. Koffler, C. Romier, A. Hamiche1, L. Tora, D. Devys, and B. Kieffer  
Histone deubiquitination by SAGA is modulated by an atypical zinc finger domain of Ataxin-7  
(manuscript under revision in EMBO rep)
2. **Y.-H. Wang**, A. Monge, J. Céraline, B. Kieffer and R. A. Atkinson  
Altered specificity of a mutated androgen receptor in prostate cancer is associated with a change in the  
surface potential of the DNA-binding domain  
(in preparation)

## Conference

2007

8th Course: Biophysics and the Challenges of Emerging Threats.

(NATO Advanced Study Institute and part of the International School of Biological Magnetic Resonance)

Ettore Majorana Center for Scientific Culture in Erice, Sicily (EMFCSC)

Selected student

Present by poster

Topic: Investigation DNA response element by androgen receptor DNA binding domain using NMR.





The logo of the University of Strasbourg, featuring two blue curved segments that form a stylized 'S' shape, positioned behind the text.

UNIVERSITÉ DE STRASBOURG