

**University of Florence**

**University Louis Pasteur Strasbourg**

**Doctorate in Chemistry, Doctorate in Life Sciences**



# **Nuclear Hormones Receptors: Structure and molecular dynamics of allosteric complexes**

Ph.D. thesis of  
**SERENA SIRIGU**

## **Directors**

Prof. Ivano Bertini

Dr. Dino Moras

## **Supervisor**

Dr. Natacha Rochel

## Acknowledgments

I am grateful to the members of the jury, Dr. William Bourguet, Prof. Stefano Ciurli, Prof. Stefano Mangani and Prof. Bruno Kieffer for having accepted to judge my work.

I am particularly grateful to my thesis directors, Prof. Ivano Bertini and Dr. Dino Moras for allowing me to undertake my studies at first rate research centers such as CERM and IGBMC, where I worked in an incredibly rich and scientifically stimulating environment.

I am very thankful to my supervisor Dr. Natacha Rochel, who proficiently and patiently guided me during the past years. Thanks to Dr. Arnaud Poterszman for his constant supervision and support.

I would also like to thank all the collaborators that provided me with their experience. Thanks to Dr. Annick Dejaegere for the Molecular Dynamic Studies, Dr. Sarah Sanglier for the Mass Spectrometry analysis and Dr. Gabrielle Lutz for the help with Plasmonic Resonance experiments. Thanks to the IGBMC Structural Biology Platform and in particular to Didier, Loubna and Edouard for their precious suggestions and practical help. Big thanks also to Virginie and Pierre who initiated me to the art of protein crystallization and to Dr. Sato Yoshiteru for collaborating with me at protein structure resolution by X-ray crystallography. Thanks to Dr. Francesca Cantini for her precious help with NMR experiments.

Thanks also to all the members of the CERM and IGBMC who contributed to render the working environment a nice and pleasant place. Thanks also very much to Anne who took care of the paperwork.

To all my dear friends in the lab, in Florence as well as in Strasbourg. I would not make everyone's name but I'm sure you'll recognize yourselves in these words. You're one of my greatest joys! Thanks for always having made me feel at home.

To my family. To my uncles Gaetano, Antonio and Pasquale who are always with me. To my mum (my greatest supporter), to my brother and to my dad who gave me the strength and the courage in life. To Robin who gives me unconditional love.

*A nonna con tutto il mio amore,*

## Summary

### Hormone Nuclear Receptors : Structure and Molecular Dynamics of Retinoic Acid Receptor (RAR) allosteric complexes

The Retinoic Acid Receptor (RAR) belongs to the large family of nuclear receptors (NRs). Nuclear receptors are ligand dependent transcription activators: they mediate the effect of specific ligands by means of interacting with the general transcription machinery and modulating the expression of a specific subset of genes.

RAR acts as a heterodimer with its functional partner, the retinoid X receptor (RXR) and exists in three main forms, the isotypes  $\alpha$ ,  $\beta$ ,  $\gamma$ . RAR shares a common modular organization with prototypical NRs with a DNA binding (DBD) domain and a ligand binding domain (LBD).

Nuclear receptors are phosphoproteins: RAR is phosphorylated at the LBD by the cyclic AMP dependent protein kinase PKA and subsequently at the N-terminal domain (AB domain) by the heterotrimeric complex CAK composed of Cdk7 kinase, Cyclin H, and MAT1 and belonging to the TFIIH general transcription factor. Phosphorylation at the ligand binding domain at position S369 for RAR $\alpha$  and S371 for RAR $\gamma$  by PKA enhances the binding of Cyclin H to the receptor therefore facilitating the phosphorylation of the AB domain by CAK and activating the transcription of retinoic target genes.

Our project aims at providing new structural insights to understand at atomic level how phosphorylation by PKA participates in nuclear receptors signalling. The crystal structures of

hRAR $\gamma$  and hCycH have already been determined in our group.

Now we aim at investigating the consequences of PKA phosphorylation on the structure of the RAR LBD. In this perspective we need to compare the high resolution crystal structures of the unphosphorylated RAR LBD with that of the phosphorylated receptor or with that of the S371E mutant in which the mutation is well known to mimic a phosphorylation event. We need as well to perform NMR studies on the unphosphorylated RAR LBD and on the phosphorylated receptor or the S371E mutant

in order to highlight differences in the structure dynamics due to the phosphorylation of the residue S371.

The work that I performed during my thesis can be divided in three main sections:

### Section 1- Samples preparation

A point mutation exchanging a serine for a glutamic acid was introduced respectively in position 369 and 371 of hRAR $\alpha$  and hRAR $\gamma$  (LBD) in order to mimic a phosphorylation event. Starting from the corresponding expression vectors, small scale expression tests were performed to optimize the conditions for the expression of the nuclear receptors hRAR $\alpha$ , hRAR $\alpha$ S369E, hRAR $\gamma$ , hRAR $\gamma$ S371E, hRXR $\alpha$ . Scaling up to large scale cultures did not require any modifications in the set up conditions. Nuclear receptor samples enriched with the NMR active isotope  $^{15}\text{N}$  were produced to perform subsequent NMR mobility experiments. A common purification protocol consisting in a metal affinity chromatography followed by a size exclusion chromatography was set up for the nuclear receptors.

The full length hCycH was expressed to perform biophysical characterization experiments of the interaction between hCycH and hRAR $\alpha$ .

The catalytic domain of the cyclic AMP dependent protein kinase (c-PKA) was expressed and purified in order to perform in vitro phosphorylation essays onto RAR. The phosphorylation of hRAR $\alpha$  onto the residue Ser 369 by PKA has been checked by immunoblotting and reconfirmed by LC mass spectrometry analysis.

### Section 2- Crystal structure of hRAR $\gamma$ S371E LBD

The ligand binding domain of hRAR $\alpha$ S369E was expressed and purified. The protein bound to the agonist ligand 9-cis retinoic acid was concentrated up to 6 mg/ml and crystallized in presence of a SRC1 co-activator peptide. Crystals diffracted on a synchrotron source up to 2.7 Å but resulted highly mosaic. Several attempts to optimize the crystallization conditions resulted unfortunately not successful. The other protein

isotype hRAR $\gamma$ S371 E (LBD) was expressed and purified with a yield of 7 mg of pure protein per 1 litre of culture. The protein was concentrated up to 6 mg/ml. The hRAR $\gamma$ S371E bound to the ligand 9-cis retinoic acid (holo receptor) crystallized over night in presence of detergents. Optimized crystals were tested on a synchrotron source and a complete data set was collected at 1.7 Å. Crystals belong to the tetragonal space group P4<sub>1</sub>2<sub>1</sub> 2 with one monomer per asymmetric unit. The structure was solved by molecular replacement using hRAR $\gamma$ WT as a starting model. After refinement the reported values for Rfactor and Rfree were respectively of 19.4% and 23.4%.

The human RAR $\gamma$ S371E-holo-LBD adopts a classical anti-parallel alpha helical "sandwich" conformation firstly described for hRAR $\gamma$ . The electron density map showed the presence of the ligand buried into the ligand binding pocket and an extra electron density around position 371 confirmed the mutation of the serine residue into a glutamic acid.

The hRAR $\gamma$ S371E LBD and hRAR $\gamma$  LBD structures are isomorphous with an r.m.s. deviation of 0.47 Å on the overall structure. Analysis of the r.m.s.d values obtained from the superposition of RARWT and RAR mutant carbon alphas of the backbone enlightened that two regions in RAR $\gamma$ S371E display larger r.m.s.d values. Interestingly one of these two regions involves residues (340-343) which are located in the region of the receptor that is responsible for the recognition and binding of hCycH and previously shown to be essential for hCycH docking. The analysis of the structures of hRAR $\gamma$ WT and the mutant hRAR $\gamma$ S371E showed very similar overall orientation of the side chains. Remarkably a significant difference in the orientation of the side chains involves again the region of the receptor responsible for hCycH binding. More precisely this difference is located at the level of the residue ASP340. In hRAR $\gamma$ S371E this residue, belonging to loop 8-9, forms a salt bridge with ARG387 belonging to helix H10. This salt bridge is not formed in the hRAR $\gamma$ WT due to the orientation of the ARG387 which now points toward the solvent. Our results were confirmed by molecular dynamic simulations performed in collaboration with Dr. Annick Dejaegere's group in Strasbourg. Simulations carried out onto hRAR $\gamma$ S371E enlightened the presence of the salt bridge which is unmodified during in the dynamics.

The same simulations carried onto hRAR $\gamma$ WT showed that the formation of the salt bridge between ARG387 and ASP340 is transient.

Section 3- Dynamic studies of RAR $\alpha$  wild type and RAR $\alpha$ S369E (LBD) by Nuclear Magnetic resonance (NMR).

X-ray crystallography and NMR spectroscopy are the two main techniques that can provide the structure of macromolecules at atomic resolution. Both techniques are well established and play already a key role in structural biology as a basis for a detailed understanding of molecular functions. X-ray crystallography requires single crystals whereas NMR measurements are carried out in solution under conditions that can be as close as possible to the physiological conditions. Furthermore, the analysis through NMR spectroscopy easily allows the characterization under several, different experimental conditions, such as different ionic strength and pH. NMR measurements not only provide structural data but can also supply information on the internal mobility of proteins on various time scales, on protein folding and on intra, as well as, intermolecular interactions. Protein mobility has an increasing relevance to our understanding of biological and chemical phenomena for the recognized importance of protein internal motions in biological processes. Macromolecular functions are often associated to energetic transition, which are intimately connected with structural changes and molecular flexibility.

The measurement of  $^{15}\text{N}$  relaxation rates in isotopically enriched proteins is particularly useful for obtaining dynamic information. In  $^{15}\text{N}$  relaxation experiments, one creates non-equilibrium spin order and records how this relaxes back to equilibrium. At equilibrium, the  $^{15}\text{N}$  magnetization is aligned along the external field, and this alignment can be changed by radio frequency pulses. The magnetization will relax back to equilibrium along the direction of the magnetic field with a longitudinal relaxation constant R1. The magnetization can also have a second component perpendicular to the external magnetic field. The relaxation constant for this spin component to return to equilibrium is called transverse relaxation constant R2.

In order to highlight differences in the dynamics of wild type RAR and its mutant we performed  $^{15}\text{N}$  longitudinal and transverse relaxation rates measurements on both proteins at 500 MHz and 298K. We focused our attention on the  $\alpha$  isotype of RAR for which the

most extensive biologic data are available. The large majority of the analyzed residues feature very similar rates except for 10 residues that in RAR $\alpha$ S369E have higher longitudinal rates with respect to the wild type. To analyze in details the difference in mobility between RAR $\alpha$  wild type and its mutant protein backbone assignment is needed. Due to RAR high molecular weight, protein assignment will require RAR $\alpha$  samples enriched with  $^2\text{H}/^{13}\text{C}/^{15}\text{N}$  isotopes.



# Table of Contents

## Chapter 1 : Introduction to Nuclear Receptors

1. The Nuclear Receptor family and its ligands
2. Modular Organization of NRs
  - 2.1 Overview
  - 2.2 The A/B Domain – Ligand independent transactivation domain
    - 2.2.1 The structure
    - 2.2.2 Coactivators mediated folding of the AF-1 domain
  - 2.3 The C Domain – DNA Binding Domain
    - 2.3.1 The structure
    - 2.3.2 DBD interaction with DNA
  - 2.4 The D Domain- Just a hinge?
  - 2.5 The E Domain- Ligand binding domain
    - 2.5.1 The structure
    - 2.5.2 The agonist conformation
    - 2.5.3 The antagonist conformation
    - 2.5.4 The ligand binding pocket
    - 2.5.5 The dimerization interface
  - 2.6 The F Domain
    - 2.6.1 The structure
    - 2.6.2 The function
3. How do NRs activate transcription?
  - 3.1 Coactivators
  - 3.2 The coregulator boxes
    - 3.2.1 The p160 family of Coactivators
    - 3.2.2 CBP/p300 Cointegrators
    - 3.2.3 TRAP and DRIP
    - 3.2.4 Corepressors
    - 3.2.5 Transcription activation
4. Posttranslational modifications on NRs
  - 4.1 Nuclear Receptors Phosphorylation
5. Cellular Localisation

## Chapter 2 : Retinoic Acid Receptors

6. Retinoids
  - 6.1 The chemistry of retinoids
  - 6.2 Vitamin A metabolism
  - 6.3 Retinoic acids

- 6.4 Retinoids acid in cancer
- 7. Retinoic Acid Receptor (RAR) and Retinoid X Receptor (RXR)
  - 7.1 RARs and RXRs transduce the signal of RA
  - 7.2 Retinoid response elements
  - 7.3 RAR posttranslational modifications
    - 7.3.1 RAR posttranslational modifications
    - 7.3.2 RAR Phosphorylation
- 8. Our project

### **Chapter 3 : from RAR gene to Crystal Structure, Molecular Dynamics and NMR Studies in Solution**

- 9. Protein expression and purification
  - 9.1 (His)<sub>6</sub> hRAR $\gamma$  WT and h(His)<sub>6</sub> hRAR $\gamma$  S371E
  - 9.2 (GST)-hRAR $\alpha$ , (GST)-hRAR $\alpha$ S369E, (GST)-hRXR $\alpha$
  - 9.3 hCyclin H
  - 9.4 PKA
- 10. Crystallization of hRAR $\gamma$ S371E (LBD) in complex with 9-cis retinoic acid and X-ray crystal structure resolution
  - 10.1 Crystallization of hRAR $\alpha$ S369E (LBD) in complex with 9-cis retinoic acid and a coactivator peptide
  - 10.2 hRAR $\gamma$ S371E (LBD) in complex with 9-cis retinoic acid
    - 10.2.1 Preliminary crystallization trials
    - 10.2.2 Optimisation and scaling up of the crystallization conditions
    - 10.2.3 Data Processing, Molecular Replacement and Structure Refinement
- 11. Molecular dynamics studies of the effects of the S371 phosphorylation onto RAR $\gamma$ 
  - 11.1 Structures preparation for molecular simulation
  - 11.2 Analysis
    - 11.2.1 RMSD
    - 11.2.2 Fluctuations RMS
    - 11.2.3 Hydrogen bonds analysis
    - 11.2.4 Dihedral angles analysis
- 12. Dynamic studies of the RAR $\alpha$  wild type and RAR $\alpha$ S371E by nuclear magnetic resonance (NMR).
  - 12.1 Preliminary 2D heteronuclear spectra
  - 12.2 RAR $\alpha$  wild type and RAR $\alpha$ S371E <sup>15</sup>N longitudinal and transverse longitudinal relaxation rates measurements

### **Chapter 4 : Results and Discussion**

- 13. X-ray crystal structure of hRAR $\gamma$ S371E (LBD)
  - 13.1 X-ray crystal structure of hRAR $\gamma$ S371E (LBD) in complex with 9-cis retinoic acid
- 14. Molecular dynamics studies on hRAR $\gamma$ S371E (LBD)
  - 14.1 Molecular dynamics studies onto RAR $\gamma$
- 15. NMR
  - 15.1 NMR studies of hRAR $\alpha$ S371E (LBD) and of hRAR $\alpha$  WT (LBD) in solution
- 16. Conclusions and Perspectives

## Chapter 4 : Other Results

- 17. Biophysical characterization of the interaction between hCycH and hRAR $\alpha$ 
  - 17.1 Surface Plasmonic Resonance (SPR)
  - 17.2 Preliminary study of the interaction of RAR $\alpha$  with PKA through NMR

## Chapter 4 : Material and Methods

### 18. Materials

#### 18.1 Molecular biology

##### 18.1.1 Plasmid Amplification and Purification : DNA Mini Prep

##### 18.1.2 LR-Gateway reactions

##### 18.1.3 Polymerase chain reaction (PCR)

###### 18.1.3.1 Point mutation at position 369 of RAR $\alpha$ and position 371 of RAR $\gamma$ (LBDs)

###### 18.1.3.2 Determination of the efficiency of the LR reaction

#### 18.2 Protein Expression

##### 18.2.1 Cell Transformation

##### 18.2.2 Small Scale Protein Expression Tests

##### 18.2.3 Small Scale Protein Purification Test

##### 18.2.4 Over-night Precultures

##### 18.2.5 Large Volume Cultures

##### 18.2.6 Large Volume Cultures Isotopically Enriched With $^{15}\text{N}$

#### 18.3 Purification

##### 18.3.1 Sonication and Lysate Clarification

#### 18.4 Metal Affinity Chromatography

##### 18.4.1 Ni $^{2+}$ affinity chromatography

##### 18.4.2 Co $^{2+}$ affinity chromatography

##### 18.4.3 GST-tag Fusion RAR

##### 18.4.4 Size Exclusion Chromatography (Gel Filtration)

#### 18.5 Buffer exchange of RAR and RXR

##### 18.5.1 Buffers for : His-RAR $\gamma$ S371E and RAR $\gamma$ WT

##### 18.5.2 Buffers for : GST-RAR $\alpha$ S371E and RAR $\alpha$ WT

##### 18.5.3 Buffers for : His-PKA

##### 18.5.4 Buffers for : His-CycH

##### 18.5.5 Phosphorylation buffer for RAR $\alpha$ WT

#### 18.6 SDS-PAGE Denaturing Gel Electrophoresis

### 19. Methods

## 19.1 X-ray crystallography

### 19.1.1 Basics

### 19.1.2 Protein crystallization

#### 19.1.2.1 Nucleation

#### 19.1.2.2 Growth

### 19.1.3 Parameters that influence Crystallization

### 19.1.4 Crystallization Techniques

#### 19.1.4.1 The Vapour Diffusion Method

### 19.1.5 Crystals Cryo-protection.

### 19.1.6 Nuclear Magnetic Resonance (NMR)

## **Public Communications**

## **Bibliography**

# **Chapter 1 : Introduction to Nuclear Receptors**

# Nuclear Receptors

Nuclear receptors (NRs) constitute a large class of structurally-related proteins that act as ligand inducible transcription factors. Nuclear receptors bind specific DNA regulatory elements (HREs) and are functionally active as monomers or homo-heterodimers. Their distinctive feature is to bind small lipophilic molecules such as hormones, vitamins derivatives and metabolic intermediates that act as signalling ligands. Ligand binding to the receptor induces radical conformational changes that modulate the receptor transcriptional activity and therefore the expression of subsets of target genes. Nuclear receptors take active part into the “cross-talk” with other signal transduction pathways thus integrating intra and extra-cellular signals and modulating a multitude of genetic programs that regulate all stages of the life of multicellular organisms such as embryonic development, differentiation of cellular phenotypes, cell proliferation and apoptosis.

Malfunctioning of the receptors leads to severe diseases such as cancer, diabetes and obesity<sup>5, 74, 100</sup>.

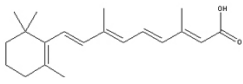
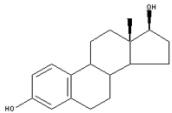
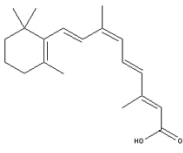
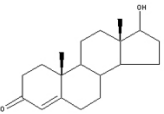
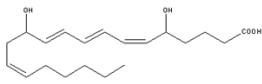
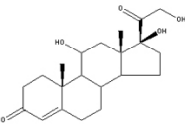
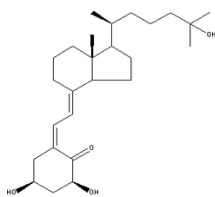
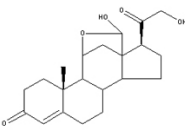
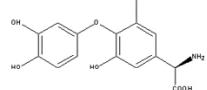
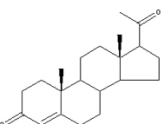

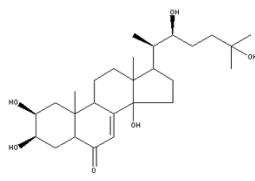
## 1. The Nuclear Receptor family and its ligands

Classical receptors were identified on the basis of their reported ability to transduce the signal of known hormones. Biochemical and structural studies revealed common structural features for these receptors and allowed a systematic screening of c-DNA libraries. This extensive research brought to the isolation of several c-DNA encoding for proteins that shared high sequence conservation with the nuclear receptors but for which the corresponding ligand was previously unknown. These new proteins were defined as “Orphan Receptors”. For long time a clear distinction was made between classical and orphan nuclear receptors. However the recent discovery of several ligands for orphan receptors and the presence of orphan receptors in all branches of NRs phylogenetic tree moderated the necessity for this discrimination. The history of the evolution of the NR

superfamily has been reconstituted on the basis of broad sequence analysis and led to nuclear receptors grouping in six subfamilies composed by the following members<sup>101, 115</sup> :

- I. The Thyroid hormone (TR), Vitamin D (VDR), Retinoic Acid (RAR), Peroxisome Proliferator-Activated (PPAR), Ecdysone (EcR) Receptors and numerous Orphan Receptors as ROR, Rev-erb.
- II. The Retinoid X receptor (RXR), the Chicken Ovalbumin Upstream Promoter Transcription Factor (COUP-TF) and the Hepatocyte Nuclear Factor 4 (HNF4).
- III. The Estrogen (ER), Mineralocorticoid (MR), Glucocorticoid (GR), Androgen (AR), Progesterone (PR), and the orphan Estrogen-Related Receptors (EER).
- IV. The Nerve Growth Factor-Induced clone B (NGFIB) orphan receptor.
- V. The Steroidogenic Factor1 (SF1) and the Drosophila Fushi Tarazu Factor1 (FTZ-F1)
- VI. The Germ Cell Nuclear Factor (GCNF1) receptor

This classification provided evidence for a correlation between the receptor position in the phylogenetic tree and their dimerization features, suggesting that most probably the common ancestral NR bound DNA as a homodimer and that the ability to heterodimerize was acquired when receptor classes I and IV differentiated. Instead no correlation can be found between chemically-related hormones and receptor position in the phylogenetic tree leading to the conclusion that the ability to bind a ligand was not inherited by a common ancestor but rather acquired independently at different stages of receptors evolution. The analysis of the phylogenetic tree suggests that nuclear receptors are found only in metazoans and that their evolution went through two “waves of duplication”: the first at the appearance of all metazoans originated the six subclasses whether the second, at the differentiation of anthropoids and vertebrates, originated different isotypes for a same family (i.e. RAR  $\alpha$ ,  $\beta$ ,  $\gamma$  or ER  $\alpha$ ,  $\beta$ )<sup>70, 100, 101</sup>

NR	Ligand	Chemical formula	NR	Ligand	Chemical formula
RAR	<i>trans</i> -Retinoic acid		ER	Oestradiol-17β	
RXR	9-cis Retinoic acid		AR	Testosterone	
PPARα	Leucotriene B4		GR	Cortisol	
VDR	1α,25-dihydroxy vitamin D <sub>3</sub>		MR	Aldosterone	
TR	3,5,3'-L-triiodo thyronine		PR	Progesterone	
HNF4	Palmitoyl -CoA		EcR	20-Hydroxy-ecdysone	

**Table.1-** Some of the Nuclear Receptors with their cognate ligand

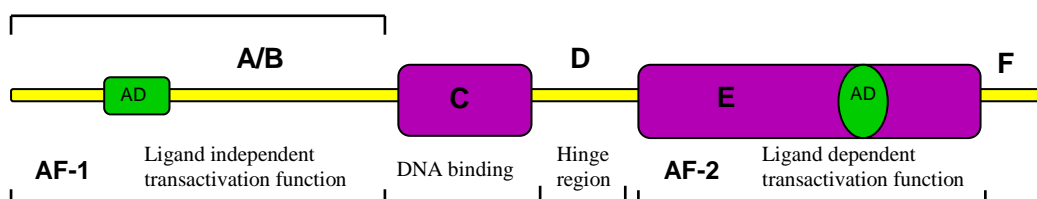


## 2. Modular Organization of NRs

### 2.1 Overview

All NRs share a common organization in functional domains, each encoding a specific function. NRs contain up to six of such regions designated from A to F (Fig.1).

The N-terminal regions A/B, commonly identified as the A/B domain, display the weakest evolutionary conservation amongst NRs and harbour one ligand-independent and constitutive active transcriptional activation function (AF-1). The DNA-binding domain (C domain) represents the most evolutionary conserved region in NRs and it notably contains a short motif responsible for the sequence specific DNA recognition. The D region does not display a high evolutionary conservation. It is classically thought to serve as a “hinge” region that reduces steric hindrance problems and allows the nuclear receptor to adopt different conformations. The E domain or ligand binding domain (LBD) is the hallmark of each nuclear receptor. This highly conserved domain bears the ligand binding pocket as well as the ligand-dependent transcription activation function (AF-2) and a large interaction surface which allows interactions with multiple partners such as homo and heterodimeric partners, corepressors and coactivators. Some receptors possess a C-terminal region called F domain with weak evolutionary conservation. The role of this region is still unknown.



**Fig.1- Structural organization of NRs in distinct domains.** The evolutionary conserved regions C/E are indicated by the purple boxes while the yellow bar represents the divergent regions A/B, D, F.

## 2.2 The A/B Domain – Ligand independent transactivation domain

### 2.2.1 Structure

The N-terminal A/B is the least conserved region among NRs. Isoforms of the same nuclear receptor, originated from the same gene by alternative splicing or different promoter usage, differ exclusively in their NH<sub>2</sub>-terminal region. This is the case, for example, for the thyroid hormone receptor (TR) isoforms TR $\beta$ 1 and TR $\beta$ 2 or for the RAR gene isoforms ( $\alpha,\beta,\gamma$ ). The A/B domain is also sensibly variable in length and it can range from 23 amino acids in vitamin D receptor, to 602 amino acids in the mineralocorticoid receptors<sup>181</sup>. Nuclear receptors contain at least two domains that possess the ability of activating transcription (AFs), a ligand-independent activation function (AF-1) generally localized in the N-terminal region and a ligand dependent activation function (AF-2) generally harboured in the ligand-binding domain<sup>5</sup>. The synergy between the AFs guarantees the full transcriptional efficiency of the nuclear receptor. Nevertheless it has been shown that the AF-1 function is able to activate transcription in a constitutive manner. Truncated estrogen (ER) or progesterone (PR) receptors containing the AF-1 function associated with the DNA binding domain efficiently stimulate transcription in a ligand independent fashion. The AF-1 transactivation function exhibits cell type, DNA-binding domain and target gene promoter specificity<sup>15, 118</sup>. The origin of this strong dependence is still unclear but might be related to the cell specific mode of action or to the presence within the cell of specific AF-1 coactivators.

To date no structure determination of the A/B nuclear receptor domain is available due to the lack of structural organization for this region. Magnetic resonance and circular dichroism studies demonstrated that the AF-1 region of glucocorticoid receptor (GR) and the N-terminal region of oestrogen receptor are disordered in solution. Proteolysis essays pointed out that progesterone receptor (PR) N-terminus is also poorly structured. Experiments performed in presence of  $\alpha$ -helical stabilizing agents as trifluoroethanol demonstrated that the AF-1 structure is incline to adopt a  $\alpha$ -helical conformation. When residues involved in the putative  $\alpha$ -helix are mutated in prolines the ability of AF-1 to

form  $\alpha$ -helices in trifluoroethanol is remarkably reduced and concomitantly the transactivation activity of this domain decreases suggesting that the formation of  $\alpha$ -helices is important for AF-1 mediated gene activation<sup>181</sup>.

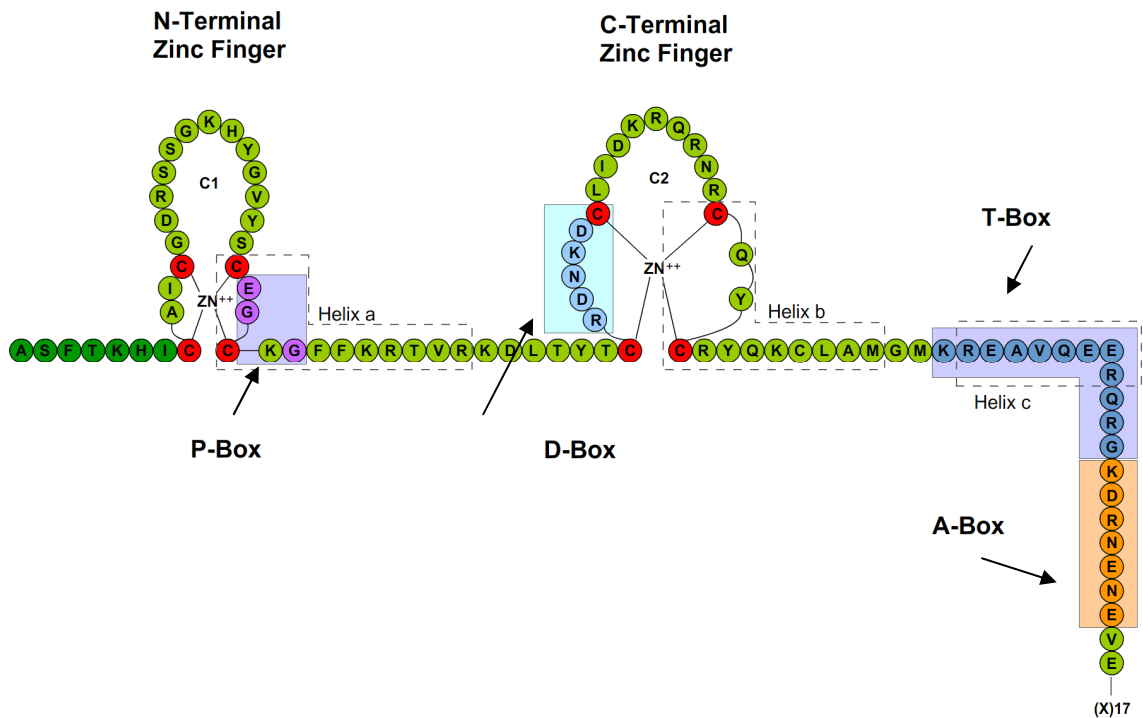
### 2.2.2 Coactivators mediated folding of the AF-1 domain

Regulation of transcriptional activity by nuclear receptors requires a battery of coregulatory proteins often constituted by subunits of larger multiprotein complexes. Coregulators are divided into coactivators or corepressors depending on whether they mediate transcription regulation or repression in a given cell type and target gene context. Binding of coactivators to NRs is mediated by one or more of its activation domains AFs. Several coactivators have been shown to interact with the AF-1 function of nuclear receptors as it is the case for the TATA box-binding protein (TBP) and the c-AMP response element binding protein that have been shown to interact directly with the AF-1 region of GR<sup>2, 78, 181</sup>. The mechanism of these interactions is still poorly understood due to lack of structural or sequence homology among the coregulators. All recently studied showed that AF-1 displays a lack of structural order in solution. Interaction of these regions with coregulators seem to induce a fold into more ordered structures as it is the case for the ER $\alpha$  terminal domain and the TATA box-binding protein (TBP). According to this folding-induced mechanism the activation domain would firstly bind to the coactivator via weak electrostatic interactions. Secondly, hydrophobic interactions would direct the formation of a more stable complex where the activation domain adopts the conformation that is optimized for that interaction partner<sup>5, 181</sup>.

## 2.3 The C Domain – DNA Binding Domain

### 2.3.1 The structure

The C domain is the highest conserved domain among nuclear receptors. It contains the DNA-binding domain (DBD) responsible for the specific-sequence DNA recognition. Several three-dimensional structures have been solved for nuclear receptors DBDs either by X-ray crystallography or nuclear magnetic resonance (RMN)<sup>100</sup>. All available NMR structures show that DBDs are monomers in solution (i.e. ER $\alpha$  and GR DBDs) and that they are poorly structured. On the contrary DBDs in complex with DNA is highly ordered<sup>104, 158</sup>. One of the features of the DNA-binding domain consist in two zinc finger motifs, each constituted of four conserved cystein residues tetrahedrally coordinating a Zn<sup>2+</sup> ion. Within the DBD four sequence elements (the P, D, A and T-boxes) have been identified to be responsible for the discrimination of the target DNA and the functionality of the domain (Fig.2). Residues constituting the P-box are responsible for selective response element recognition as in the case of the discrimination between the EREs and the GREs<sup>157</sup>. Three-dimensional structures demonstrated that in GR and ER homodimers the D-box stabilizes the dimerization interface<sup>57</sup> and selects the length of the spacer between the two hemi-sites of the palindrome<sup>117</sup>. The C-terminal extension of the C-domain (CTE) harbours the A and T boxes and is critical for monomer DNA binding<sup>117</sup> and discrimination of the spacing distance between the two halves of the palindrome as demonstrated for the RXR-RAR and the RXR-TR heterodimers<sup>74, 143</sup>. The core of the DBD is constituted by two amphipathical  $\alpha$ -helices (N-terminal H1 and C-terminal H2), packed at right angles and responsible for the recognition of the hemi-site of the DNA response element<sup>100</sup>. Side-chains amino acids of the H1 establish interactions with each DNA major groove half site whereas H2 contributes to the stabilization of the overall structure<sup>8, 104</sup>. The NMR structure of RXR (DBD) and crystal structure of TR $\beta$ -RXR $\alpha$  (DBD) show that the CTE forms an additional  $\alpha$ -helix H3<sup>103</sup>.



**Fig.2- Schematic illustration of the RXR DNA binding domain (DBD).** The regions involved in the response element selection are represented by the highlighted boxes (P, D, T, A) and involve two zinc-finger motifs and a C terminal extension (CTE). The P box is part of the helix H1, the D box is localized in the second zinc finger while the T and A boxes are located in the CTE.

### 2.3.2 DBD interaction with DNA

All nuclear receptors regulate transcription by binding to specific DNA sequences, known as Hormone Response Elements (HREs) and located at the 5' end of target genes. Extensive sequence analysis of HREs revealed that nuclear receptors recognize derivatives of the six base pair DNA sequence 5'-PuGGTCA (Pu =A, G). Nuclear receptors can bind to their response elements as monomers, homodimers or heterodimers. At present only orphan receptors such as NGFI-B and SF-1 and steroid nuclear receptors such as TR and GR are known to bind to DNA as monomers<sup>70, 88, 98, 100, 142, 185</sup>. In this case the DNA core motif contains an extra A/T-rich sequence at the 5'end and two subfamilies of NRs binding as monomers can be identified by their discrimination for 5'-AAA or 5'-TCA extension (NGFI-B and SF-1 subfamilies respectively). The Helix H3 at the CTE of the DNA-binding domain plays an important role for the stabilization of the monomeric DNA-binding nuclear receptors as it

provides additional interactions with the DNA's minor groove upstream of the hemi-site recognition core<sup>194</sup>.

Most of nuclear receptors bind to DNA as homo-heterodimers. Their RE is constituted by a duplication of the core recognition motif and the two half-sites can arrange as Palindromes (PAL<sub>x</sub>), Direct Repeats (DR<sub>x</sub>) or Everted Repeats (ER<sub>x</sub>) where the x indicates the base pairs separating the two half-sites.

Steroid receptors as GR specifically bind as homodimers to REs that are the duplication of the 5'-PuGG (T/A) CA core motif, as palindromes or everted repeats with 3 base pairs spacing. Estrogen Receptor is an exception as it can bind to widely spaced direct repeats. The D-box in this case forms a DNA induced dimerization interface and is responsible for the length selection of the spacer between the two half sites of the palindrome<sup>171</sup>.

Non steroid receptors as VDR, RAR, RXR, and TR can bind as homodimers or heterodimers to the response element 5'-PuGGTCA arranged as direct or inverted repeats. The formation of heterodimers is strongly favourable as it brings new elements into the response elements repertory, increases the efficiency of DNA binding and transcriptional activity in cells in vivo<sup>50</sup>. RXR plays a double role as it can either form homodimers or act as promiscuous partner in the formation of heterodimers with VDR, RAR, RXR, and TR. The nature of the RXR partner in the heterodimer will dictate the preference for the length of the spacer in the direct repeat response elements as schematically represented by the 1-to-5 rule<sup>115</sup> (Table2).

It must be pointed out that the same receptors can adopt a specific binding fashion according to the specific response element.

Spacer=X (NTs)	RE	Acronym	Receptor Complex
1	DR1	RXRE, PPARE, etc.	PPAR-RXR, etc.
2	DR2	RARE	RAR-RXR
3	DR3	VDRE	RXR-VDR
4	DR4	TRE	RXR-TR
5	DR5	RARE	RXR-RAR

**Table2-** Schematic representation of the 1-to-5 rule<sup>115</sup>.

## 2.4 The D Domain- Just a hinge?

The D region of nuclear receptors is less conserved than the adjacent DNA-binding and ligand-binding domains. D domain is classically believed to serve as a flexible hinge allowing the rotational flexibility of the DBD to respect to the LBD and the formation in both of the domains of dimerization interfaces. Nevertheless an increasing number of studies provide evidences for the D region implication into ligand-induced enhancement of transcription and cellular localisation.

The cellular localization of a nuclear receptor is a result of the shuttling between the nucleus and the cytoplasm. Functional nuclear localisation signals (NLS) are responsible for the shuttling of cytoplasmic proteins to the nucleus. Deletion mutants of PR were observed to display a different cellular localization thus suggesting NLS implication in nuclear receptors shuttling<sup>76, 116</sup>. Nuclear receptors hinge region also contains a lysine motif that is conserved between phylogenetically related NRs. Lysine residues within this region are acetylated in vivo in ER $\alpha$  and are selectively acetylated

by the coactivator p300 in vitro. Mutations of the ER $\alpha$  and AR conserved lysine motif have been shown to occur respectively in breast cancer and in prostate cancer<sup>179</sup>.

## 2.5 The E Domain- Ligand binding domain

### 2.5.1 The structure

The C-terminal half of nuclear receptors encompasses the Ligand Binding Domain (LBD). The LBD is a highly structured module that encodes for different functions most of which have a ligand-dependent activity. Most importantly it harbours the ligand-dependent transactivation function (AF-2), but it also contains large dimerization interfaces and cognate boxes responsible for the interaction with coregulator proteins<sup>74</sup>. Despite displaying a low sequence identity for the LBD domain (below the 30% for RAR $\gamma$  and RXR $\alpha$ ) all available structures of nuclear receptors LBDs describe a highly conserved folding<sup>25, 31, 74, 146, 177</sup> consisting in a three-layered sandwich of 11-12 antiparallel  $\alpha$ -helices (H1, H3-H12). The central layer, formed by helices (H4, H5, H6, H8, and H9), is comprised between the first layer (H1 and H3) and the third one (helices H7 and H10). H2, followed by an omega loop, establishes the connection between H1 and H3 but is not retrieved in the folding of all nuclear receptors. This is the case for RAR $\gamma$  and TR where the H1-H3 connection is made by a longer  $\Omega$ -loop. Short loops are found also in the region between H5 and H10 and at the C-terminal end in proximity of H12.

A “common signature” in the LBD of all nuclear receptors, represented by a conserved 20 amino-acid sequence (F, W) AKxxxxFxxLxxxDQxxLL, is found in the region between the end of H3 and H5.

The carboxy-terminal region of the LBD encompasses the ligand dependent transactivation function AF-2. Within this region, an activation domain AF-2 AD has been identified (corresponding to the amphipathical helix H12) which demonstrated to possess a weak autonomous transcriptional activity in the absence of the ligand and to be structurally and functionally conserved among NRs. Mutations in this region do not affect ligand or DNA binding nor dimerization but prevent ligand dependent



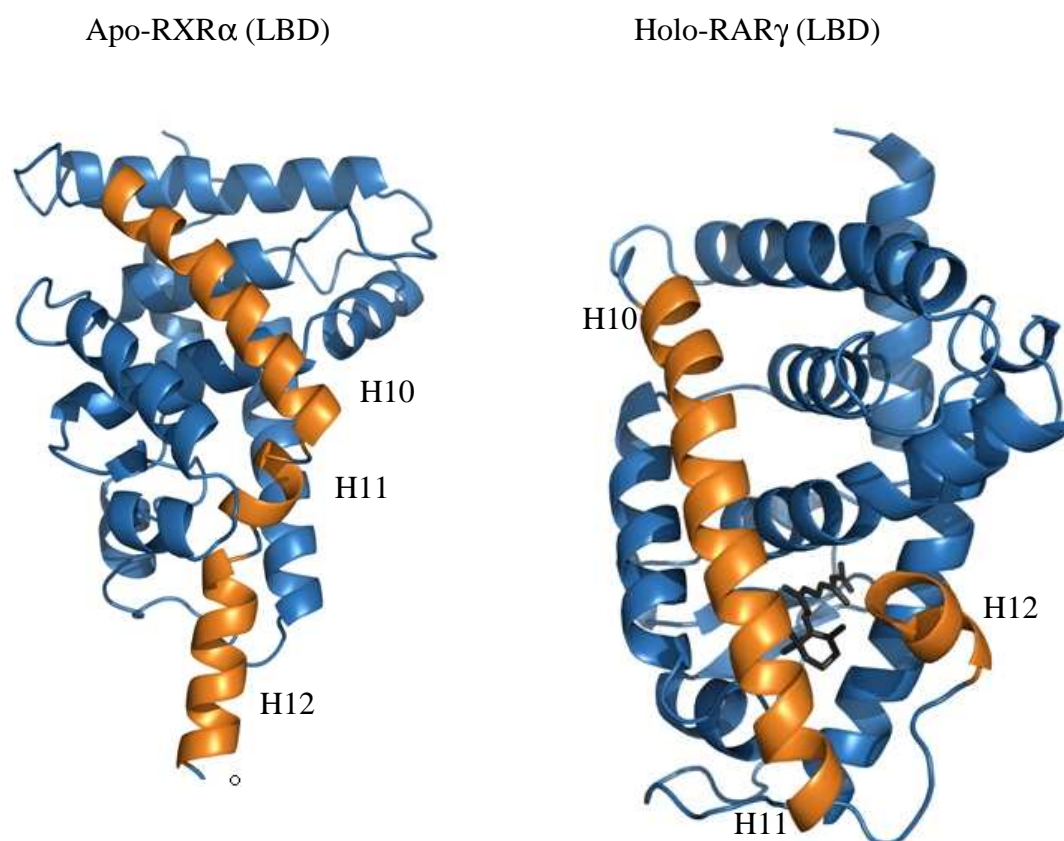
transcriptional activation of AF-2 and efficient recruitment of coactivators to the receptor<sup>44, 45, 49, 74</sup>.

The first crystal structures for unliganded (apo) RXR $\alpha$  and liganded (holo) RAR $\gamma$  LBDs revealed that hormone binding induces profound conformational changes in the receptor leading to the repositioning of H12 in the “agonist conformation”: H12 is folded up on H4 and seals the ligand binding pocket stabilizing the ligand in a “mouse-trap” model fashion (Fig.3). Repositioning of H12 generates a hydrophobic groove where coactivators can bind and activate the receptor for transcription<sup>25, 31, 71, 74, 146, 187</sup>.

### 2.5.2 The agonist conformation

Crystal structures for apo-RXR $\alpha$ , holo-TR and holo-RAR $\gamma$  LBDs reveal a common overall folding<sup>25, 51, 74, 146</sup>. Nevertheless several important differences can be pointed out when comparing the structures for the apo- to the holo- receptors. Ligand binding to the receptor triggers a drastic conformational rearrangement that leads to the alignment of H10 with H11, the flipping-up of H12 that now is tighten against H4, and the release of the L11-12 loop. According to the “mouse-trap model”, H12 in its final agonist position, seals the ligand binding cavity and orients its hydrophobic residues toward the buried ligand<sup>5, 92, 100, 122, 187</sup>.

Positioning of H12 in its agonist conformation generates the correct surface for coactivators recognition and binding and therefore is essential for nuclear receptor transcriptional activation.



**Fig.3- The mouse trap model.**

A comparison between the crystal structures of the apo-RXR $\alpha$ (LBD) and the holo-RAR $\gamma$ (LBD) shows the ligand-induced conformational change leading to the transcriptionally active form of NRs. According to this model, the ligand binding triggers a “mousetrap” mechanism in which H11 is repositioned in the continuity of H10 and the H12 folds up and seals the ligand binding cavity.

### 2.5.3 The antagonist conformation

Crystal structures of the estrogen receptor ER $\alpha$  LBD bound to raloxifen (RAL)<sup>31</sup> or to the 4-hydroxytamoxifen (OHT)<sup>160</sup>, as well as the crystal structure of RAR $\alpha$  in complex with the ligand BMS614 showed that helix H12 can adopt an “antagonist conformation” in which it is rotated by 130° and shifted by 10Å towards the receptor body to respect to the agonist conformation. Sterical hindrance of some ligands seems to induce this antagonist conformation in which H12 occupies the cleft originated by H5 and the C-terminal end of H3, a position that is reserved to the NR box peptide in the agonist bound LBD receptor<sup>31, 100</sup>. Ligands that induce non-holo H12 conformations are

“antagonists”. They prevent the formation of the NR-coactivator interface and the receptor transcription activation.

Cases of partial antagonism in which the ligand displays only a part of the “bona fide antagonist” features, have been reported. In these cases the receptor H12 adopts an antagonist position upon ligand binding nevertheless a weak transcriptional activity is detected. This is the case for RAR $\alpha$ /RXR $\alpha$ F318A bound to oleic acid<sup>26</sup>.

#### 2.5.4 The ligand binding pocket

The ligand binding pocket (LBP) is buried inside the receptor within the bottom half of the LBD. LBP is delimited by a three helices layer composed by H3, H7 and H10 and it is sealed, in the holo-receptor, by H12, at the completion of the mouse trap mechanism. Most of the residues that look onto the LBP contribute to the mainly hydrophobic environment. Only few polar residues are present at the end of the cavity and are conserved among the same

family of nuclear receptors, suggesting their role in guiding the ligand anchoring at specific positions or in conferring ligand selectivity to the pocket<sup>31, 175</sup>. As an example, an arginine residue, located in H5, is conserved among RARs, RXRs and steroid receptors. This residue establishes hydrogen bonds with the carboxyl group of retinoids in RAR and RXR or with the A-ring phenolic hydroxyl group in ER and serves as a guide for the ligand correct positioning into the cavity<sup>25, 31, 146</sup>. The crystal structures of the LBDs of multiple nuclear receptors have been solved. Information provided by the crystal structures of holo-RAR $\gamma$  LBD and apo-RXR $\alpha$  LBD revealed a common fold for the LBD that adopts an anti-parallel alpha helical "sandwich" conformation as previously described. Crystal structures for RAR $\gamma$ , TR $\beta$  and VDR LBDs show that the ligand binding pocket fits perfectly the cognate ligand shape thus offering the maximum number of hydrophobic contacts and contributing to the stability and the selective recognition of the ligand. In some cases RAR ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) only three residues are responsible for the receptors isotype-selectivity towards retinoids<sup>146</sup>. The ligand binding pocket for steroid hormones such as PPAR  $\gamma$  is much wider than the ligand and accounts for the great variety of ligands that can be bound. The ligand recognition is driven, in these cases, by specific contacts rather than a concert of hydrophobic interactions<sup>128</sup>.

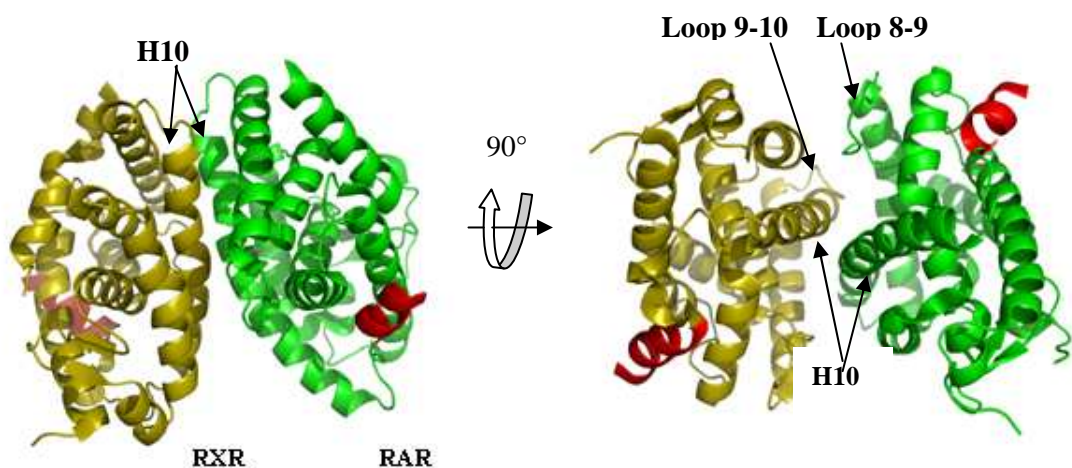
### 2.5.5 The dimerization interface

Crystal structures for LBD dimers of several NRs have recently been solved.

As predicted by early mutagenesis studies H10 plays a central role in the formation of the dimers. Homo and heterodimers display a canonical structure with H10 constituting the backbone of the interface. Further stabilization is provided by interactions involving H7, H9, and H11 as well as loops L8–9 and L9–10(Fig.3b). Due to their intrinsic asymmetry the heterodimers interface covers approximately the 10% of the entire nuclear receptors surface compared to the 15% of the homodimers<sup>26, 31, 64, 92, 128</sup>.

Structural information was integrated in an extensive sequence alignment analysis aimed at elucidating the mechanism that regulates the association of nuclear receptors as homo or heterodimers. Two classes of nuclear receptors have been identified that selectively conserve either the residue at position 93 (R93 in 91% and K93 in 5% in homodimers) or the arginine at position 62 (R in 98% in heterodimers). Homodimeric NRs belong to class 1, heterodimeric NRs belong to class 2.

Class-specific salt bridges network stabilize the receptors structures. Salt bridge connecting H8 to the loop 9-10 (C-term for class1, N-term for class 2) is conserved in both classes and confers stability to the dimer by blocking the loop in the correct conformation. Class 2 is also provided with a salt bridge connecting the conserved residue E42 (kink H4-H5) with R62 (loop 8-9). This specific interaction could explain why RXR heterodimers are more stable than the homodimers<sup>26, 28</sup>.



**Fig.3b-** Structure of the RARβ/RXRα heterodimer. (a) H10 constituting the backbone of the interface, (b) Further stabilization is provided by interactions involving loops L8–9 and L9–10.

## 2.6 The F Domain

### 2.6.1 The structure

The F domain is located at the C-terminal part of the nuclear receptor and it corresponds to the region extending from the LBD helix H12 right to the end of the receptor. Only some nuclear receptors possess this domain which displays a low degree of conservation among NR and even within the same family. Few structures have been solved for this domain so far. Crystal structure for progesterone receptor (PR) indicates that the F domain extended conformation is stabilized by interactions with residues belonging to the ligand binding domain H10<sup>184</sup>.

### 2.6.2 The function

The biological function of the F domain within NR is not well known. Nevertheless recent studies carried out on the oestrogen receptor (ER) have demonstrated that truncation of F domain does not affect transcriptional activation in response to estradiol E2. Following investigations revealed that the F domain deletion does not affect binding to the receptor of estrogens or antiestrogens nor the binding of the receptor to their DNA REs<sup>120, 134</sup>. A large F domain is also demonstrated to prevent the interaction between the nuclear receptor hepatocyte nuclear factor 4  $\alpha$ 1 (HNF4 $\alpha$ 1) and its coactivator GRIP1<sup>161</sup>. Analysis of the interaction between GR or TR $\beta$  and the p160 cofactor also suggests that their extended F domains prevent the receptors from recognizing the coactivator NR-box<sup>45</sup>.

### 3. How do NRs activate transcription?

The ability of DNA-bound NRs to interact with the basal transcriptional machinery and activate the transcription of target genes is mediated by Transcription Intermediary Factors (TIFs), namely corepressors and coactivators.

#### 3.1 Coactivators

Early squelching experiments, meant to investigate protein-protein interactions, were carried out onto nuclear receptors. These experiments demonstrated that several proteins interacted with nuclear receptors. Coactivators were identified as the common limiting factors capable of inhibiting the transcriptional activity of an agonist-bound NR when this interacts with another (or an excess of the same) NR. Corroboration of this hypothesis came from the observation that the coactivators acted in ligand-dependent manner and did not exert their activity onto transcriptional inactive NR (mutants or non holo- receptors)<sup>71</sup>.

Some of these proteins have the character of prototypical coactivators. They enhance transcription activation by scaffolding the assembly of basal transcription factors into a stable preinitiation complex whereas other coactivators modulate nuclear receptors activity by exert different functions.

#### 3.2 The coregulator boxes

The interaction between NRs and coactivators is mediated mainly by the AF-2 region of nuclear receptors<sup>10</sup>. Extensive biochemical analyses enlightened that H12 integrity is essential for transactivation activity of NRs and that any deletion or mutation in this region would abrogate transcription of target genes<sup>31, 44</sup>.

Previous mutagenesis experiments demonstrated that additional residues belonging to helices H3, H5, H6 and H11 are also involved in the recruitment and stabilization of coactivators to NRs<sup>79, 131</sup>. Subsequent crystal structure of TR $\beta$  LBD bound to the

coactivator GRIP1 revealed that in the agonist-bound receptor helix H12 and conserved residues from H3, H4, and H5 of the nuclear receptor LBD originate a hydrophobic groove that interacts with the highly conserved motif (LxxLL) of the coactivator identified as “NR box peptide”.

Two highly conserved residues (Lys 301 (H3) and a Glu 471 (H12)), at the surface of the LBD, make hydrogen bonds with Leu 1 and Leu5 from the LxxLL motif serving as a clamp that orients and locks the LxxLL peptide in the correct position and allows the hydrophobic interaction between the conserved leucines of the coactivator motif and the receptor hydrophobic groove. The interaction between amphipathical  $\alpha$ -helices LxxLL motifs and the hydrophobic groove is highly conserved among NR-coactivator suggesting a common mechanism for coactivators binding to the receptors<sup>5, 25, 31, 45, 128, 146, 160, 187</sup>, whereas sidechain sequences outside the hydrophobic groove of the NR LBD or at double ends of the LxxLL motif generate receptor and binding mode selectivity<sup>45</sup>. The coactivator SRC-1 binds to ER with a single LxxLL motif whereas combinations of two motifs are required for the activation of AR, TR, RAR, and PPAR. Co-repressors binding to nuclear receptors shows close similarities to co-activators binding fashion. Co-repressors are recognized by a short amphipathical  $\alpha$ -helical motif called “CoRNR” and mutational studies reveal that they interact with NR surfaces that are topological very similar to that recognized by NR coactivators but do not involve H12<sup>124, 132</sup>.

The discovered interaction between TIF2, a member of the p160 family, and the N-terminal region of ER and AR as well as the ligand-dependent interaction of different components of the coactivator TRAP/DRIP with both the AF-1 and AF-2 domains of GR suggested the identification of a second binding site for coactivators, located in the A/B domain of nuclear receptors. Synergy between the AF-1 and AF-2 activation domains with coactivator would result in an enhancement of the transcriptional activity of NRs<sup>14, 81</sup>. Due to the high sequence variability of the A/B domain, binding of coactivators to this region involves only glutamine-rich regions of the coactivator<sup>16</sup>.

### 3.2.1 The p160 family of Coactivators

The most abundant class of nuclear receptors-interacting proteins is composed by proteins with a molecular weight of 160 KDa, the p160 family. P160 coactivators share a high sequence similarity and possess intrinsic histone acetyltransferase activity (HAT),<sup>38, 164</sup>. HAT activity addresses the carboxy-terminal region of the coactivator and histones H3 and H4. Prototypical coactivators are molecules that help the interaction between nuclear receptor and the basal transcription machinery. It is commonly believed that recruitment of coactivators upon ligand binding to the receptor leads to histones acetylation and chromatin remodelling and allows the formation of a stable preinitiation complex that promotes in turn transcription activation.

Cloning allowed the partition of the p160 family in three distinct classes: SRC-1/NCoA-1, TIF-2/GRIP-1/NCoA-2 and p/CIP/ACTR/AIB1/TRAM1/RAC3<sup>71</sup>.

The steroid receptor coactivator SRC-1 was the first to be identified. It acts as bona fide coactivator with several nuclear receptors as GR or ER, and non-steroid group II receptors such as VDR, PPAR, TR, and RXR and exerts his function in an agonist and AF-2-dependent manner, stimulating the transcriptional activity of the receptor. The mouse homolog of SRC-1 is known as NCoA-1 exhibits a NH-terminal extension.

Immunoprecipitation experiments showed that a large part of the p160 family is constituted by SRC-2 coactivators, the TIF-2 in humans and the homologous GRIP-1 or NCoA-2 in mouse. This second class of p160 coactivators also acts prototypical coactivators enhancing the transcriptional activity of the cognate nuclear receptor. This two classes of coactivators are 30% identical overall and a high degree of homology has been observed in correspondence of three regions: the PAS domains at the N-term end of SRC-1 and TIF-2 which mediate homo and heterodimeric interactions, the central region of SRC-1 and TIF-2 containing a high percentage of serine and threonine residues and the C-term end has been found to interact with PR in the case of SRC-1.

The third class of p160 proteins, named SRC-3, is composed by ACTR, AIB1, RAC3, or TRAM-1in humans and p/CIP in mice. The members of the third family are very similar to the coactivators from the first two subfamilies but additionally they act as



bona fide coactivators for several different transcription factors including activators of transcription (STAT-1) and the cAMP response element binding protein (CREB).

### 3.2.2 CBP/p300 Cointegrators

CBP and p300 are large proteins that were previously discovered to work as essential coactivators for the transcription by AP-1 proteins. Subsequently it was found that CBP and p300 also act as coactivators for many classes of transcription factors as ER, RAR, RXR and TR<sup>5</sup>. The strong implication of CBP and p300 in nuclear receptor transcription activation is witnessed by several experimental results. Overexpression of these coactivators enhances the transcriptional activation of the cognate nuclear receptor and retrieves the inhibition of GR and RAR onto AP-1 activity as the result of the competition between AP-1 and the nuclear receptor for CBP/ p300.

As CBP/p300 binds nuclear receptors at the C-terminus and p160 coactivators at the N-terminus, ternary complexes can be formed between coactivators and nuclear receptor as in the case of CBP and SRC-1 which interact together in the activation of PR and ER. CBP and p300 display acetyltransferase activity on free and nucleosomal histones. The removal of the histone acetyltransferase activity (HAT) domain from CBP and p300 results in the loss of the transcriptional activity for many transcription factors<sup>5, 38, 71, 164</sup>.

### 3.2.3 TRAP and DRIP

TRAP and DRIP are multiprotein complexes containing 14-16 proteins and ranging from 70 to 230 KDa in size. Binding of these coactivators as a single unit (DRIP205/TRAP220) occurs at the holo-receptor AF-2 domain and is mediated by a LxxLL motif, identical to that found in p160 coactivators<sup>5</sup>.

Different components of the DRIP/TRAP complex interact in a hormone dependent fashion with both AF-1 and AF-2 domains of GR. DRIP150 binds to AF-1 surface whereas DRIP205 interacts with the AF-2 domain. Transcription activity of the receptor is enhanced by the cooperation of the two coactivators suggesting a functional link between GR activation domains.

The complex DRIP/TRAP does not display HAT activity; nevertheless it has been shown to enhance VDR and TR transcriptional activity in a cell free system containing chromatin. As DRIP/TRAP components are found in the SMACC mediator complex that interacts with RNA Pol II together with SRB proteins, it is believed that DRIP/TRAP work to recruit RNA Pol II to the target promoter<sup>58, 141</sup>.

### 3.2.4 Corepressors

Mutagenesis experiments performed onto TR $\beta$  LBD identified a domain called “CoR” box, encompassing H1, which is essential for binding of corepressors to the receptor. Some of the residues in this box are well conserved among nuclear receptors and mutation of three of them in particular (AHT) resulted in the abrogation of active repression<sup>83</sup>. All available crystal structures showed indeed that the conserved residues do not make direct contacts with the corepressor suggesting that the AHT mutation more likely affects the dissociation of the H1 from the LBD core<sup>122</sup>. The first corepressors to be identified were the silencing mediator of retinoic and thyroid hormone receptor (SMRT)<sup>39</sup> and the nuclear receptor corepressor NCoR (RIP-3)<sup>83, 191</sup>. NCoR and SMRT interact with a variety of receptors such as TR, RAR and VDR. A couple of novel corepressors such as Alien and Hairless have more recently been isolated. Alien interacts with receptors as VDR, TR but not with retinoic acid receptors. Hairless, a corepressor required for the hair cycling, shows to interact with both TR and VDR<sup>90, 138, 139</sup>.

Structural analysis of the first corepressors leads to the identification of N-terminal repression domains (RDs). RDs contain the highly conserved “CoRNR” consensus-sequence (LXXI/HIXXXI/L) and are responsible for the interaction of the corepressor with several HDACs and the activation of gene silencing. C-terminal conserved receptor interaction domains (IDs) have been discovered to be responsible for the interaction of the corepressor with unliganded receptors<sup>5</sup>.

Both corepressor and coactivator binding motifs have an amphipathical  $\alpha$ -helical character. In addition mutation in H3, H4, H5, involved in the coactivator binding site, are shown to affect corepressor binding to the receptor. All these aspects suggest that

corepressors and coactivators bind to overlapping motifs as confirmed by the superposition of the agonist bound and antagonist bound LBD of PPAR $\alpha$ <sup>124, 132, 188</sup>.

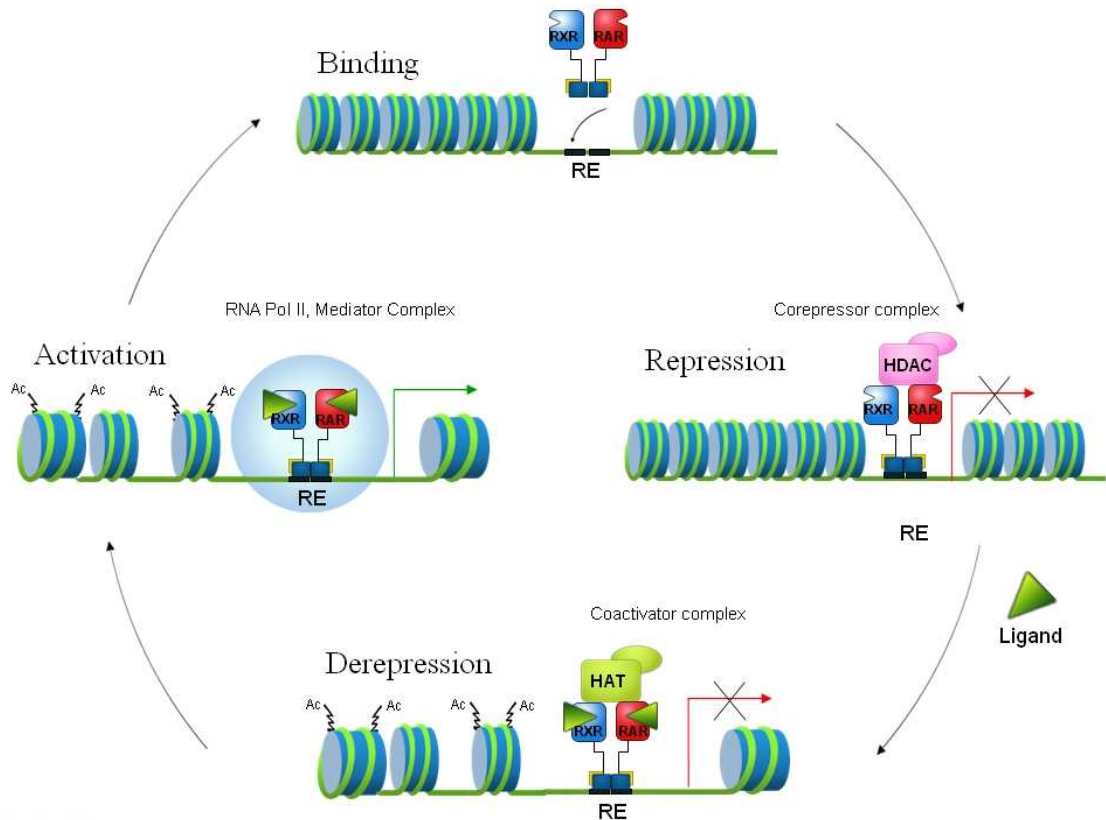
It has been reported that integrity of H12 is inhibitory for corepressor binding and that ligand binding alone is not sufficient to induce the dissociation of the corepressors from the receptor.

### 3.2.5 Transcription activation

Activation of target genes transcription can be schematically described as an hypothetical three-step mechanism (Fig.4). Some receptors as RAR and TR can active repress transcription of target genes in the absence of the ligand<sup>9, 34</sup>. This “repression” step represents the first step in the mechanism that regulates nuclear receptor mediated transcription. In the absence of ligand apo-NRs are bound to HRE and are associated with NCoR/SMRT corepressors possessing intrinsic or acquired histone deacetylase activity (HDAC)<sup>39, 97</sup>. Deacetylated histones are associated with silent regions of the genome leading to gene silencing. In a second “derepression” step the ligand binds to the receptor and induces major conformational changes leading to corepressors dissociation and recruitment of coactivators belonging to the family of TIF2, SRC-1, RAC3 and CBP, p300 cointegrators. Coactivators display histone acetylase activity (HAT) and acetylate specific residues in the N-terminal tails of different histones leading to chromatin decondensation. The p160 coactivators may rather recruit histone deacetylase activity by physical association with histone acetyl transferases, such as CBP or p300, or with complexes containing such activities. Specifically, the activation domain AD1 of TIF2 has been demonstrated to function via the recruitment of cointegrator CBP<sup>176</sup>, which in turn acetylates TIF2. Besides HAT activities, NR coactivator complexes possess other enzymatic activities. TIF2 proteins are able of interacting functionally via their activation domain AD-2 with a protein methyltransferase<sup>38</sup>.

Once the chromatin environment at the promoter of target genes has been decondensed by coactivator complexes containing members of the TIF2 and CBP families, the nuclear receptor recruits the RNA Pol II holoenzyme via association with the TRAP220/DRIP205 subunit, belonging to the large multisubunit protein complex

known as mediator complex SMCC. The switch between coactivators and the SMCC complex might be regulated by the acetylation of coactivators within the HAT complex, resulting in their dissociation from the nuclear receptor and in the recruitment of factors such as SMCC, via the LxxLL motif of the TRAP220/DRIP205 subunit<sup>72, 100</sup>



**Fig.4- Schematic mechanism by which NRs activate the transcription of target genes.** In the absence of the ligand RAR/RXR heterodimers, located in the cell nucleus, are bound at the RE of the target genes, in complex with co-repressors possessing histone deacetylase activity. Histone deacetylation leads to DNA condensation and repression of transcription. Binding of ligand induces the release of the HDAC complex and induces the recruitment of co-activators possessing histone acetyltransferase activity (HAT). The subsequent chromatin decondensation due to histone acetylation leads to the recruitment of the RNA Polymerase II holoenzyme and the Mediator complexes (MEDs) and leads to transcription activation.

## 4. Posttranslational modifications on NRs

Nuclear Receptors are targets for multiple posttranslational modifications such as acetylation, ubiquitylation and phosphorylation<sup>55, 61, 179</sup> which act in concert to regulate NR-mediated transcription.

Nuclear receptors such as AR and ER $\alpha$  are target for histone acetyltransferase in vitro and in vivo. The ER $\alpha$  acetylation motif is conserved between different species including vertebrates, arthropods, and nematodes and between phylogenetically related nuclear receptors such as TR, RAR, PPAR, VDR, GR, PR, HNF4 and SF1, suggesting that direct acetylation of nuclear receptors may participate to additional signalling pathways<sup>60, 62, 179</sup>.

Transcription regulation of NRs appears to be coupled with degradation of these proteins by the ubiquitin–proteasome pathway<sup>18, 22, 43, 55</sup>. Ubiquitination of steroid receptors can depend on the ligand binding to the receptor as for MR<sup>190</sup>, ER ( $\alpha$ ,  $\beta$ )<sup>33</sup>, GR<sup>178</sup> and PPAR $\alpha$ <sup>18</sup> or act constitutively as for PR and AR. Ubiquitylation of PR depends on the phosphorylation of the residue S294<sup>140</sup> whereas for AR poly-ubiquitylation impairs the activity of the receptor by directing it to the proteasome<sup>109</sup> and monoubiquitination of the receptor leads to AR stabilisation and transcriptional activity enhancement<sup>33</sup>.

Recent investigations showed that nuclear receptors are also target for methylation. ER $\alpha$  is methylated in the hinge region by Set7 at the consensus recognition sequence [R/K][S/T]K<sup>42</sup>, which displays high similarity with the conserved acetylation motif found in nuclear hormone receptors<sup>60</sup>. ER $\alpha$  is also methylated at the DNA-binding domain at residue R260 by PRMT1. Methylation of ER stabilises the receptor allowing its efficient recruitment to target genes<sup>166</sup>. The hepatocyte nuclear factor 4 is also methylated by PRMT1 at the DNA-binding domain leading to enhanced binding to target genes<sup>11</sup>. Recent studies enlightened that mouse RAR $\alpha$  trimethylated at Lys347 located in the ligand binding domain (LBD) leading to enhanced interactions with coactivators as well as its heterodimeric partner RXR<sup>85</sup>.

## 4.1 Nuclear Receptors Phosphorylation

Phosphorylation has been the most extensively studied posttranslational modification due to its established participation to the signalling cross-talk of almost all NRs.

Both the N-terminal and C-terminal regions of NRs contain several consensus sites for phosphorylation by different kinases such as cyclin-dependent kinases (CDKs), mitogen activated protein kinases (MPKs), and Jun N-terminal kinases (JNKs).

There is increasing evidence that phosphorylation can finely tune nuclear receptors activity by disrupting the complex with other components of the transcription machinery, weakening the affinity for the ligand binding domain or activating the receptor degradation thus reducing or abolishing the receptor response to the ligand. Nevertheless the exact mechanism through which phosphorylation would regulate nuclear receptors' activity is still unclear<sup>23, 148, 149</sup>.

Moreover there is increasing evidence indicating that NRs phosphorylation is determinant in the development of several types of cancer. Most of these cancers such as breast, ovarian and prostate cancers are characterized by up-regulated kinase activity and ligand-independent transcription activation of target genes. Hyperactivity of MAPKs leads to the ligand-independent phosphorylation of ER $\alpha$  and AR and to anomalous growth of breast and prostate cells. Hyper-phosphorylation of RXR prevents transcription of genes responsible for the control of cell growth and it is probably linked to development of hepatocellular carcinoma. Finally hypophosphorylation of RARs by the cyclin-dependent kinase cdk7 results in a deficient response to retinoic acid and leads to a severe genetic disease known as Xeroderma Pigmentosum characterized by photosensitivity, premature skin aging and skin tumours development<sup>23</sup>. Up to date no structure is available for a phosphorylated receptor<sup>19, 150</sup>.

Most of nuclear receptor phosphorylation sites are localized within the N-terminal region. The number of phosphorylation sites is highly variable depending on the type of nuclear receptor. As an example, Progesterone receptor (PR) contains up to 13 phosphorylation sites in the N-terminal A/B region, Retinoic acid receptors (RARs) and PPARs contain one or two phosphorylation sites whereas VDR represents an exception and is not phosphorylated at his NTD. Most of the phosphorylation sites are constituted by serine and proline-rich motifs, corresponding to consensus sites for proline

dependent kinases such as cyclin-dependent kinases (CDKs) and MAP kinases (Erks, JNKs and p38MAPK). Phosphorylation at the N-term can occur in response to hormone binding or in the absence of ligand as the response to a variety of signals. As an example, unliganded estrogen receptor alpha (ER $\alpha$ ) is phosphorylated at Ser 118, located in AF-1 by the mitogen-activated protein kinases (MAPK). This phosphorylation affects the receptor ability to dimerize, bind the ligand and the DNA, interact with coregulators and finally activate the transcription of target genes<sup>32, 155</sup>. The AF-1 domain of PR, PPARs, and RARs has also been reported to be target for phosphorylation by MAPKs. In particular phosphorylation of AF-1 of RAR $\gamma$ 2 is required for RA-induced differentiation into primitive endoderm, whereas phosphorylation of AF-1 and AF-2 in RAR $\alpha$ 1 and RAR $\gamma$  is required for differentiation into parietal endoderm<sup>148, 169</sup>. RXR $\alpha$  is targeted by c-Jun N-terminal kinases (JNKs) which belong to the MAPKs family and ER $\alpha$  and AR are phosphorylated at their N-terminal region by Akt or PKB kinases whose activation is determinant in cell survival and proliferation<sup>27, 149</sup>. Ligand-bound NRs including ER $\alpha$ , RAR $\alpha$ , RAR $\gamma$  and AR have been reported to interact with the general transcription factor TFIID, a multiprotein complex that acts as mediating transcription activator. TFIID is composed of nine subunits, including cdk7 which displays a cyclin-dependent kinase activity. As the consequence of the interaction with TFIID, ER $\alpha$  and RAR ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) are phosphorylated at their N-terminal region by TFIID subunit cdk7<sup>13, 37, 86, 149</sup>. In most cases (i.e. ER $\alpha$ , PPAR $\alpha$  and AR) the phosphorylation at the N-terminal region helps the assembly of a functional transcription machinery by favouring the recruitment of coactivators, chromatin remodellers, and general transcription factors that enhance the response to the ligand and consequently the expression of retinoid target genes<sup>149</sup>. Nevertheless there are cases in which phosphorylation of the N-terminal region has been correlated with the inhibition of the transcriptional activity of the receptor, as it is the case for VDR, or with the promotion of NRs degradation by the ubiquitin proteasome pathway as it is the case for GR, PPAR $\gamma$ , RAR<sup>20, 23, 30, 32, 68, 69, 159</sup>. In the latter case nuclear receptor phosphorylation would allow tuning the length and the intensity of the receptor response to the ligand.

Nuclear receptor LBD is also a target for phosphorylation. Phosphorylation of this domain involves several kinases. Tyrosine kinases are responsible for the

phosphorylation of ER $\alpha$  and RXR $\alpha$  whereas cAMP-dependent protein kinase A (PKA) and MSK-1 share the S369 phosphorylation site of RAR $\alpha$ . The cAMP-dependent protein kinase A (PKA) is also responsible for RAR $\gamma$  phosphorylation at S371<sup>30, 148, 149</sup>. Phosphorylation at the C-term domain is most often directly involved in the enhancement of the receptors transcriptional activity as in the case of ER and PKA. ER $\alpha$  phosphorylation at tyrosine 537 by Src kinases enhances ER $\alpha$  transactivation probably due to conformational changes that modify ligand binding properties. RAR $\alpha$  phosphorylation by PKA also enhances the transcriptional activity of the receptor and has been shown to favour receptor heterodimerization and coregulators binding. Phosphorylation of coregulators is involved in the crosstalk between NRs and other signalling pathways. It has been demonstrated that phosphorylation of coactivators such as p300/CBP<sup>149</sup> and SRC-1<sup>151</sup> favourites their interaction with the liganded receptor and the recruitment of chromatin modifying complexes. Phosphorylation of corepressors, on the other hand, has been related to translocation of the receptor out of the nucleus and inhibition of transcriptional activity<sup>82</sup>.

## 5. Cellular Localisation

NRs are classically divided in two classes depending on their localization within different subcellular compartments.

The “translocating” receptors, as most of the steroid hormones (AR, GR, PR, MR but not ER), are predominantly found in the cytoplasm in the absence of ligand and they move to the nucleus upon cell stimulation by the hormone. In the cytoplasm and in the absence of hormone stimulation, the receptor is inactive and it forms a large complex with chaperones proteins, notably members of the heat shock protein Hsp90 machinery, which appear to help the correct folding of the receptor LBD<sup>77</sup>. Binding of the cognate ligand to the receptor induces dissociation of the receptor-chaperone complex and exposes the receptor nuclear translocation signals (NLS). Interaction between NLS and the nuclear pore complex would determine the subsequent translocation of the receptor



to the nucleus. The “Nuclear” receptors as RAR, VDR and TR are mostly found in the nucleus either in presence or absence of the ligand.

Recently several exceptions to this widely accepted model have been documented. The steroid receptor ER $\alpha$  exchanges its Hsp-complexed and free form in a ligand dependent manner, nevertheless its presence is confined to the cell nucleus<sup>77, 136</sup>. TR is as a steroid receptor and its constant presence in the nucleus was always explained by the receptor ligand-independent ability to bind DNA. Recent experiments in living cells demonstrated indeed that a large amount of the total TR present in the cell is found in the cytoplasm<sup>77, 129</sup>. Finally several receptors such as PR, TR and GR receptors have been shown to gather in “accumulation points” within the nucleus upon cell stimulation with hormones. All these results suggest that nuclear receptors localization within cell does not simply depend on the receptor NLS recognition or the ligand-independent capacity of the receptor to bind to DNA but is rather a dynamical and multifactor process.

## **Chapter 2 : Retinoic Acid Receptors**

## 6. Retinoids

Vitamin A and derivatives are commonly referred to as retinoids.

Retinoids are important signalling molecules that explicate their function through the binding to specific nuclear receptors (RARs and RXRs) and their activation for transcription of retinoid target genes. Retinoids are essential for the life of all organisms. They control vital processes such as cell growth, differentiation and death and regulate functions such as vision, reproduction, embryonic growth and development.

Vitamin A deficiency (VAD) has been correlated to severe ocular features known as Xerophthalmia and reduced resistance to infections. Excess of vitamin A also has deleterious effects on the organisms and affects skin, bones, nervous and circulatory system and embryonic development<sup>152, 165</sup>.

Several studies based on animal models demonstrated that retinoids are efficient chemotherapeutic agents. Retinoids and especially retinoic acid compounds are currently used in the treatment of acute promyelocytic leukaemia (APL), several types of cancers and precancerous lesions such as oral leukoplakia, cervical dysplasia and xeroderma pigmentosum (an inherited predisposition to ultraviolet-induced cancers)<sup>3, 96, 121, 165</sup>.

### 6.1 The chemistry of retinoids

According to the IUPAC definition the term Vitamin A encompasses all compounds displaying the biological activity of retinol.

Long controversies have characterized the definition of the term “retinoids”. The currently accepted definition is an integration of the original IUPAC definition and considers the class of retinoids constituted by “*retinol analogs (with or without biologic activity) but also of several compounds that are not closely related to retinol but elicit biological vitamin A or retinoid activity*”.

Retinoids are derivatives of the primary alcohol all-trans retinol. All-trans retinol is a lipophilic ligand that is carried in plasma by binding proteins but it possesses a partial

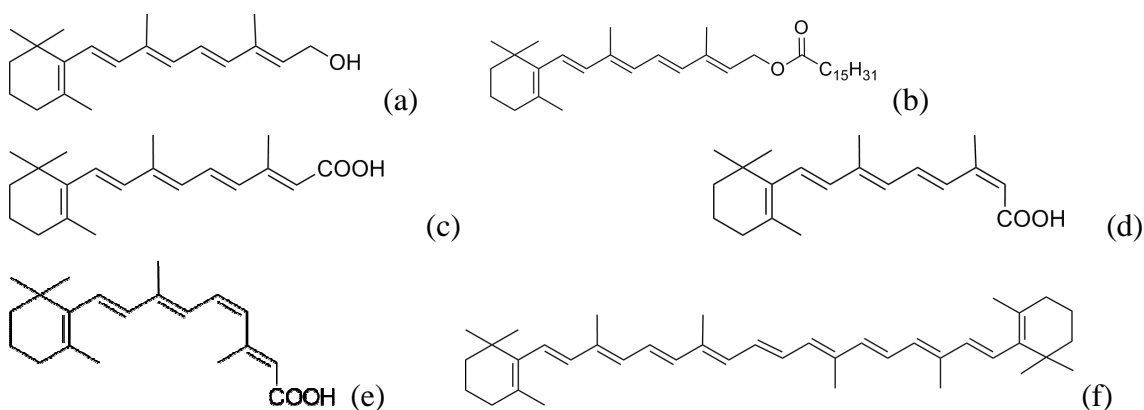
hydrophilic character. These features allow all-trans to efficiently diffuse through hydrophilic cellular fluids as well as hydrophobic membranes<sup>19</sup>. Retinoids are mostly present in animal tissues as retinyl palmitate or other fatty acid esters such as retinyl stearate or retinyl oleate and occur predominantly in the all-trans configuration (Fig.5). The role of all these retinoids is not known yet but they are likely to represent intermediates in the synthesis or degradation of retinol.

Retinoids are unstable in presence of oxygen and light. Exposure to either of the two elements leads respectively to degradation or isomerization. For these reason retinoid solutions must be handled in inert atmosphere and must be sheltered from light <sup>75</sup>.

## 6.2 Vitamin A metabolism

Animals cannot synthesize vitamin A. They guarantee the appropriate uptake of vitamin A by dietary assumption of carotenoids, as the product of the metabolism of fungi, algae, micro organisms and bacteria, subsequently converting them into retinal. As an alternative, animals can obtain retinoids directly from other animals that have already carried out the conversion into retinol or retinyl esters. Retinyl esters and retinol accumulate especially in fish, mammalian liver, eggs yolk and milk.

Carotenoids are absorbed in the small intestine cells (enterocytes) by passive diffusion. Here  $\beta$ -carotene is enzymatically converted into retinal which is in turn reduced to retinol<sup>107</sup>. Dietary retinyl esters instead are enzymatically reduced to retinol before being absorbed by the enterocytes. In the enterocytes retinol is bound to a protein called cellular retinol-binding protein type II (CRBP-II). This protein solubilizes retinol and promotes the enzymatic re-esterification of the large majority of the retinol to retinyl esters. Retinyl esters are stored and secreted into the intestinal lymph clustered into large lipoprotein complexes called chylomicrons<sup>21</sup>. In this form retinyl esters and unesterified retinol are uptaken into the hepatocytes and stored in the so-called stellate cells. Retinol is mobilized by these cells in a controlled manner and oxidized to all-trans retinoic acid which has a particular relevance as it activates the RAR and RXR transcription factors. Other metabolise of Vitamin A include 9-cis retinoic acid, 13-cis retinoic acid, 13-cis-4-oxo retinoic acid, all-trans-4-oxo retinoic acid, all present in the plasma<sup>126, 145</sup>.



**Fig.5- Some of the retinoids present in human body:** (a) all-trans retinol, (b) retinyl palmitate, (c) all-trans retinoic acid, (d) 13-cis retinoic acid, (e) 9-cis retinoic acid, (f) β-carotene

### 6.3 Retinoic acids

Retinoic acid (RA) is a low molecular weight molecule of 300 Da.

Despite the predominant hydrophobic character retinoic acid is partially soluble in intra and extra-cellular fluids. This amphipathical nature allows retinoic acid to rapidly diffuse and renders it a prototypical morphogen that is synthesized in localized centres but then explicates its function at a distance. Retinoic acid plays a central role in a multitude of physiological mechanism such as cellular differentiation and reproduction, development of the central nervous system<sup>113</sup> and haematopoiesis<sup>41</sup>, and is implicated in pathological conditions such as cancer<sup>127, 174</sup> and skin diseases<sup>56, 95, 189</sup>. The simplest mechanism leading to 9-cis-retinoic acid biosynthesis consists in the isomerization of all-trans retinoic acid. Despite having been accepted as Vitamin A pathway signalling molecule, there is no experimental evidence for 9-cis retinoic acid to be an endogenous compound.

Endogenous characterization of this isomer should include multiple techniques such as MS, NMR and UV spectroscopy and should take in account artefacts due to the isomerization of all-trans retinoic acid to the cis-isomer induced by day light and thiol-containing compounds such albumin and GSH.

In one case the presence of 9-cis retinoic acid has been detected in the human body and that is following administration of a 100 fold recommended dose of retinol. These results do not prove the existence of endogenous retinol but could suggest that all-trans retinoic acid is the natural activator of the retinoid receptors while 9-cis retinoic acid would be produced in response to a massive administration of vitamin A and would account for vitamin A toxicity and aberrant RAR and RXR gene activation<sup>3, 19</sup>.

## 6.4 Retinoids acid in cancer

Strong experimental evidences show that retinoids prevent cancer by inhibiting progression of preneoplastic lesions to malignant stages<sup>4, 40, 53, 162, 163</sup>. Experimental studies based on animal models demonstrated that they possess a wide range of clinical applications and display different pharmacokinetic and pharmaceutical features.

Isotretinoin (13-cis retinoic acid) induces the differentiation and apoptosis of tumour cells and inhibits the progression of pre-malignant lesions<sup>102</sup>. Isotretinoin finds numerous clinical applications and is used in combination with interferon alpha (IFN $\alpha$ ) to treat patients with pre-malignant diseases such as leukoplakia, cervical intraepithelial neoplasia and xeroderma pigmentosum as well as to arrest advanced skin and cervix carcinomas<sup>46, 95, 119</sup>.

ATRA or Isotretinoin (All-trans retinoic acid) binds to RXR receptors. This retinoid is currently used to treat acute promyelocytic leukaemia (APL) in which myelopoiesis is arrested at the promyelocytic stage. In APL a chromosomal translocation is responsible for the formation of the oncofusion protein PML–RAR $\alpha$  resulting in the deregulation of RAR $\alpha$  gene. Retinoic acid combined with chemotherapy restores the normal functionality of RAR $\alpha$  and reactivates all the functions mediated by the retinoid receptor such as the induction of cell differentiation and the control of cell growth<sup>40</sup>.

Alitretinoin (9-cis-retinoic acid) activates both RAR and RXR receptors. Experiments based on animal models demonstrated that 9-cis retinoic acid has a more powerful antitumoral activity than the all-trans isomer. It strongly promotes cell differentiation and suppresses the progression of breast tumours in several animal models<sup>73, 186</sup>. In transgenic mice expressing the oncoprotein Hras1 and treated with the tumour promoter PHORBOL ESTER 12,13- tetradecanoyl phorbol acetate (TPA), 9-cis retinoic acid

demonstrated to prevent the progression of the early stage tumours<sup>3</sup>. Despite experimental evidence showing that the benefits from retinoids are reversible upon suspension of the treatment, clinical administration of retinoids can occur only for short periods of time due to their high level of toxicity. Despite all-trans retinoic acid activates only RAR, 9-cis retinoic acid binds to both RAR and RXR and 13-cis retinoic acid binds to the receptors with very low affinity, the three retinoic acids interconvert in the human body via isomerization and act as non selective (pan agonists) activators for the RAR-RXR heterodimer, incrementing retinoids toxicity<sup>46</sup>. For this reason development of receptor-selective and isotype-selective synthetic retinoids is a major priority in retinoids pharmaceutical industry for clinical applications. More selective retinoids with lower toxicity and fewer side effects have recently been discovered. As an example LGD1069 (bexarotene) is the first synthetic retinoid displaying highly selectivity for retinoid-X-nuclear receptor to be authorized in the treatment of cutaneous T-cell lymphoma. LGD1069 represents a valuable substitution to the classical combinatorial treatment based on 13-cis and all-trans retinoic acid which demonstrated to be less effective and more toxic. LGD 1069 is now in Phase II clinical trials for the treatment of non- small-cell lung cancer<sup>93, 180</sup>.

## 7. Retinoic Acid Receptor (RAR) and Retinoid X Receptor (RXR)

### 7.1 RARs and RXRs transduce the signal of RA

Retinoids explicate their biological function through binding to specific families of nuclear receptors: the RARs and RXRs. Each family is composed by three genetic isotypes ( $\alpha$ ,  $\beta$ ,  $\gamma$ ).

Each RAR and RXR isotype has two major isophorms originated from the use of different promoters or alternative splicing and that differ in their N-term regions. The high degree of sequence homology between the isophorms of each isotype suggests that each isophorm has specific function<sup>105</sup>.

The retinoid acid receptor RAR binds both all-trans RA and 9-cis RA with K<sub>d</sub> values in the 0.2-0.7 nM range<sup>1</sup> and is transcriptionally active in the form of RAR-RXR heterodimers. RXR binds 9-cis RA with K<sub>d</sub> value of 10 nM range but display very low affinity for all-trans RA<sup>1, 80, 106</sup>.

## 7.2 Retinoid response elements

According to the 1-to-5 rule<sup>115</sup> RAR, in the form of RXR-RAR heterodimer, binds to DR5 response elements. In these elements RXR occupies the 5' half-site whereas RAR occupies the 3' one<sup>97, 114</sup>. The selectivity of the binding to the DR5 elements is determined by the interaction between the D-box in the second zinc finger of RXR and the tip of the second zinc finger of RAR<sup>193</sup>. RXR-RAR heterodimers also bind DR1 and DR2 response elements. In DR1 elements the polarity of the heterodimers is inverted (5'-RAR-RXR-3'). This is thought to account for the lack of response to retinoids characteristic of RXR-RAR heterodimers bound to DR1 elements<sup>97</sup>. In RXR-RAR heterodimers bound to DR2 element the RAR T-box forms a surface that interact with and the RXR second zinc-finger<sup>192</sup>.

## 7.3 RAR posttranslational modifications

### 7.3.1 RAR posttranslational modifications

Recent studies performed through mass spectrometry revealed that mouse RAR $\alpha$  is target for methylation at lysine residues 347, 109 and 171 located respectively in the receptor LBD, DBD and hinge region. Trimethylation of RAR $\alpha$  at Lys347 enhances the LBD sensitivity to retinoic acid, helps the recruitment of the cofactors p300/CREB and favours the interaction with the heterodimeric partner RXR<sup>84, 85</sup>. Methylation at Lys109 involves the homologous sequence (CEGCKGFFRRS) which is conserved within the NR super family and different species. Recent studies enlightened that methylation at Lys109 enhances the full length receptor response to retinoic acid<sup>84</sup>.



RAR is target for ubiquitylation within the heterodimer RAR/RXR. This process leads to the both partners degradation by the proteasome in response to retinoids<sup>22, 68, 69, 94, 168, 195</sup>. Nevertheless ubiquitylation of RAR $\gamma$  within RAR $\gamma$ /RXR revealed that proteasome ubiquitylation could play a double role, controlling the recruitment of a functional transcription machinery on one hand<sup>109</sup> and leading to the receptors degradation on the other hand. On the basis of this dual function the proteasome ubiquitylation has been purposed to regulate the functionality of RAR $\gamma$ /RXR heterodimers, as recently demonstrated for other nuclear receptors such as the estrogen and androgen receptors<sup>59, 144</sup>.

### 7.3.2 RAR Phosphorylation

The N-terminal domain of RARs is target for phosphorylation by the cyclin-dependent cdk7 kinase. This kinase is a subunit of CAK, a ternary complex composed by cdk7, cyclin H and MAT1 and belonging to the general transcription factor TFIIH. Cdk7 phosphorylates RAR $\alpha$  and RAR $\gamma$  respectively at S77 and S79<sup>150</sup>. Recent studies demonstrated that the correct positioning of cdk7 and the efficiency of the receptors phosphorylation by cdk7 is controlled by the docking of cyclin H at a specific region of RARs LBD, corresponding to the loop 8–9 and the N terminus of helix 9 (aa 339–368 in RAR $\gamma$ )<sup>24, 63</sup>. The presence of a specific binding site for cycH is thought to elicit the specificity of cdk7 kinase interaction with the N-terminal proline rich region of NRs that could in principle be phosphorylated by any proline-tyrosine kinase. The importance of the TFIIH mediated phosphorylation at the N-term domain of RARs has been enlightened by studies performed onto Xeroderma Pigmentosum (XP) patients suffering from an inherited predisposition to develop UV-induced skin-cancers. XP is a rare genetic disorder in which inherited mutations in the XPD subunit of the general transcription and repair factor TFIIH prevent the correct positioning of the cycH thus leading to hypophosphorylation of RAR $\alpha$  NTD by cdk7 and to the receptor deficient response to RA.

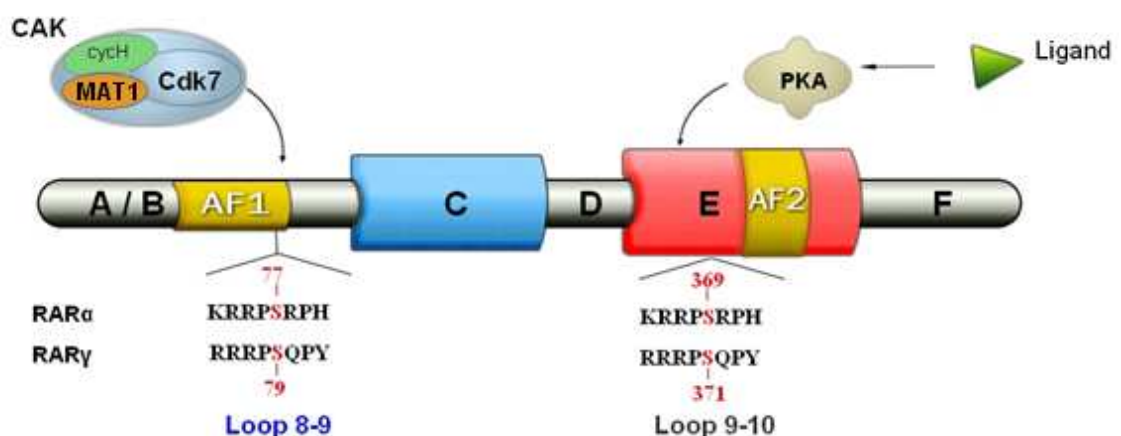
RARs are also target for phosphorylation by the cAMP-dependent protein kinase (PKA) which phosphorylates RAR $\alpha$  and RAR $\gamma$  respectively at S369 and S371 (Fig.6). These

serine residues are located in the loop 9-10 of the LBD, in close vicinity to the cyclin H-docking site<sup>63</sup>.

Gaillard and al. recently demonstrated that phosphorylation of these residues by PKA enhances cych binding to the receptor. Consequently, both the phosphorylation of NTD by Cdk7 and the transcription of RA-target genes are enhanced<sup>63</sup>. According to this novel molecular pathway phosphorylation by PKA would propagate c-AMP signal from the AF-2 domain to the AF-1 domain and would represent the first example of allosteric controlled communication between two domains of the same receptor.

RAR $\gamma$  (but not RAR $\alpha$ ) phosphorylation at the NTD affects its interaction with the actin-binding protein Vinexin $\beta$  whose function is still not fully understood. Vinexin $\beta$  is a protein containing SH3 motifs and it is thought to act as a corepressor for RAR $\gamma$ . In the absence of the ligand Vinexin $\beta$  and RAR $\gamma$  are both bound to the promoter region of RAR $\gamma$  target genes and RAR $\gamma$  is not phosphorylated at the NTD. In response to the ligand, TFIIH is recruited at the receptor surface and Vinexin $\beta$  is released allowing transcription initiation<sup>23</sup>.

In presence of the ligand the N-term domain of RAR $\gamma$  is phosphorylated by the p38MAPK<sup>30</sup> whereas RAR $\alpha$  is not directly targeted by this kinase which phosphorylates instead the p160 coactivator SRC-3. Phosphorylation by p38MAPK marks the receptor for ubiquitynilation and proteosomal degradation providing a switch to the receptor transcriptional activation.



**Fig.6- RAR is target for phosphorylation at different domains.**

The N-ter domain of RAR is phosphorylated by the cyclin-dependent kinase cdk7. This kinase is a subunit of CAK, a ternary complex composed by cdk7, cyclin H and MAT1 and belonging to the general transcription factor TFIIH. Cdk7 phosphorylates RAR $\alpha$  and RAR $\gamma$  at S77 and S79 respectively. RAR is

also target for phosphorylation by the cAMP-dependent protein kinase (PKA) which phosphorylates RAR $\alpha$  and RAR $\gamma$  at S369 and S371 respectively.

## 8. Our project

This work aims at providing new structural insights to understand at atomic level how phosphorylation by PKA participates in nuclear receptor signalling. More specifically we will investigate the impact of the phosphorylation at the serine residue located in the loop 9-10 of the LBD on the structure of the whole RAR ligand binding domain and onto hCycH binding to the receptor. To this respect we will integrate X-ray Crystallography and Nuclear Magnetic Resonance spectroscopy (NMR) studies. Both these techniques can provide structure of macromolecules such as proteins and nucleic acids at atomic resolution and play a key role in structural biology for a detailed understanding of molecular functions. It is important to underline that X-ray Crystallography and Nuclear Magnetic Resonance spectroscopy are not competing techniques, they rather complement each other.

X-ray crystallography requires single crystals and can be applied to very large macromolecular complexes (>100KDa), as long as suitable crystals are available.

The most serious limitation of X-ray crystallography is given by its restriction to static structures. In order to acquire information about the dynamics of the molecule of interest it is therefore essential to combine X-ray crystallography with other techniques capable of providing such information.

Nuclear magnetic resonance (NMR) represents the most suitable technique to obtain information concerning molecular dynamics at atomic level. NMR measurements are performed under conditions that can be as close as possible to the physiological conditions. NMR measurements not only provide structural data but can also supply information on the internal mobility of proteins, on protein folding and on intra- as well as, intermolecular processes. NMR is limited to relatively small proteins. Larger molecular structures (more than 30 KDa) give rise to overlapping signals that cannot be resolved. In these cases protein samples enriched with active isotopes  $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^2\text{H}$  are required.

The crystal structure of hRAR $\gamma$  has already been determined in our group<sup>146</sup>. We need to compare the high resolution crystal structures of the unphosphorylated RAR (LBD) with that of the phosphorylated receptor or with that of the S371E mutant in which the mutation is well known to mimic a phosphorylation event<sup>17, 89, 110</sup>. Differences in the structures dynamics due to the phosphorylation of the residue S371 will be then highlighted by nuclear magnetic resonance (NMR), the most suitable technique to provide information about the dynamics of macromolecules in solution at atomic resolution.

**Chapter 3 : from RAR gene to Crystal Structure,  
Molecular Dynamics and NMR Studies in  
Solution**

## 9. Protein expression and purification

### 9.1 (His)<sub>6</sub> hRAR $\gamma$ WT and h(His)<sub>6</sub> hRAR $\gamma$ S371E

The LBDs of hRAR $\gamma$  WT (178-423) and hRAR $\gamma$ S371E (178-423) were separately expressed in the bacterial host (*E. coli*, BL21 (DE3)). Both the proteins were expressed in pET-15b vectors with ampicillin resistance and fused with a hexahistidine (His)<sub>6</sub> tag for protein purification. In hRAR $\gamma$ S371E a mutation was introduced at position 371 exchanging a serine for a glutamic acid to mimic the phosphorylation of the residue<sup>17, 89, 110</sup>.

A common purification protocol was set up for the two proteins consisting in the following consecutive steps:

- 1- Metal (Ni<sup>2+</sup>) affinity chromatography
- 2- Size exclusion chromatography

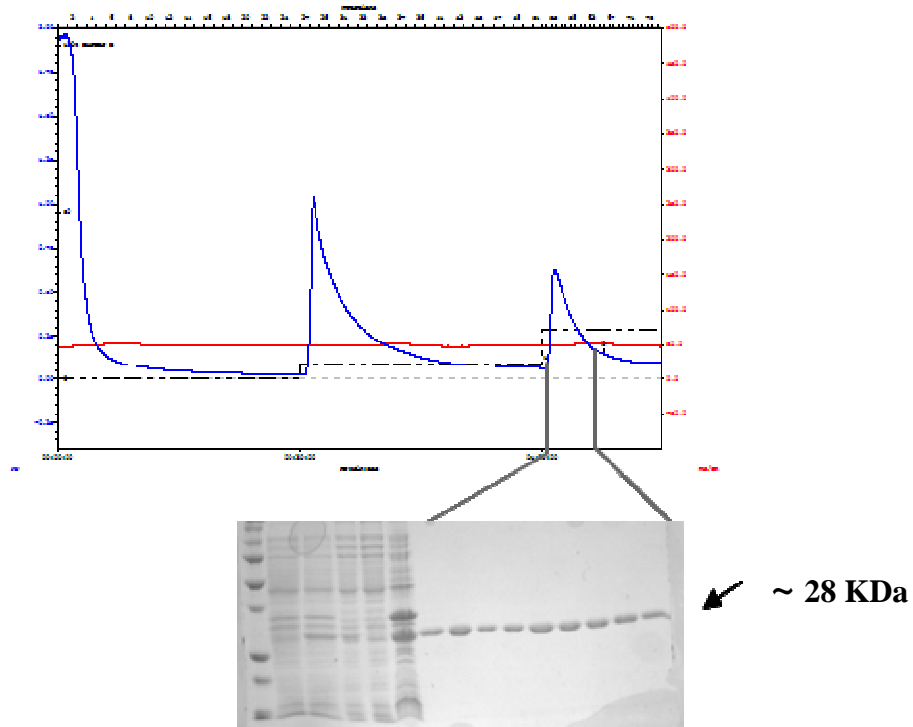
The yield of the pure protein was respectively of 6 mg per litre of culture for hRAR $\gamma$  WT and of 7 mg per litre of culture for hRAR $\gamma$  S371E. Ligand binding was confirmed by ESI-mass spectrometry and the purity of the proteins was checked by SDS-gel electrophoresis. The polydispersion of the purified proteins was analyzed by dynamic light scattering (DLS) measurements and resulted of the 22% for both hRAR $\gamma$  and hRAR $\gamma$  S371E.

Below I reported the chromatograms concerning the two-step purification of hRAR $\gamma$  WT and hRAR $\gamma$ S371E.

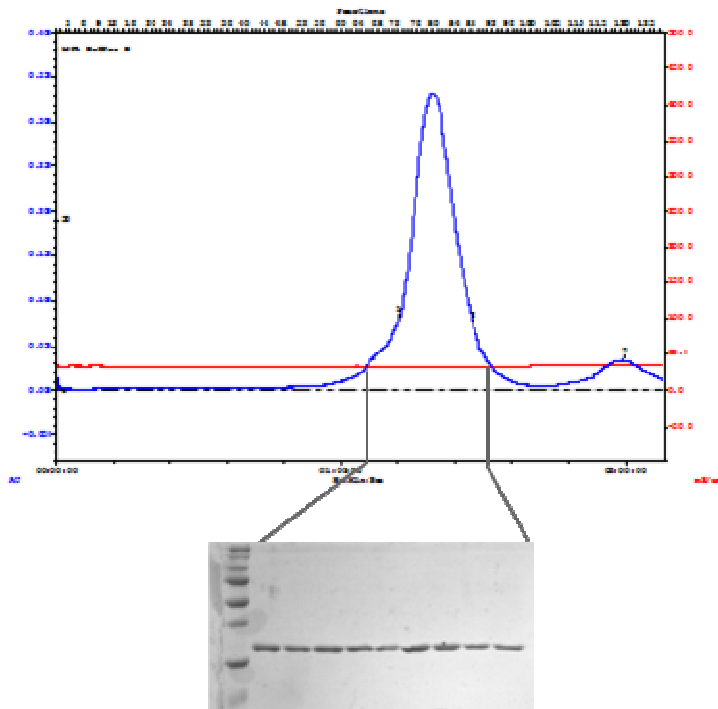
The typical metal affinity chromatogram I registered consisted of several peaks corresponding to different steps of the elution by increasing concentrations of imidazole. The size exclusion chromatogram consisted in a single major peak corresponding to the elution of RAR $\gamma$  as a monomer homogeneous pure sample.

Both the hRAR $\gamma$  WT and hRAR $\gamma$ S371E are stable in the absence of ligand. For this reason the proteins were purified in the absence of ligand and 9-cis retinoic acid was added to the purified protein, which allowed to spare a considerable amount of material.

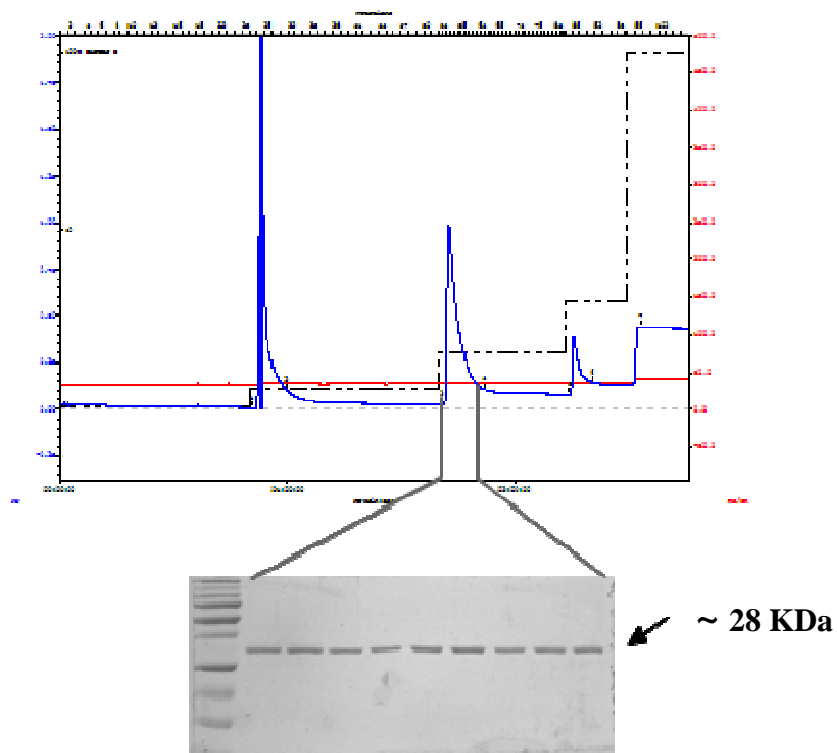
The ligand is added to a two fold number of moles to respect to the moles of protein. The sample, sheltered form light, was let incubate for 30 min at 4°C to ensure the binding.



**Fig.7-** Metal affinity chromatogram and SDS-PAGE of hRAR $\gamma$  wild type (apo)

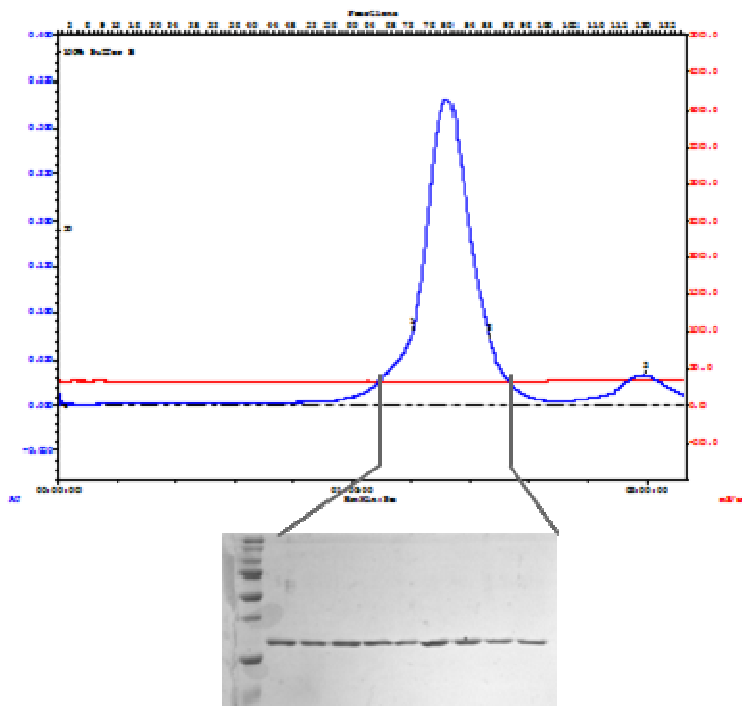


**Fig.8-** Size exclusion chromatogram and SDS-PAGE of hRAR $\gamma$  wild type (apo)



**Fig.9-** Metal affinity chromatogram and SDS-PAGE of hRAR $\gamma$  S371E (apo)





**Fig. 10-** Size exclusion chromatogram and SDS-PAGE of hRAR $\gamma$  S371E (apo)

## 9.2 (GST)-hRAR $\alpha$ , (GST)-hRAR $\alpha$ S369E, (GST)-hRXR $\alpha$

The LBDs of hRAR $\alpha$  WT (176-421), hRAR $\alpha$ S369E (176-421) and hRXR $\alpha$  (223-462) were separately expressed in the bacterial host (*E. coli*, BL21 (DE3)). The three proteins were expressed in pGex-4T3-Gateway vectors carrying ampicillin resistance and a Glutathione-S-transferase (GST) tag.

The hRAR $\alpha$  WT and hRAR $\alpha$ S369E were either expressed in rich medium or minimal media for  $^{15}\text{N}$  isotopic enrichment. In hRAR $\alpha$ S369E a mutation was introduced at position 369 exchanging a serine for a glutamic acid to mimic the phosphorylation of the residue<sup>17, 89, 110</sup>.

GST fusion cleavage was performed according to experimental requirements.

When GST cleavage was needed the ligand was added before purification to reduce the destabilisation of protein upon tag digestion and a common protocol for the three receptors was adopted consisting in the three following steps:

- 1- Glutathione (GSH) affinity chromatography
- 2- Cleavage of the GST fusion tag by thrombin digestion
- 3- Size exclusion chromatography

When experimental requirements did not include GST cleavage the purification protocol, adapted to the three receptors consisted in the following alternative procedure:

- 1- Glutathione (GSH) affinity chromatography
- 2- Elution of the GST fusion protein by glutathione (GSH) digestion
- 3- Size exclusion chromatography

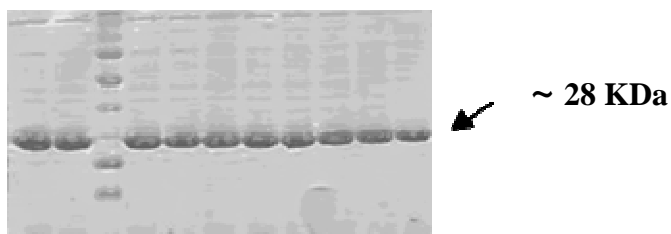
The yield of the pure proteins was of 7 mg per litre of culture for both hRAR $\alpha$  WT and hRAR $\alpha$ S369E, either in unlabelled or the  $^{15}\text{N}$  enriched medium. The yield of pure RXR was of 30 mg per litre of culture.

The binding of the ligand was checked by ESI-mass spectrometry and the purity of the proteins was checked by SDS-gel electrophoresis. The polydispersion of the purified proteins was analyzed by dynamic light scattering (DLS) measurements and resulted respectively of the 20% and 21% for hRAR $\alpha$  and hRAR $\alpha$ S369E.

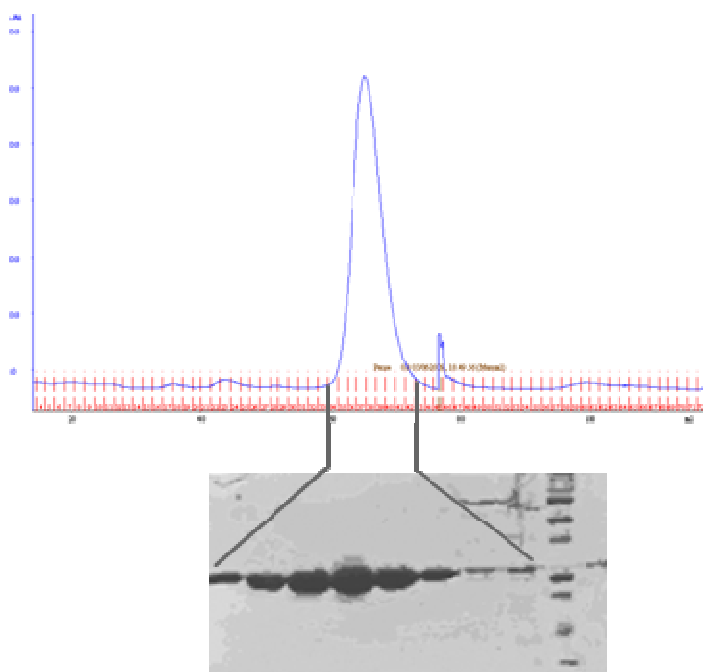
Here below I reported the SDS-PAGEs concerning the three step purification of hRAR $\alpha$  WT (same for hRAR $\alpha$ S369E and hRXR $\alpha$ ).

The GSH affinity chromatography consisted in the wash of the GSH beads, to which the GST-fusion protein is bound, and it was performed onto peristaltic pump. Enzymatic cleavage of the GST tag was carried on beads in presence of thrombin. The size-exclusion chromatogram consisted in a major peak corresponding to the elution of RAR $\alpha$  as a monomer homogeneous pure sample, as shown by the SDS gel and the DLS measurement. When the GST fusion was cleaved the ligand was added from the cell lysis to increase protein stability in solution. As retinoids undergo UV-light induced

isomerization, the purification was carried and the final protein sample was handled under dim light.



**Fig.11-** Elution of hRAR $\alpha$  WT (holo) after thrombin digestion on beads



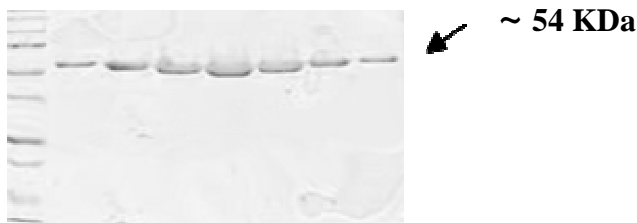
**Fig.12-** Size exclusion chromatogram and SDS-PAGE of hRAR $\alpha$  WT (holo)

Here below I reported the final SDS-PAGEs concerning the purification step of GST-hRAR $\alpha$ S369E (same for hRAR $\alpha$  WT and hRXR $\alpha$ ).

The GSH affinity chromatography consisted in the wash of the GSH beads to which the GST-fusion protein and it was performed onto peristaltic pump.

Elution of the GST fusion protein was carried out in presence of GSH.

Size exclusion chromatogram consisted in a major peak corresponding to the protein of interest.



**Fig.13-** SDS-PAGE of GST- hRAR $\alpha$ S369E (apo) after size-exclusion chromatography

### 9.3 hCyclin H

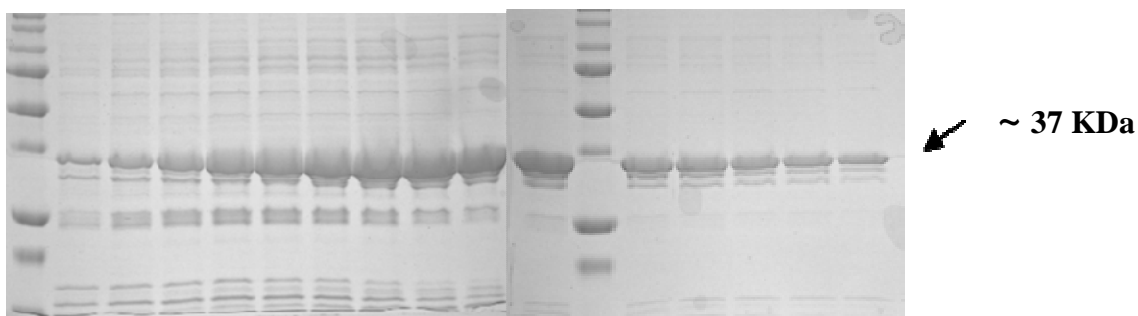
The full length hCycH (1-323), was expressed in the bacterial host (*E. coli*, BL21 (DE3)). The expression was carried out in pET-15b vectors with ampicillin resistance and hexahistidine (His)<sub>6</sub> tag for protein purification.

The purification protocol consisted in three main steps:

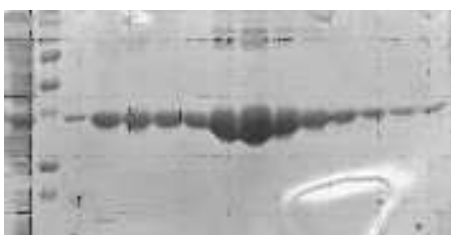
- 1- Metal (Co<sup>2+</sup>) affinity chromatography
- 2- Cleavage of the His tag by thrombin digestion
- 3- Size exclusion chromatography

The yield of the pure protein was of 12 mg per litre of culture. The purity of the protein was checked by SDS-gel electrophoresis and MALDI-TOF mass spectrometry and resulted of the 95%.

As previously discussed the metal affinity chromatogram consisted of several peaks corresponding to different steps of the elution by increasing concentrations of imidazole. The size exclusion chromatogram consisted in a single major peak corresponding to the protein of interest (chromatograms not shown). Below I reported the SDS-PAGEs concerning the two chromatography steps of the purification of hCycH.



**Fig.14-** SDS-PAGE of hCycH after metal affinity chromatography



**Fig.15-** SDS-PAGE of hCycH after size-exclusion chromatography

## 9.4 PKA

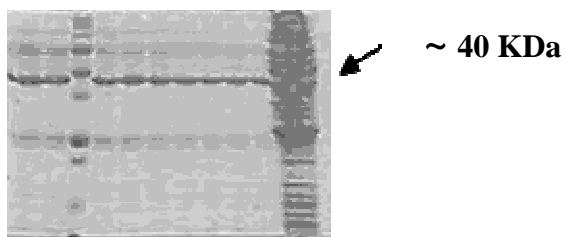
The catalytic subunit of the c-AMP dependent protein kinase (2-351) was expressed in the bacterial host (*E. coli*, BL21 (DE3)). The expression was carried out in the Gateway vector pHGWA containing Ampicillin resistance and hexahistidine (His)<sub>6</sub> tag.

The purification protocol consisted in one main step:

### 1- Metal (Co<sup>2+</sup>) affinity chromatography

The yield of the pure protein was of 1 mg per litre of culture. For our purposes the level of protein purity gained with the metal affinity chromatography was sufficient and, considered the low yield of the protein expression, a further purification was not performed.

The metal affinity chromatogram consisted of different peaks corresponding to different steps of the elution by increasing concentrations of imidazole. Below I reported the SDS-PAGEs concerning the single step purification of PKA.



**Fig.16-** SDS\_PAGE of PKA after metal affinity chromatography

## 10. Crystallization of hRAR $\gamma$ S371E (LBD) in complex with 9-cis retinoic acid and X-ray crystal structure resolution

### 10.1 Crystallization of hRAR $\alpha$ S369E (LBD) in complex with 9-cis retinoic acid and a coactivator peptide

The ligand binding domain of hRAR $\alpha$ S369E was expressed and purified. The protein bound to the agonist ligand 9-cis retinoic acid was concentrated up to 3, 6 or 8 mg/ml. The protein at a concentration of 6 mg/ml crystallized in presence of a SRC1 coactivator peptide. Crystals were tested and diffracted on a synchrotron source up to 2.7 Å but resulted highly mosaic. Several attempts to optimize the crystallization conditions resulted unsuccessful.

## 10.2 hRAR $\gamma$ S371E (LBD) in complex with 9-cis retinoic acid

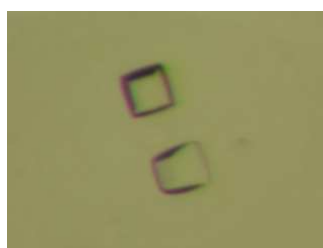
### 10.2.1 Preliminary crystallization trials

The other protein isotype hRAR $\gamma$ S371 E LBD was expressed and purified. The protein, bound to the ligand 9-cis retinoic acid, was concentrated up to 3, 6 or 8 mg/ml. Preliminary crystallization screenings were carried out using the sitting drop vapour diffusion method. Crystallization trials were performed in 1+1  $\mu$ l, 96 wells Hampton Research Innovaplate crystallization plates. Screening around four hundred conditions resulted in three hits for crystallization condition optimization: the first is 0.2 M Magnesium nitrate hexahydrate and 20% Peg 3350, the second is 0.2 M Magnesium chloride, 0.1 M Tris HCl pH 8.0 and 20% Peg 6K and the third is 0.2 M Calcium chloride, 0.1 M Tris HCl pH 8.0 and 20% Peg 6K (Table 3)

0.2 M Magnesium nitrate  
hexahydrate, 20% Peg 3350



0.2 M Magnesium chloride,  
0.1 M Tris HCl pH 8.0, 20%  
Peg 6K



0.2 M Calcium chloride, 0.1  
M Tris HCl pH 8.0, 20% Peg  
6K



**Table.3-** Preliminary crystallization conditions and crystals

Crystals from these conditions appeared overnight at 17°C and kept growing over the next 3-4 days. The protein crystallizes in presence of the zwitterionic detergent 3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS) and the neutral detergent n-dodecyl-β-d-maltopyranoside (C12M). Efforts to crystallize hRARγS371E in the absence of detergents and in presence of coactivator peptides were unsuccessful.

### 10.2.2 Optimisation and scaling up of the crystallization conditions

Optimisation screenings around the initial conditions were successfully performed by scaling up the size of the crystallization drops from 1 + 1 μl (protein+ reservoir) to 2 + 2 μl using the hanging drop technique in Hampton Research VDX plates.

This finer screening led to the growth of single crystals with improved visual quality.

The best crystals were obtained from the following optimized conditions:

- (1) 0.3 M Magnesium nitrate hexahydrate, 18% Peg 3350
- (2) 0.2 M Magnesium chloride, 0.1 M Tris HCl pH 7.5, 14% Peg 6K
- (3) 0.3 M Magnesium chloride, 0.1 M Tris HCl pH 8.0, 14% Peg 6K
- (4) 0.2 M Calcium chloride, 0.1 M Tris HCl pH 7.5, 14% Peg 6K
- (5) 0.4 M Calcium chloride, 0.1 M Tris HCl pH 8.0, 14% Peg 6K

Further improvements in the quality and size of the crystals were obtained on the basis of the optimized crystallization conditions by varying the ratio of protein and reservoir buffer constituting the drop as well as the volume of buffer in the reservoir. The best crystals were obtained from drops of 5 μl, constituted by a ratio 3: 2 of protein and reservoir buffer, and a reservoir volume of 1 ml. Single crystals appeared overnight at 17°C and kept growing over one week to an average size of 200x200x100 μm<sup>3</sup>. All the optimized crystallization conditions produced bi-pyramidal crystals, very much similar in shape to the diffracting crystals obtained previously for hRARγ WT (LBD) for which the structure has already been solved<sup>92</sup>.

One crystal grown in the condition with 0.2 M Calcium Chloride, 0.1 M, Tris HCl pH 7.5, 14% Peg 6K with a size of 300x200x200 μm<sup>3</sup> produced the best diffraction at a



resolution of 1.7 Å. Diffraction from this crystal produced clean and strong reflections. A total of 450 frames with 0.4 ° of oscillation angle were collected at the ESRF synchrotron BEAMLINe ID14-1 at 100 K.

### 10.2.3 Data Processing, Molecular Replacement and Structure Refinement

The data obtained for hRAR $\gamma$ S371E (LBD) in complex with 9-cis retinoic acid were processed with the program HKL2000<sup>130</sup>. Results from data processing suggested the space group P4<sub>1</sub>2<sub>1</sub>2 with one monomer per asymmetric unit and similar lattice parameters as for the hRAR $\gamma$  wild type LBD (PDB-ID: 3LBD<sup>92</sup>). Following data processing, molecular replacement was carried out to find the solutions for the new complex using the hRAR $\gamma$  wild type LBD structure as a starting model. This calculation was done by using AMoRe<sup>125</sup>, integrated in the graphical interfaced version of the CCP4 (Collaborative Computational Project, 1994). A unique solution was obtained and refinement of the structure was started with the rigid body. Manual building of the structure was carried out with COOT<sup>52</sup> with iterative cycles of structure refinement performed through REFMAC5<sup>123</sup>. The parameters and the topology coordinates for the ligand and the detergent were generated from PRODRG<sup>156</sup> as REFMAC5 uses PRODRG files. The final steps of the refinement included the addition of the ligand, the detergent and the water molecules. The final structure reported values for R<sub>work</sub> and R<sub>free</sub> that were respectively of 19.4% and 23.4%. Refer to table 4 for an overview of data collection parameters and refinement statistics. Ramachandran plot for the geometry of the main chain is shown in Figure 17.

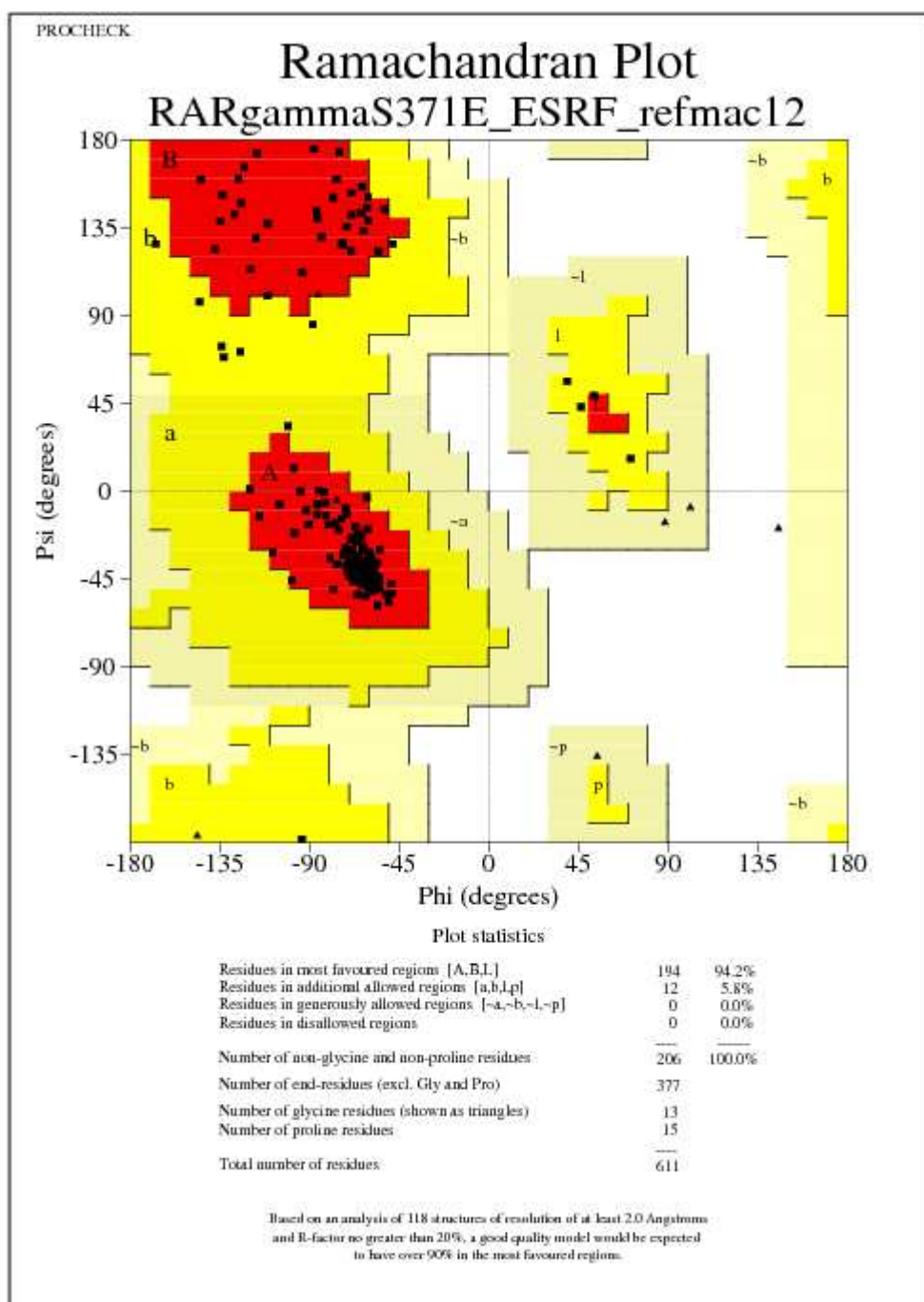
hRAR $\gamma$ S371E (LBD) in complex with 9-cis retinoic acid	
<i>Data processing</i>	
Resolution (Å)	50-1.69 (1.72-1.69)
Crystal space group	P4 <sub>1</sub> 2 <sub>1</sub> 2
Cell parameters (Å)	$a = 157.4, b = 59.9, c = 157.4$
Unique reflections	32495 (1589)
Mean redundancy	13.4 (12.3)
$R_{\text{sym}}$ (%) <sup>a</sup>	4.4 (31.8)
Completeness (%)	100 (99)
Mean $I/\sigma$	70.17 (11.81)
Wilson $B$ factor (Å <sup>2</sup> )	22.1
<i>Refinement</i>	
Resolution (Å)	50-1.69
Number of non-hydrogen atoms	
RAR-LBD	1852
Ligands	22
Water molecules	157
RMSD bond length (Å)	0.01
RMSD bond angles (°)	1.11
$R_{\text{cryst}}$ (%) <sup>b</sup>	19.4
$R_{\text{free}}$ (%) <sup>c</sup>	23.4
-----	
Ramachandran plot (%)	
Core	94.2
Allow	5.8
Generous	0

**Table 4. Data collection and refinement statistics**

<sup>a</sup>  $R_{\text{sym}} = 100 \times \sum_{\mathbf{h}j} |I_{\mathbf{h}j} - \langle I_{\mathbf{h}} \rangle| / \sum_{\mathbf{h}j} I_{\mathbf{h}j}$ , where  $I_{\mathbf{h}j}$  is the  $j$ th measurement of the intensity of reflection  $\mathbf{h}$  and  $\langle I_{\mathbf{h}} \rangle$  is its mean value.

<sup>b</sup>  $R_{\text{cryst}} = 100 \times \sum (|F_o| - |F_c|) / \sum |F_o|$ , where  $|F_o|$  and  $|F_c|$  are the observed and calculated structure factor amplitudes, respectively.

<sup>c</sup> Calculated using a random set containing 5% of observations that were not included throughout refinement [Brünger A. T., (1992) The Free R Value: a Novel Statistical Quantity for Assessing the Accuracy of Crystal Structures, *Nature* **355**, 472-474].



**Fig.17-** Ramachandran plot of the structure of hRAR $\gamma$ S371E (LBD) in complex with 9-cis retinoic acid

## 11. Molecular dynamics studies of the effects of the S371 phosphorylation onto RAR $\gamma$

Molecular dynamics studies have been carried out in collaboration with Dr. Annick Dejaegere's group. These studies aimed at the investigation of the effects of the phosphorylation of the serine located into helix H9 of the RAR $\gamma$  LBD with particular regard to the effects of this phosphorylation onto hCycH binding region. Molecular dynamics simulations have been carried out onto hRAR $\gamma$  WT (LBD) and its mutant, hRAR $\gamma$  S371E, on the basis of the respective crystal structures. The structures of phosphorylated mutants of hRAR $\gamma$ S371 carrying respectively one (S371P1) or two (S371P2) negative charges were built and analysed. Mutants of hRAR $\gamma$  S371E in which the glutamic acid was substituted by a phosphorylated serine residue carrying respectively one (RAR-mut- pho 1) or two (RAR-mut- pho 2) negative charges were also built and analysed.

### 11.1 Structures preparation for molecular simulation

Force fields to describe the nuclear receptors LBDs, the 9-cis retinoic acid and the hydrogen atoms were introduced<sup>29, 111, 112</sup>. The protonation state of the residues, determining the total charge of the protein and the local distribution of the charges around the amino acids, was established on the basis of two different programs PROPKA and UHBD<sup>12, 108, 154, 173</sup>.

The dynamics were produced with the program NAMD<sup>135</sup>. The hydrogen bonds were fixed with the algorithm SHAKE to improve calculation times<sup>153</sup>. The temperatures were adjusted at each step and the pressure was constantly controlled with a Langevin piston.

To delete the wrong contacts between the protein and the solvation water molecules, an energy minimisation of 1000 steps with conjugated gradient (CG) was carried out.

During this step the protein, the ligand and the water molecules were fixed. Then the system was warmed up to 600°K for 23 steps. A shorter energy minimisation (CG) was carried out followed by a warming up step to 300°K. In a second moment the protein, the ligand and the crystallographic waters were not kept fixed anymore and an energy minimisation (CG) of 2000 steps was carried out, followed by a warming up of 13 steps up to 300°K. An equilibration step was then followed by the observation time during which the conformations of the structure was investigated. The dynamics of each system was studied in isotherm and isobar conditions for a total of 6800 hours of calculations on 512 processors.

## 11.2 Analysis

### 11.2.1 RMSD

The Root Mean Square Deviation (RMSD) value provides a description of the differences in the structure of the protein at a given time of the simulation to respect to the initial protein structure taken as reference. The structures of the protein at different times were superimposed to the structure of the reference protein and the corresponding RMSD value calculated according to the formula:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N_{atomes}} d_i^2}{N_{atomes}}}$$

where N represents the number of atoms for which the RMSD is calculated and  $d_i$  represents the distance between the coordinates of the atom i in the two superposed structures.

### 11.2.2 Fluctuations RMS

This value provides an estimation of the mobility of a structure at a given time of the simulation to respect to the averages structure. The average structure out of all the possible conformations is calculated on the basis of the average position  $\langle r_i \rangle$  of each atom.

$$\text{Fluctuations RMS of the atom } i = \sqrt{\frac{1}{n} \sum_{k=1}^n (r_{ik} - \langle r_i \rangle)^2}$$

### 11.2.3 Hydrogen bonds analysis

Interactions between an acceptor and a donor atom were considered as hydrogen bonds within 3.4 Å. The number and the life time of the hydrogen bonds originated during the simulations were analysed for each system with the program Multivac (M. Ferrario, G. Marchetti, M. Oberlin, A. Dejaegere, not published).

### 11.2.4 Dihedral angles analysis

The dihedral angle adopted by the protein backbone and by each residue during the dynamic simulation was analysed with the program Multivac (M. Ferrario, G. Marchetti, M. Oberlin, A. Dejaegere, not published).

## 12. Dynamic studies of the RAR $\alpha$ wild type and RAR $\alpha$ S371E by nuclear magnetic resonance (NMR).

### 12.1 Preliminary 2D heteronuclear spectra

Preliminary heteronuclear 2D experiments were carried out onto  $^{15}\text{N}$ -hRAR $\alpha$ S369E and  $^{15}\text{N}$ -hRAR $\alpha$ WT in complex with the ligand 9-cis retinoic acid. The alpha isotype of RAR was used in these experiments due to the lack of monodispersity of the RAR $\gamma$  isotype in NMR suitable buffer conditions. The samples were placed in a 3mm NMR tube and spectra were acquired on a Bruker Avance 800 MHz spectrometer.

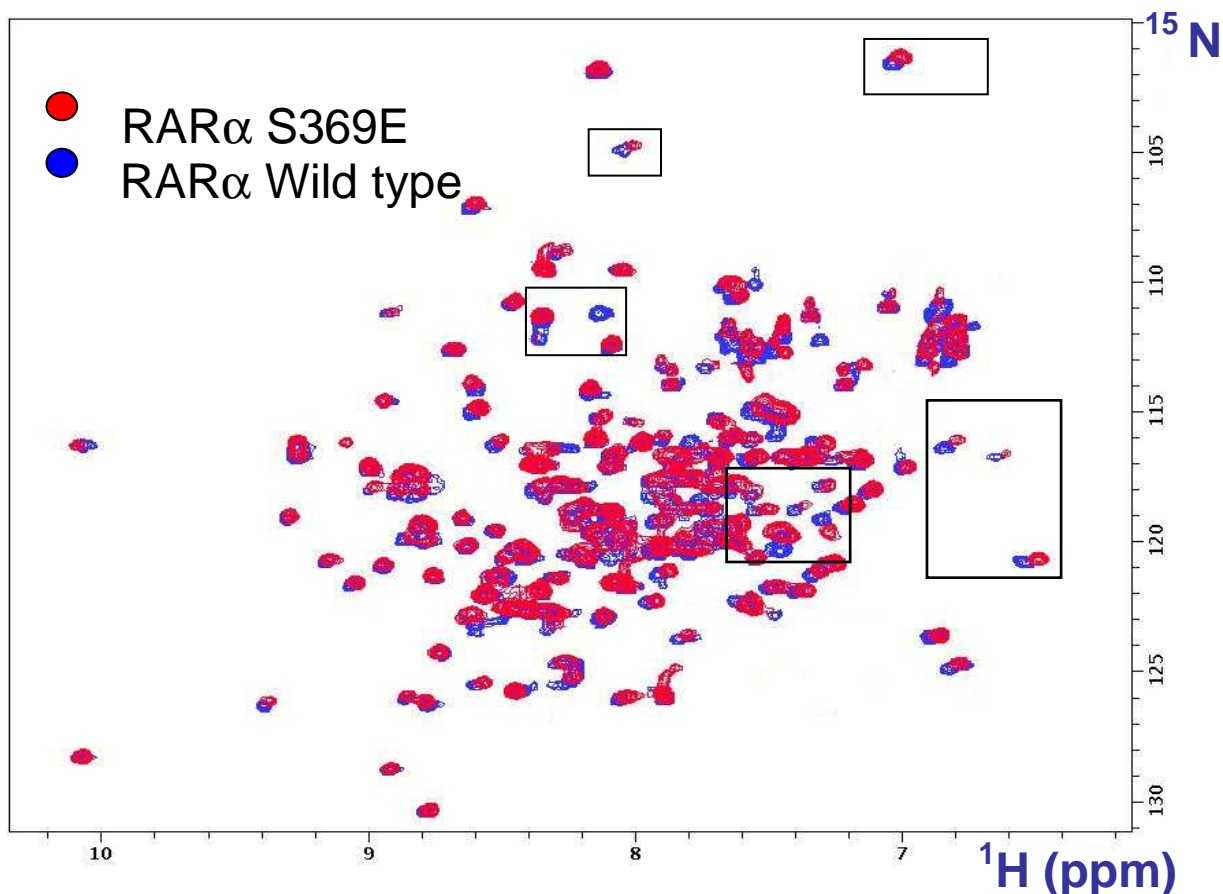
The HSQC spectra were registered at different temperatures, pH and buffer compositions to find the optimal conditions for the acquisitions.

Difference of probe temperature between 298K and 303 K did not affect significantly the spectra while a decreasing value of the pH from 7.5 to 6.5 increased the spectra quality as at pH 6.5 a considerable number of missing signals were recovered.

The best condition for spectra acquisition resulted to be common to the two proteins and is described here below:

Buffer composition	Bis-Tris <sub>10</sub> NaCl <sub>300</sub> Chaps <sub>1.5</sub> TCEP <sub>1 (mM)</sub> pH 6.5
Protein concentration	$\approx 400\mu\text{M}$
T probe	303 K

$^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of  $^{15}\text{N}$ -hRAR $\alpha$ S369E and  $^{15}\text{N}$ -hRAR $\alpha$ WT were acquired in the previously described conditions and the spectra were superimposed. Significant shifts for several peaks are enlightened in the black boxes (Fig. 18)



**Fig.18-** <sup>1</sup>H-<sup>15</sup>N HSQC spectra of <sup>15</sup>N-hRARαS369E and <sup>15</sup>N-hRARαWT acquired on a Bruker Avance 800 MHz spectrometer, at 303K.

## 12.2 RARα wild type and RARαS371E <sup>15</sup>N longitudinal and transverse longitudinal relaxation rates measurements

Protein backbone <sup>15</sup>N spin relaxation rates can be measured by solution NMR and provide useful dynamic information with a residue-specific resolution. In a <sup>15</sup>N relaxation experiment, one creates non-equilibrium spin order and records how this relaxes back to equilibrium. At equilibrium, the <sup>15</sup>N magnetization is aligned along the external field. This alignment can be altered by radio frequency pulses. The magnetization will relax back to equilibrium along the direction of the magnetic field with longitudinal relaxation rate  $R_1$ .



Outside the equilibrium the magnetization also has a component perpendicular to the external magnetic field. The constant rate for this spin component to return to equilibrium is called transverse relaxation rate  $R_2$ . The conventional method to perform relaxation rates measurements is to record a series of 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra with varied relaxation delays and derive  $^{15}\text{N}$  relaxation rates by numerical fitting of the relaxation free induction decay data (FID).

In order to highlight differences in the dynamics of wild type RAR and its mutant relaxation experiments were performed on Avance Bruker 500 MHz spectrometer at 298K onto samples of  $^{15}\text{N}$ -hRAR $\alpha$ S369E and  $^{15}\text{N}$ -hRAR $\alpha$ WT in complex with the ligand 9-cis retinoic acid. Both the protein samples were prepared in previously found optimised conditions and had a final concentration of about 300  $\mu\text{M}$ .

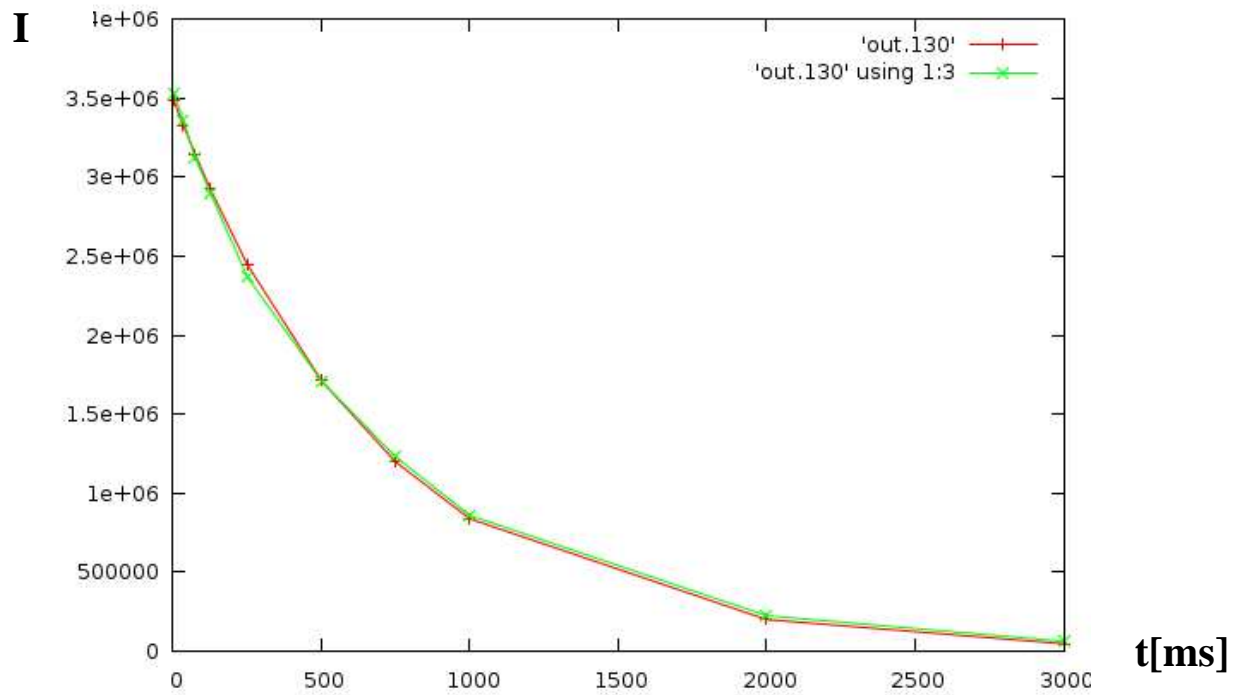
$^{15}\text{N}$  longitudinal relaxation rate  $R_1$ , transverse relaxation rate  $R_2$ , and steady-state heteronuclear NOEs, were measured with pulse sequences described by Farrow et al.<sup>54</sup>.  $R_2$  were measured using a refocusing time of 450 ms. In all experiments the water signal was suppressed with 'water flip-back' scheme. The  $^{15}\text{N}$  longitudinal relaxation rates,  $R_1$ , were measured using delays in the pulse sequence of 2.5, 35, 75, 125, 250, 500, 750, 1000, 2000 and 3000 ms for both the protein samples. The relaxation delays used for  $^{15}\text{N}$  transverse relaxation rates  $R_2$  were of 16.96, 33.92, 50.88, 67.84, 84.8, 118.72, 169.6, 220.48, 288.32 ms for both the protein samples.

The heteronuclear  $^1\text{H}$ - $^{15}\text{N}$  NOEs were calculated considering the ratio of the peak volumes in spectra recorded with and without  $^1\text{H}$  saturation. Each experiment was comprised of 32 scans for  $R_1$  and  $R_2$  experiments and 64 scans for  $^1\text{H}$ - $^{15}\text{N}$  NOE experiments. All 2D data consisted of 2K data points in the acquisition dimension and of 256 experiments in the indirect dimension.

Relaxation rates  $R_1$  and  $R_2$  were determined by fitting the cross-peak intensities ( $I$ ) as a function of the delay ( $t$ ) within the pulse sequence, to a single exponential decay curve according to the following equation (Fig.19) :

$$I(t) = A \cdot \exp[-R \cdot t]$$

where A, and R were adjustable fitting parameters. The value for the correlation time  $\tau_c$  was estimated from the R2 and R1 ratio and resulted of 17ns.



**Fig.19-** Intensity decay vs. time delay between two consecutive pulse sequences plotted for peak number 130.

## **Chapter 4 : Results and Discussion**

## 13. X-ray crystal structure of hRAR $\gamma$ S371E (LBD)

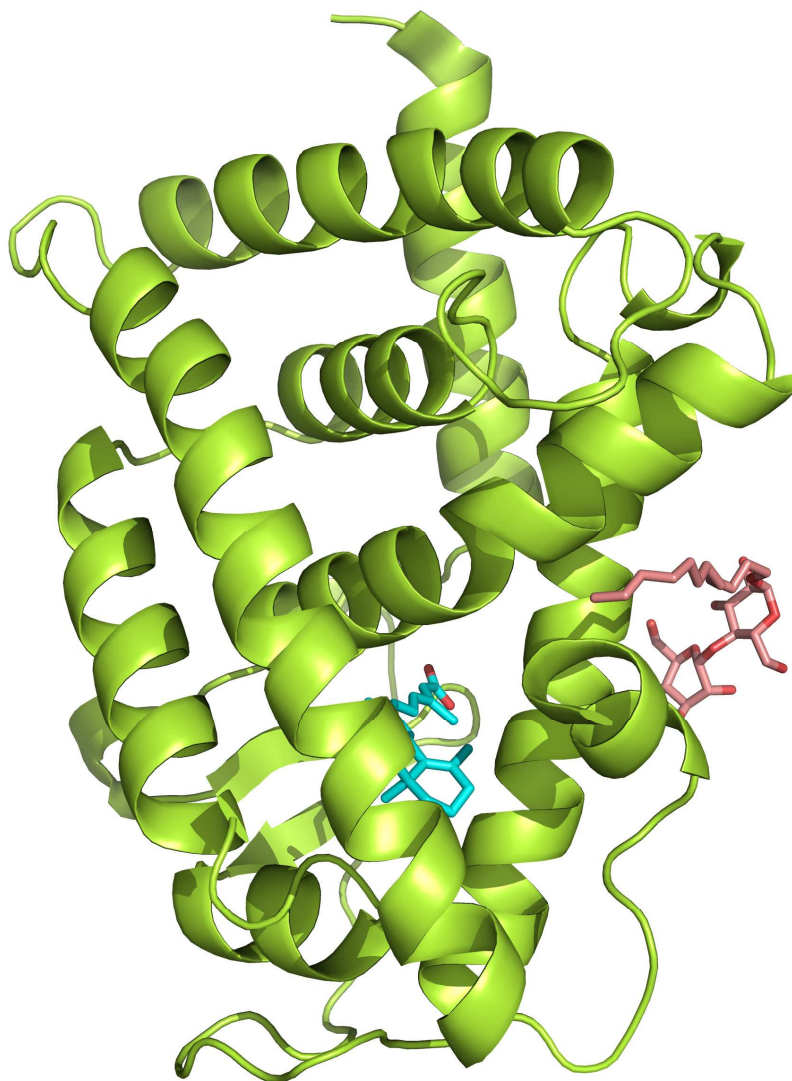
### 13.1 X-ray crystal structure of hRAR $\gamma$ S371E (LBD) in complex with 9-cis retinoic acid

Phosphorylation at the RAR $\gamma$  ligand binding domain (AF-2) by PKA enhances the binding of hCyclinH to the receptor and therefore facilitates the phosphorylation of RAR $\gamma$  (AF-1) domain by CAK<sup>24</sup>.

We investigated the consequences of the phosphorylation on the structure of RAR $\gamma$  ligand binding domain by comparing the high resolution crystal structures of the RAR $\gamma$  WT (LBD) with that of the S371E mutant in which the mutation reproduces in large part the effects of a phosphorylation event.

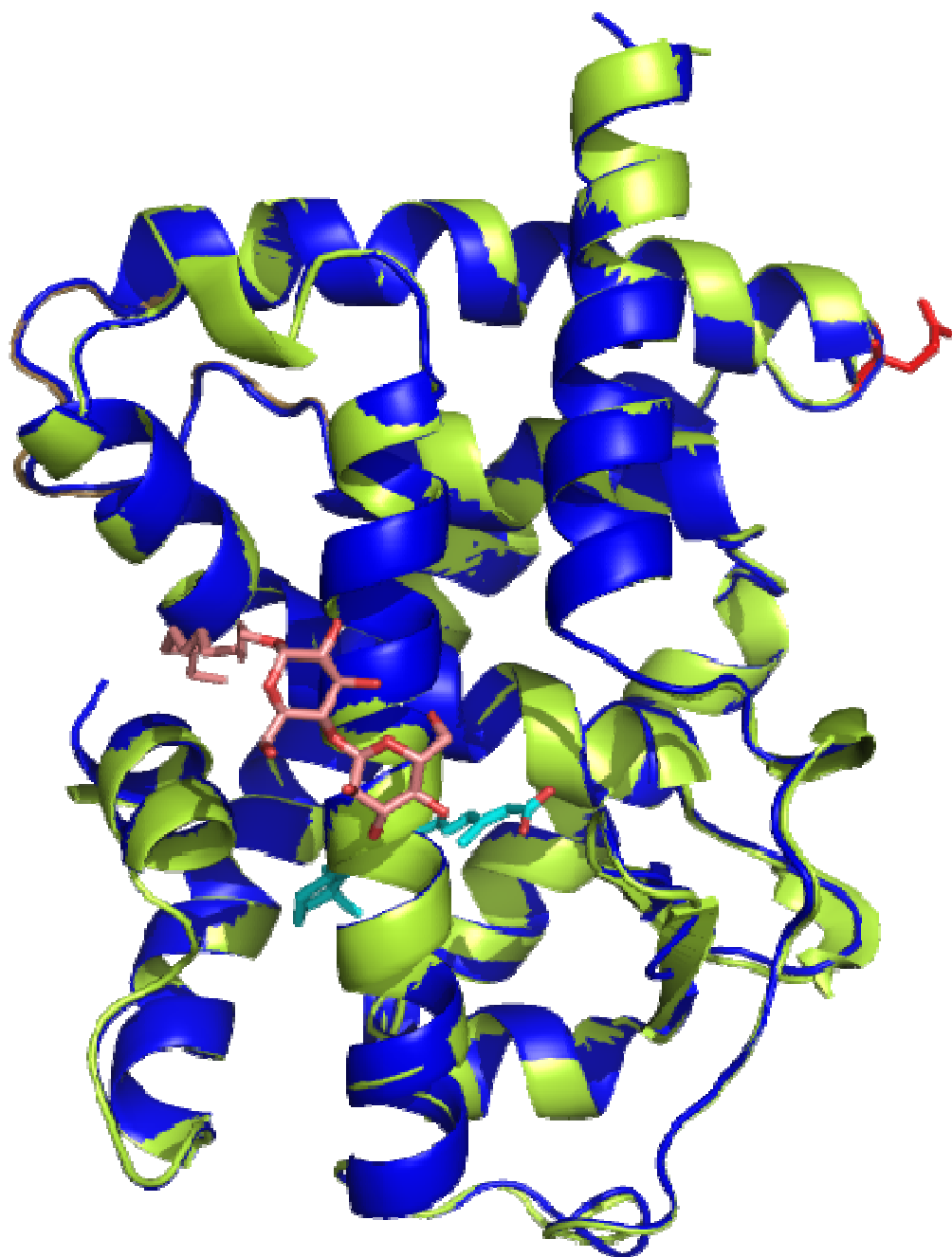
The ligand binding domain of hRAR $\gamma$ S371E bound to the ligand 9-cis retinoic acid was crystallized in presence of the neutral detergent n-dodecyl- $\beta$ -d-maltopyranoside (C<sub>12</sub>M), required for stabilization of hRAR $\gamma$ S371E (LBD)<sup>91</sup>. Optimized crystals were tested on a synchrotron source and a complete data set was collected at 1.7 Å. Crystals belong to the tetragonal space group P4<sub>1</sub>2<sub>1</sub>2 with one monomer per asymmetric unit. The structure was solved by molecular replacement using hRAR $\gamma$  WT as a starting model<sup>92</sup>.

After refinement the reported values for R<sub>factor</sub> and R<sub>free</sub> were respectively of 19.4% and 23.4%. The human RAR $\gamma$ S371E-holo-LBD adopts a classical anti-parallel alpha helical "sandwich" (Fig.20) similar to that of the wild type RAR $\gamma$  LBD structure conformation firstly described for hRAR $\gamma$ <sup>146</sup>. The electron density map showed the presence of the ligand buried into the ligand binding pocket and an extra electron density around position 371 confirmed the mutation of the serine residue into a glutamic acid. The detergent mediates a crystal contact in the region of the receptor involved in coactivator binding. The dodecyl moiety of the detergent is buried in a hydrophobic channel, whereas the maltoside head group establishes a hydrogen bond with water molecules and polar residue side of the chains.

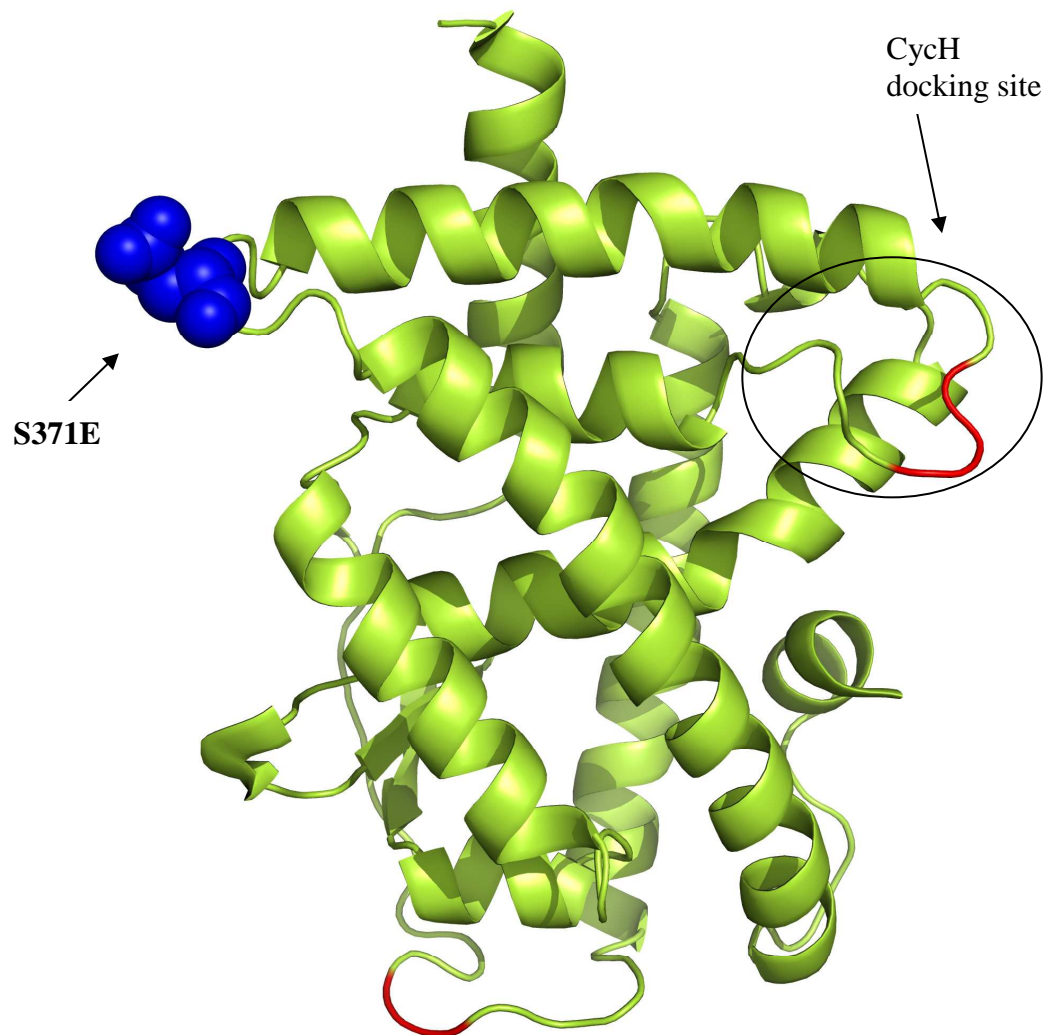


**Fig.20- X-ray Crystal structure of hRAR $\gamma$ S371E in complex with 9-cis retinoic acid.** The ligand, buried in the ligand binding pocket, is in blue; the detergent C<sub>12</sub>M is in purple.

The hRAR $\gamma$ S371E (LBD) and hRAR $\gamma$  (LBD) structures are isomorphous (respectively in green and blue in Fig 21). The r.m.s. deviation calculated on the carbons alpha of the backbone is of 0.47 Å. Interestingly one of the two highest r.m.s.d values, equal to 1Å, is localised in the loop 8-9 involved in cychH recognition and binding rather than in the mutation region (Fig. 22).

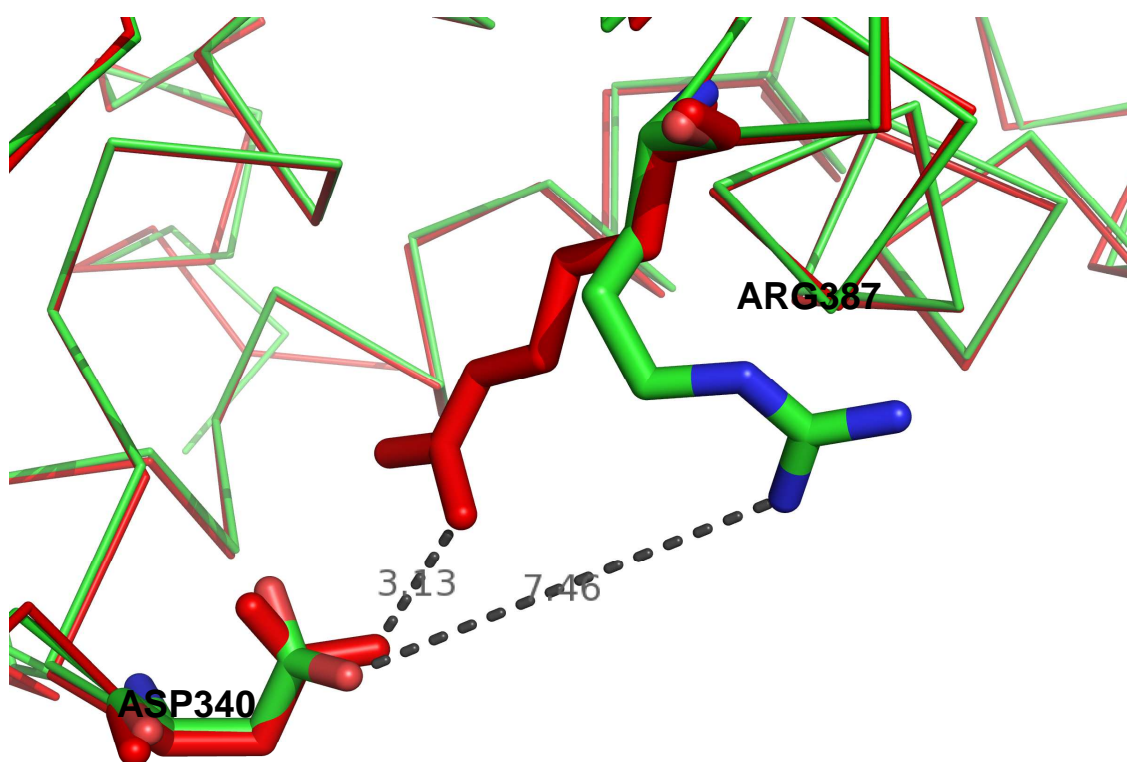


**Fig. 21– Superposition of the crystal structures of the LBDs of hRAR $\gamma$ WT (blue) and hRAR $\gamma$ S371E (green), in complex with the agonist ligand 9-cis retinoic acid (light blue). In pink the detergent C12M. The two structures are isomorphous.**



**Fig.22–** Two regions in RAR $\gamma$ S371E display r.m.s.d values of 1 Å (in red). One of these regions belongs to the part of RAR $\gamma$ S371E involved in the hCycH docking (circled). In blue the mutation S371E.

The analysis of the three-dimensional structures of hRAR $\gamma$ WT and the mutant hRAR $\gamma$ S371E showed a very similar orientation of the side chains. Remarkably a significant difference in the orientation of the side chains involves the loop 8-9 responsible for the hCycH binding. More precisely this difference is located at the level of the residue ASP340. In hRAR $\gamma$ S371E this residue forms a salt bridge with ARG387 belonging to H10 (Fig.23). This salt bridge is not formed in the hRAR $\gamma$ WT due to the orientation of the ARG387 which now points toward the solvent. Moreover in the transcriptional active heterodimer formed by hRAR $\beta$  and RXR $\alpha$  in complex with 9-cis retinoic acid, the ARG residue (ARG378 in RAR $\beta$ ) is implicated in a salt bridge with other residues belonging to RXR and adopts an orientation that resembles that of the WT rather than that of the mutant<sup>137</sup>.



**Fig. 23– Superposition of hRAR $\gamma$ S371E and hRAR $\gamma$ WT crystal structures. In RAR $\gamma$ S371E (in red) a salt bridge involves ASP340 (loop 8-9) and ARG387 (H10). This salt bridge is not present in RAR $\gamma$ WT (in green).**



## 14. Molecular dynamics studies on hRAR $\gamma$ S371E (LBD)

Molecular dynamics studies of the ligand binding domain of hRAR $\gamma$  WT and its mutants were performed in collaboration with Dr. Dejaegere's team and aimed at contributing to the understanding of the effects of the phosphorylation of the RAR ligand binding domain onto the region of the receptor involved in the recruitment of hCycH.

### 14.1 Molecular dynamics studies onto RAR $\gamma$

Molecular dynamic simulations enlightened that the phosphorylation region of RAR and the region involved in the recruitment of hCycH, both rich in charged amino acids, possess peculiar electrostatic characteristics. In particular the two methods used for the calculation of the protonation state of RAR residues, PROPKA and UHBD provided contradictory results for few amino acids: ASP324, GLU327 could be neutral and the CYS338 could be charged. Interestingly the ambiguous residues are located respectively in spatial proximity of the phosphorylation site (ASP324, GLU327) or of the CycH binding site, loop 8-9 (CYS338). Two protonation states were tested and the distances between the residues having an ambiguous degree of protonation were measured during the dynamic simulation. Significantly none of the tested conditions reproduces exactly the whole set of the crystallographic distances indicating the existence of multiple possible conformations corresponding to different protonation states of the residues.

Hydrogen bond formation was analysed during the dynamic simulation as any stable modification of the life time of a hydrogen bond can be part of a regulation mechanism. Hydrogen bonds modifications were monitored in the region close to RAR phosphorylation site and to hCyc binding site. Alternative hydrogen bonds in the RAR phosphorylation site can originate depending on the protonation state of the residues located in proximity of S371 and on the modifications occurring on S371. Interestingly formation of alternative hydrogen bonds in the phosphorylation site and involving residues located in the loop 8-9 were observed for any of the tested protonation states.

The existence of a salt bridge between the conserved ARG387, located in H10, and ASP340 belonging to the CycH binding region comprising loop 8-9 was observed in hRAR $\gamma$ S371E but not in hRAR $\gamma$ WT. When the S371 residue is phosphorylated the formation of the salt bridge between ARG387 and ASP340 arises spontaneously and rapidly. When the simulation is carried out onto hRAR $\gamma$ S371E the existing salt bridge between ARG387 and ASP340 is unmodified during in the dynamics. Molecular dynamics onto hRAR $\gamma$ WT showed, on the contrary, that the formation of the salt bridge between ARG387 and ASP340 is transient during the simulation. These results demonstrated for the first time the existence of a dynamic coupling between the regions involved in PKA phosphorylation and that involved in the recognition of hCycH.

## 15. NMR

### 15.1 NMR studies of hRAR $\alpha$ S371E (LBD) and of hRAR $\alpha$ WT (LBD) in solution

High resolution three-dimensional structure of proteins can be solved either in solid state through X-ray crystallography, or in solution through NMR spectroscopy. NMR has the advantage of studying proteins under conditions as close as possible to their physiological state.

NMR spectroscopy provides information ranging from the protein folding to the internal motions of the protein. All these aspects are essential for a deep understanding of the protein biological functions.

NMR structural studies of hRAR $\alpha$ S371E LBD and hRAR $\alpha$  LBD in solution required the isotopic enrichment of the proteins with  $^{15}\text{N}$ . HSQC (heteronuclear single quantum correlation) experiments were carried out on both proteins.

The HSQC experiment is the most important inverse NMR experiment. In a HSQC the nitrogen atom of an NHx group is directly correlated with the attached proton. In the resulting spectrum each HSQC signal represents a H<sup>N</sup> proton of the protein backbone and, since there is only one backbone H<sup>N</sup> per amino acid, each cross-peak corresponds to one single amino acid.

<sup>1</sup>H-<sup>15</sup>N HSQC spectra for <sup>15</sup>N-hRAR $\alpha$ S369E and <sup>15</sup>N-hRAR $\alpha$ WT were performed in previously optimised and common conditions. The HSQC spectra of both the proteins showed well dispersed resonances that suggest the proper folding of the protein.

Superposition of <sup>1</sup>H-<sup>15</sup>N HSQC spectra of <sup>15</sup>N-hRAR $\alpha$ S369E and <sup>15</sup>N-hRAR $\alpha$ WT enlightened significant shifts in the resonances of several residues as showed (Fig.18, black boxes) but identifying the shifting residues requires the assignment of the protein backbone.

Measurements of <sup>15</sup>N relaxation rates of the proteins were used for obtaining information about the dynamics of hRAR $\alpha$ S369E and hRAR $\alpha$ WT in solution.

Spin relaxation comprises a series of processes leading magnetization to decay over time and is composed by two components: T1 and T2 (commonly we refer to the inverse of the T1 and T2 as the relaxation constant rates, R1 and R2 respectively). T1 is the longitudinal relaxation time and it corresponds to the restoration of equilibrium between the numbers of nuclei in the high and low energy spin states. T2 represents the time needed to re-gain phase coherence of the spins.

Specific pulse sequences gather structural information by transferring magnetization from one nucleus to another and allowing magnetization to evolve over time. <sup>15</sup>N relaxation time is physically correlated to the overall rotational correlation time of the protein (tumbling rate), as well as to the internal flexibility of the polypeptide chain. The most important factor influencing T2 is the tumbling rate of the protein which is a function of the size of the protein. Protein tumbling generates oscillating magnetic fields. Slow tumbling generates fields in the right frequency range to interact with the fields generated by the magnetic nuclei allowing an exchange of energy and fast relaxation times. On the contrary, fast tumbling determine long T2. Because relaxation time has to be long enough to allow the recovery of a useful signal, nuclear magnetic resonance is limited to the study of small molecules. Relaxation measurements performed onto the <sup>15</sup>N-hRAR $\alpha$ S369E and <sup>15</sup>N-hRAR $\alpha$ WT enlightened that the large

majority of the analyzed residues feature very similar relaxation rates except for ten residues in the hRAR $\alpha$ S369E which have higher longitudinal rates with respect to the hRAR $\alpha$ WT. This result indicates residue-specific differences in the dynamics of the two proteins. Relaxation rates are consistent with a protein in a monomeric state possessing a correlation time of 17 ns.

## 16. Conclusions and Perspectives

Recent studies highlighted that nuclear receptors (NRs) mediated transcription is regulated not only by the ligand binding but also through post-translational modification events such as phosphorylation. Deregulation of receptor phosphorylation leads to several types of cancers such as breast, ovarian and prostate cancer but up to date nothing is known about NR regulation through phosphorylation and no structure of phosphorylated nuclear receptor is available.

The human retinoic acid receptor gamma (hRAR $\gamma$ ) belongs to the NR superfamily and is phosphorylated at the N-terminal domain by the TFIID associated form of CAK and at the ligand binding domain by the cyclic AMP-dependent protein kinase (cAMP-PKA). Phosphorylation at the ligand binding domain by PKA enhances the binding of Cyclin H, a subunit of the CAK, to the receptor and therefore facilitates the phosphorylation of hRAR $\gamma$  domain by CAK and the transcription of cognate target genes. We aimed at providing new structural insights to understand at atomic level the impact of PKA phosphorylation at RAR ligand binding domain on the structure of RAR LBD.

We expressed, purified and crystallized the ligand binding domain of hRAR $\gamma$ S371E in which the mutation is well known to mimic a phosphorylation event. The X-ray crystal structure of hRAR $\gamma$ S371E (LBD) was solved at 1.7 Å. Superposition of hRAR $\gamma$ S371E (LBD) and hRAR $\gamma$  WT (LBD) crystal structures enlightened that the two structures are isomorphous with an r.m.s.d value, calculated onto the carbon alphas of the backbone, of 0.47 Å. Our X-ray crystallography structure details enlightened structural differences between RAR wild type and mutant, located at the level of the Cych docking site.

Analysis of the r.m.s.d values showed that one of the two regions displaying higher r.m.s.d values involves residues from 340 to 343, located in the region of receptor that is responsible for the recognition and binding of Cych (loop 8-9). A remarkable difference in the orientation of the side chains also involves the region of the receptor responsible for Cych binding. This difference is located at the level of the residue ASP340 which in hRAR $\gamma$ S371E, but not in RAR $\gamma$ WT, forms a salt bridge with ARG387 (H10). Our results demonstrated that the effects of the phosphorylation on the residue SER371 of RAR, mimed by the mutation S371E, do not affect profoundly the structure of RAR but they could finely tune the structural dynamics of the receptor. Our hypothesis were confirmed by molecular dynamic simulations, performed on the basis of the crystal structures of the wild type and mutant receptor, which highlighted for the first time a dynamic interaction between the mutation region of RAR and the Cych docking site.

To understand whether the mutation at S371 of RAR $\gamma$  is responsible for the structural modifications localized at the Cych binding site of the receptor we performed NMR mobility studies on both RAR WT and mutant.

Differences in the dynamics of wild type RAR and its mutant were pointed out through  $^{15}\text{N}$  longitudinal and traverse relaxation rates measurements performed onto the  $^{15}\text{N}$  isotopically enriched proteins. The large majority of the analyzed residues featured very similar relaxation rates when comparing hRAR $\alpha$  wild type to the mutant but a few residues in hRAR $\alpha$  mutant showed higher longitudinal relaxation rates with respect to the wild type. To analyze in greater details the difference in mobility between RAR wild type and its mutant protein backbone assignment is required. The power of NMR results from the fact that the overlapping signals originated by biological macromolecules can be resolved by multi-dimensional NMR techniques. This becomes more difficult for larger macromolecules (more than 30 KDa) and establishes a limit to the molecular size of the molecules that can be studied by NMR. In these cases protein samples enriched with NMR active isotopes  $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^2\text{H}$  are required. RAR molecular weight is approximately of 28KDa. Therefore a more detailed analysis of the difference in mobility between RAR $\alpha$  wild type and its mutant, which needs spectra assignment of the protein backbone, will most probably require the production of protein samples enriched with  $^2\text{H}/^{13}\text{C}/^{15}\text{N}$  isotopes.

## **Chapter 4 : Other Results**

## 17. Biophysical characterization of the interaction between hCycH and hRAR $\alpha$

### 17.1 Surface Plasmonic Resonance (SPR)

Surface Plasmonic Resonance experiments were initially performed to study the interaction between hCycH and hRAR $\alpha$  and monitor the effects of the mutation S369E on the binding.

The hRAR $\alpha$ , hRAR $\alpha$ S369E, hRXR $\alpha$  ligand binding domains were purified in presence of the ligand 9-cis retinoic acid and immobilized onto a four channel biacore chip by means of the GST fusion tag.

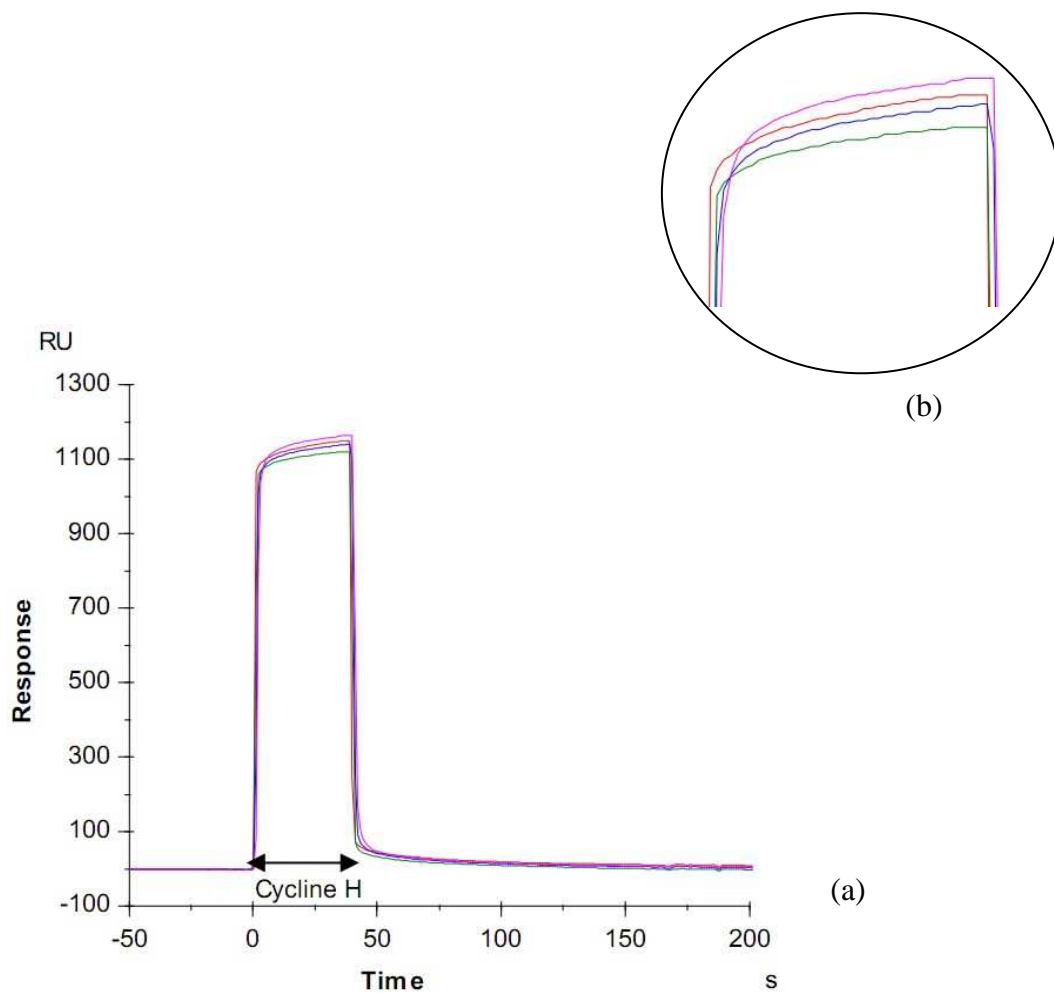
Either RXR $\alpha$  or pure GST should not provide any specific interaction with hCycH and were therefore taken as negative controls.

The LBD domains were stably captured on the chip as confirmed by the rapid association and dissociation kinetics. A concentrated sample of hCycH (115 $\mu$ M) was then passed over the four channels at the same time.

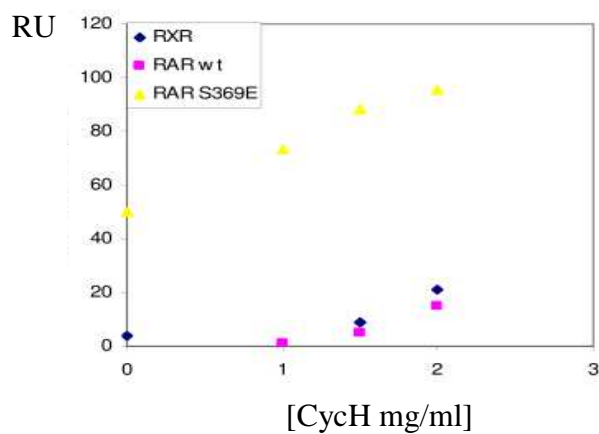
Considering the GST signal as baseline, a positive signal of 15 RU (1 RU, Response Unit, corresponds to 1pg mm<sup>-2</sup> of bound protein) was measured for the interaction between RARS369E (LBD) and hCycH, corresponding to a binding affinity in the  $\mu$ M range. No specific signal was detected, instead, for the interaction between hRAR $\alpha$  or RXR $\alpha$  and hCycH, as shown by the binding profiles (Fig.24)

In later experiments an hCycH concentration-dependent signal confirmed the specificity of hCycH-hRAR $\alpha$ S369E interaction (Fig. 25).

SPR results were consistent with our predictions and indicated that hCycH interacts specifically with hRAR $\alpha$ S369E but not with hRAR $\alpha$  or hRXR $\alpha$ .



**Fig. 24-** (a) Binding profiles of the interaction between the immobilized hRAR $\alpha$ S369E (purple), hRAR $\alpha$ WT (blue), hRXR $\alpha$  (green), LBDs or GST (red) and hCycH (115 $\mu$ M). In (b) a zoom of the plateau region.



**Fig. 25-** Immobilized GST- hRAR $\alpha$ S369E (yellow), hRAR $\alpha$ WT (purple), hRXR $\alpha$  (blue), LBDs titrated with increasing concentration of hCycH (0-60  $\mu$ M)



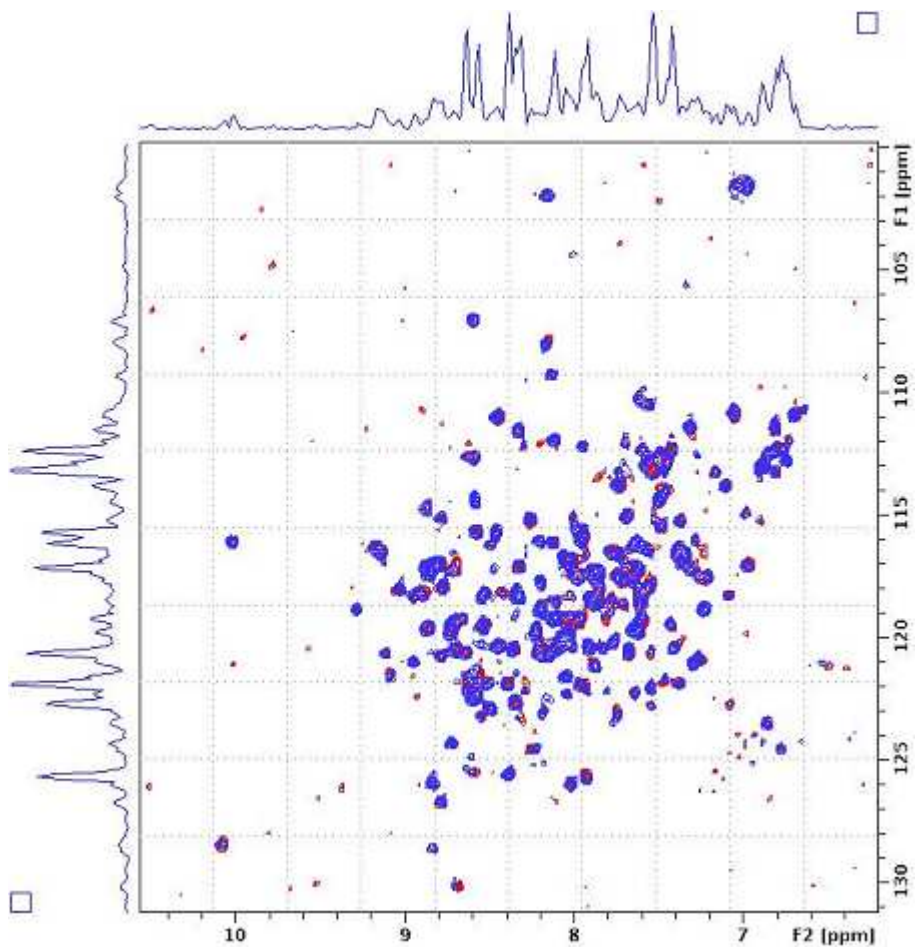
## 17.2 Preliminary study of the interaction of RAR $\alpha$ with PKA through NMR

The RAR $\alpha$  WT (LBD) was incubated with the phosphorylation buffer containing ATP and Mg<sup>2+</sup> at pH 6.5. A 1H-15N HSQC was acquired on a Bruker Avance 800 MHz spectrometer, at 303K. (Fig.26, red peaks)

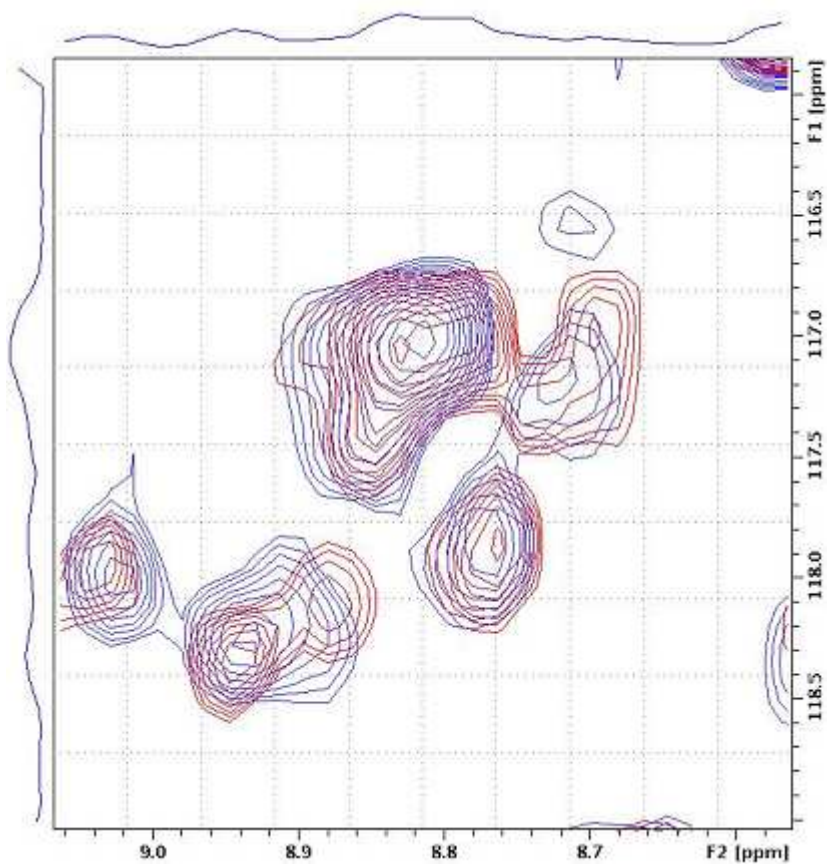
The protein sample was subsequently incubated in presence of PKA kinase and phosphorylation of RAR $\alpha$  at position S369 was checked by immunoprecipitation. <sup>1</sup>H-<sup>15</sup>N HSQC spectra were acquired on the phosphorylated RAR $\alpha$  WT (Fig.26, blue peaks).

Upon addition of PKA to the sample, we enlightened chemical shift variations for several residues (Fig.26, black boxes and Fig.27).

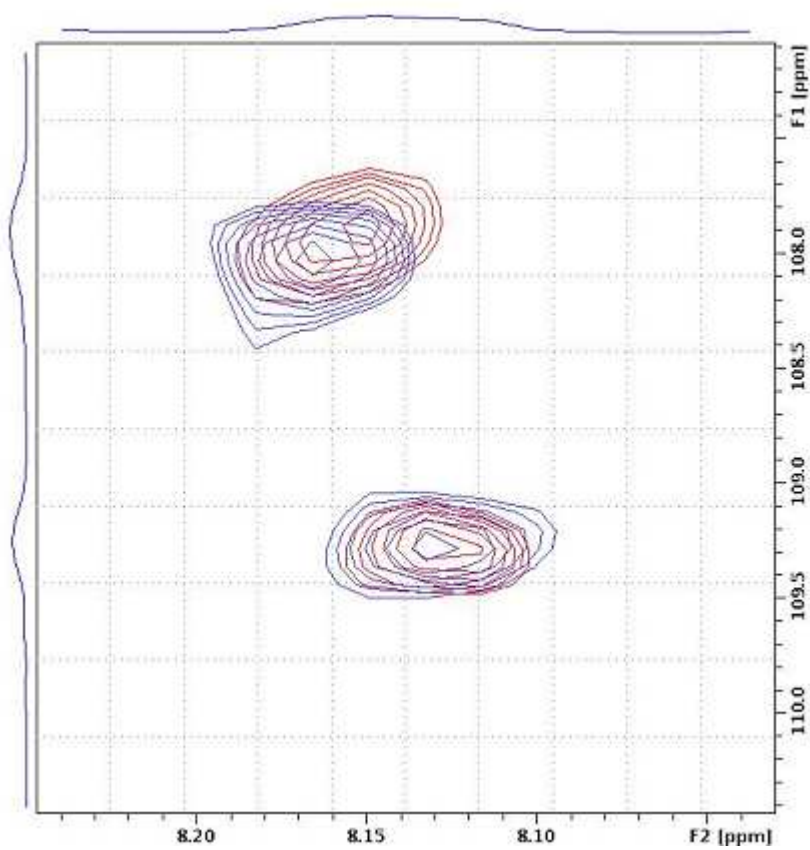
Identification of the residues resonating at different values of the chemical shift will need protein backbone assignment.



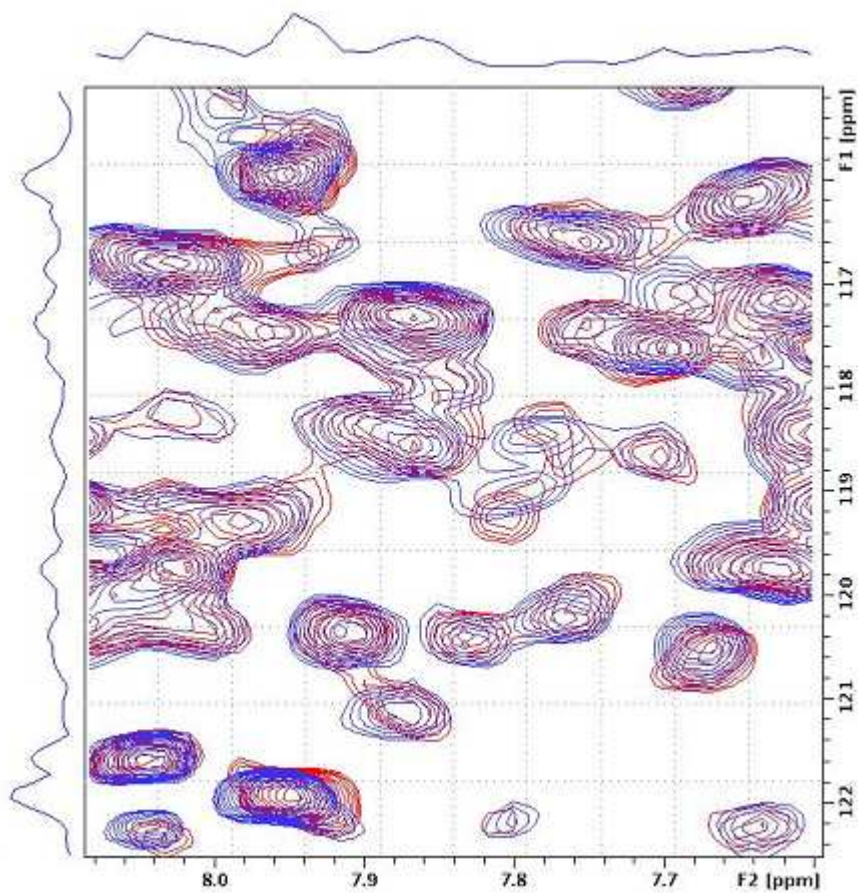
**Fig.25-** Superposition of <sup>1</sup>H-<sup>15</sup>N HSQC spectra of RARαWT (in red) and RARαWT after incubation with PKA (in blu). The spectra were acquired on a Bruker Avance 800 MHz spectrometer, at 303K.



(a)



(b)



(c)

**Fig-26** – (a), (b), (c): Different regions of the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra showing the superposition between RAR $\alpha$ WT (in red) and RAR $\alpha$ WT after incubation with PKA (in blue). The three images enlighten chemical shift variations for several residues.

## **Chapter 4 : Material and Methods**

## 18. Materials

### 18.1 Molecular biology

#### 18.1.1 Plasmid Amplification and Purification: DNA Mini Prep

The purification of the plasmids was performed with “NucleoSpin plasmid” kit from Macherey-Nagel. *E. coli* (DH5 $\alpha$ ) cells were transformed with the plasmids containing the genes of interest and 5 ml over-night cultures were set up. Cells from these cultures were collected by centrifugation and resuspended in 250  $\mu$ l buffer A1. Two hundred fifty  $\mu$ l buffer A2 was added inverting the tube four times and the suspension incubated at room temperature for 5 min. Three hundred  $\mu$ l buffer A3 was added gently inverting the tube four times. The suspension was centrifuged at 11000 rpm for 10 min. The supernatant was transferred into the NucleoSpin plasmid column and centrifuged for 1 min at 11000 rpm. The column was washed with 600  $\mu$ l buffer A4 and centrifuged at 11000 rpm for 1 min. The flowthrough was discarded and the membrane dried up by centrifuging for 2 min at 11000 rpm. Fifty  $\mu$ l sterile H<sub>2</sub>O was added and the plasmid eluted by centrifugation at 11000 rpm for 1 min.

#### 18.1.2 LR-Gateway reactions

Gateway Technology provides a high throughput technology to transfer simultaneously the gene of interest into different cloning vectors. The Gateway cloning technology is based onto a two step reaction. In the first step (BP reaction) the gene of interest, amplified with primers flanked by attB sites, recombines with a donor vector containing attP sites. The resulting entry vector contains the gene of interest flanked by attL sites and possesses resistance to the antibiotic kanamycin. In the second step (LR reaction)

the entry vector is mixed with the destination vector containing attR sites. Recombination between attL and attR sites occurs in presence of Gateway clonase enzyme and generates the expression vector containing the gene of interest and resistance to ampicilline (Invitrogen, Gateway Technology Manual, 2003).

Starting from pDON207 entry vectors containing respectively hRAR $\alpha$ , hRAR $\alpha$ S369E and hRXR $\alpha$  (LBDs) genes, LR recombination reactions were performed in presence of pGex-4T3-Gateway destination plasmid. One  $\mu$ l from each LR reaction was transformed in *E. coli* DH5 $\alpha$  cells and positive clones were selected onto solid medium in presence of antibiotic.

One colony from each transformation was transferred into 100  $\mu$ l LB in presence of antibiotic and incubated for 3h at 37° C under constant agitation. One and a half  $\mu$ l of the respective precultures was used to test the efficiency of the LR reactions by PCR.

### 18.1.3 Polymerase chain reaction (PCR)

#### 18.1.3.1 Point mutation at position 369 of RAR $\alpha$ and position 371 of RAR $\gamma$ (LBDs)

In vitro site-directed mutagenesis at position 369 of RAR $\alpha$  LBD, exchanging a serine with a glutamic acid, was carried out in a single PCR step. The pDon207 vector containing the gene of RAR $\alpha$  LBD was used as a template and the following set of primers:

5'- CGGAAGCGGAGGCCCGAACGCCCCACATGTTC

5'- GAACATGTGGGGGCGTTCGGGCCTCCGCTTCCG

were respectively used as forward and reverse primers. Table 5 describes the composition of one PCR reaction.

<b>Reagent</b>	<b>Volume (<math>\mu</math>l)</b>
Plasmid DNA (17 ng/ $\mu$ l)	0.6
Polymerase buffer HF (5x)	5
dNTP (2mM each)	10
Forward primer UP1 (100 nM)	3
Reverse primer GWR (100 nM)	3
DMSO	5
H <sub>2</sub> O	Up to 49
PFU Polymerase	1

Table 5- Composition of one PCR reaction

The PCR program that was used is reported below:

30 sec at 98° C

7 sec at 98° C  
 20 sec at 55° C  
 15sec/Kb at 72° C

} 30 Cycles

7 min at 72° C

Sequence analysis of the resulting plasmid was performed to ensure that the desired mutation was successfully introduced.



### 18.1.3.2 Determination of the efficiency of the LR reaction

Polymerase chain reaction (PCR) was carried out to test the efficiency of the LR Gateway cloning step of RAR $\alpha$ , RAR $\alpha$ S369E and RXR $\alpha$  (LBDs) genes into pGex-4T3-Gateway expression vectors.

The reagents listed in table 6 were mixed to perform one PCR reaction.

<b>Reagent</b>	<b>Volume (<math>\mu</math>l)</b>
Preculture	1.5
Polymerase buffer 10x	2
dNTP (2mM each)	2
Forward primer SP136 (10 $\mu$ M)	1
Reverse primer GWR (10 $\mu$ M)	1
H <sub>2</sub> O	Up to 19.5
Taq Polymerase	0.5

Table 6- Composition of one PCR reaction

The PCR reaction was carried out according to the program below and two  $\mu$ l of the reaction were analysed by agarose gel electrophoresis.

4 min at 94° C

1 min at 94° C  
45 sec at 55° C  
1 min/Kbp at 72° C

} 30 Cycles

5 min at 72° C

## 18.2 Protein Expression

### 18.2.1 Cell Transformation

A hundred ng of the expression vector was transformed into 50  $\mu$ l of thermo competent *E. coli* BL21 (DE3) cells for protein expression or DH5 $\alpha$  for plasmid purification. The cells were incubated on ice for 30 min and then osmotic shock was carried out for 90 seconds at 42°C, followed by incubation on ice for 2 min. Two hundred 200  $\mu$ l LB-medium was added and the cells incubated at 37°C for 1 hour under constant shaking at 250 rpm.

The cells containing the gene of interest were selected onto LB-agar plates in presence of the correspondent antibiotic. Thirty or eighty  $\mu$ l of cell suspension was plated out respectively from BL21 (DE3) or DH5 $\alpha$  transformations in order to obtain isolated colonies. The plates were incubated overnight at 37°C.

### 18.2.2 Small Scale Protein Expression Tests

One single colony from the LB-agar plate was inoculated in 5 ml of LB-medium in presence of antibiotic and incubated at 37°C overnight shaking at 250 rpm. Twenty-five ml LB-medium containing antibiotic were inoculated with of the overnight culture to have a starting OD<sub>600</sub> of 0.15. The culture was incubated at 37°C shaking at 250 rpm till OD<sub>600</sub> value reached 0.6. Protein expression was induced with 0.8 mM IPTG (Isopropyl-beta-thio-galactoside). The cultures were incubated for four hours at 25° C shaking at 250 rpm. Five hundreds  $\mu$ l of culture was collected before induction, as a negative control for protein over-expression. The cells from this aliquot were collected, resuspended in 100  $\mu$ l 2x Laemli buffer and heated to 100° C for 10 min. Ten  $\mu$ l was loaded on the 15% SDS-page.

### 18.2.3 Small Scale Protein Purification Test

Cells from the 25 ml culture were collected by centrifugation at 4000 rpm for 10 minutes. The cell pellet was resuspended in 1.5 ml lysis buffer (T20mM, NaCl500 mM) and 150  $\mu$ l Lysozyme (25 mg/ml) was added. The pellet was homogenized and the suspension incubated on ice for 15 min and put at 37° C for 5 min. Sonication was carried out 3 times for 30 seconds at 1 min intervals with a 3 mm tip sonicator at amplitude of 20%. The suspension was centrifuged for 10 min at 14000 rpm and the supernatant recovered.

Fourty  $\mu$ l Talon or Nickel Sepharose metal chelating resin or 30  $\mu$ l Glutathione Sepharose 4B resin was washed with deionised water. The suspension was centrifuged at 1000 rpm for 2 min and the beads equilibrated with 200  $\mu$ l lysis buffer for four times. At each step the supernatant was discarded upon centrifugation. The clarified lysate was added to the resin and incubated respectively for 30 min (Talon or Nickel Sepharose) and 1 hour (Glutathione Agarose). The resin was collected by centrifugation at 1000 rpm for 2 min and washed 5 times with 200  $\mu$ l of lysis buffer, the supernatant discarded. For His-tagged hRARs, hRXR, hCycH and PKA 50  $\mu$ l elution buffer (T20mM, NaCl500 mM, Imidazole 1M) were added to the resin, the resin centrifuged at 1000 rpm for 2 min and the supernatant collected. Three  $\mu$ l of the eluted protein was mixed with 6  $\mu$ l 2x Laemli buffer and loaded on SDS-PAGE. For GST-tagged hRARs and hRXR, 100  $\mu$ l 2x Laemli buffer was added to the resin and incubated at 100° C for 10 min. The resin was spun down by centrifugation and three  $\mu$ l of the suspension was loaded on 15% SDS-PAGE.

### 18.2.4 Over-night Precultures

One single colony was added to 500 ml pre-autoclaved LB-medium in presence of the required antibiotic. The culture was incubated over-night at 37° C shaking at 200 rpm.

### 18.2.5 Large Volume Cultures

The components for one litre of LB-Home were dissolved in 800 ml of water in a 5 liter Erlenmeyer flask and autoclaved. Sterile sucrose was successively added up to one litre. The adapted volume of overnight preculture was added in presence of antibiotics to the 1 litre culture in order to obtain an initial value of OD<sub>600</sub> of 0.15. The culture was incubated at 37° C and shaken at 250 rpm until OD<sub>600</sub> reached the value of 0.6. Protein expression was induced with 0.8 mM IPTG and culture incubated at 25° C for four hours. Cells were harvested by centrifugation, at 4000 rpm for 20 min. Pellet was resuspended in 25 ml fresh LB and centrifuged for 20 min at 4000 rpm. The supernatant was discarded and the pellet stored at -20° C.

### 18.2.6 Large Volume Cultures Isotopically Enriched With <sup>15</sup>N

The chemically defined minimal medium M9, isotopically enriched with the isotope <sup>15</sup>N, was used to express <sup>15</sup>N labelled hRAR $\alpha$  and hRAR $\alpha$ S369E. Five hundred ml of minimal media was prepared in a 5 liter Erlenmeyer flask, added with the required antibiotic. The volume of overnight preculture was added in order to obtain an initial value of OD<sub>600</sub> of 0.15. For the composition of the M9 minimal media see Table 7.

<b>Chemical</b>	<b>1 Litre Culture</b>
M9 5X – NH <sub>4</sub>	200 ml
Glucose 20%	20 ml
( <sup>15</sup> NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	1.25 g
MgSO <sub>4</sub> 1M	2 ml
CaCl <sub>2</sub> 1M	100 $\mu$ l

FeSO <sub>4</sub> 3.5 mM	100 µl
Thiamine 10 mg/ml	1 ml
Biotine 10 mg/ml	1 ml
Vitamines T.t	1 ml
H <sub>2</sub> O	Up to 1 litre

<b>M9 5X Solution</b>	<b>1 Litre Culture</b>
Na <sub>2</sub> HPO <sub>4</sub> (H <sub>2</sub> O) <sub>12</sub>	85g
KH <sub>2</sub> PO <sub>4</sub>	15g
NaCl	2.5g

Table 7- Composition of media

## 18.3 Protein Purification

### 18.3.1 Sonication and Lysate Clarification

To one litre pellet, 25 ml lysis buffer containing 1.5 ml of Lysozyme (25mg/ml) and 1 tablet of the Protease inhibitor Cocktail PIC (EDTA free Sigma) was added and the pellet resuspended. The pellet was homogenised with a Turax mixer and then incubated on ice for 15 min. The suspension was warmed up in a water bath at 37° C for 5 min before sonication. Cells were sonicated on ice for 3 min in continuous with amplitude of 20%. Twenty-five ml of fresh lysis buffer was added and the 3 min sonication repeated as previously. Soluble protein was extracted by ultra-centrifugation for 50 min at 45000 rpm in a Beckman L-70 ultracentrifuge, 50.2 Ti rotor. Cell debris was discarded and the supernatant filtered through a Millipore filter.

## 18.4 Metal Affinity Chromatography

### 18.4.1 Ni<sup>2+</sup> affinity chromatography

A 5 ml HisTrap FF crude column (GE Healthcare) Ni<sup>2+</sup> affinity chromatography column was prepared for the purification of His-tagged proteins from 2 litres of culture. The column was washed with 10 column volumes (CV) of deionised water and then charged with nickel with 10 CV of NiSO<sub>4</sub>. The column was finally washed with 10 CV of deionised water and equilibrated with 10 CV of the required binding buffer until conductivity and UV absorbance at 280 nm reached a stable baseline.

The clarified and filtered protein soluble extract was loaded onto the column at a flow rate of 5 ml/min and the flowthrough was collected to determine binding efficiency of the protein on a 15% SDS-PAGE gel. The FPLC apparatus (Bio-Rad) was programmed according to the following protocol (Table8):

<b>BB Buffer (%)</b>	<b>EB Buffer (%)</b>	<b>Imidazole concentration</b>	<b>RunVolume (CV)</b>	<b>Flowrate (ml/min)</b>	<b>Fractionsize (ml)</b>
100	0	5mM	30	5	6
95	5	50 mM	30	5	6
85	15	150 mM	20	5	3
70	30	300 mM	10	5	6
0	100	1M	10	5	6

Table 8- Protocol for metal affinity chromatography (adapted for 2 litres of culture)

The concentration of the protein was determined by the protein colorimetric Bradford assay (Bio-Rad). A volume corresponding to 1 µg of protein was loaded onto 12.5%

SDS-PAGE to determine the purity of the eluted protein for each elution fraction. The purified protein pool was dialyzed against the lysis buffer to eliminate imidazole and then concentrated.

#### 18.4.2 Co<sup>2+</sup> affinity chromatography

A 7.5 ml TALON Metal Affinity column was packed for the purification of His-tagged hCycH from 3 litres of culture. The column was washed with 10 column volumes (CV) of deionised water and then equilibrated with 10 CV of the required binding buffer until conductivity and UV absorbance at 280 nm reached a stable baseline.

The clarified and filtered protein soluble extract was loaded onto the column at a flow rate of 5 ml/min and the flow trough was collected to determine binding efficiency of the protein on SDS-PAGE gel. The FPLC apparatus (Bio-Rad) was programmed according to the following protocol (Table9):

<b>BB Buffer (%)</b>	<b>EB Buffer (%)</b>	<b>Imidazole concentration</b>	<b>Run Volume (CV)</b>	<b>Flowrate (ml/min)</b>	<b>Fraction size (ml)</b>
99	1	10mM	30	5	6
99-75	1-25	10-250 mM	12	5	3

Table 9- Protocol for metal affinity chromatography (adapted for 3 litres of culture)

The concentration of the protein was determined by the protein colorimetric Bradford assay. A volume corresponding to 1 µg of protein from each elution fraction was loaded onto a 12.5% SDS-PAGE to determine the purity of the eluted protein. The purified protein pool was dialyzed against the lysis buffer to eliminate imidazole and then concentrated according to experimental requirements.

### 18.4.3 GST-tag Fusion RAR

Five ml Glutathione Agarose 4B resin was used for the purification of GST-tagged protein from 2 litres cultures. The resin was washed with 10 CV of deionised water and then equilibrated for 2 times with 50 ml binding buffer. After each washing step, the material was centrifuged for 2 min at 1000 rpm and the supernatant discarded. Clarified and filtered protein soluble extract was added to the resin and allowed to incubate for 60 min at 4° C with slow and continuous mixing. The resin was recollected by centrifugation at 1000 rpm for 5 min, and the supernatant was collected to determine binding efficiency of the protein on SDS-PAGE gel. Roughly 5 ml supernatant was let on top of the resin, gently resuspended and transferred into an FPLC column. Ten ml lysis buffer was added to the column, the resin was allowed to settle and the column was sealed. The column was loaded onto a FPLC apparatus (Bio-Rad) and the non-specifically bound proteins were washed from the column with GST lysis buffer until UV absorbance at 280 nm reached a stable baseline.

GST fusion enzymatic cleavage was carried out with on-column thrombin digestion. Glutathione Agarose 4B resin has a protein binding capacity of 8 mg/ml of resin. Sixty mg of pure protein could be purified with a 7.5 ml column. Considering the dead volume of tubes and other mechanic parts connected to the column, 8.5 ml lysis buffer containing 1.2 unit/ mg protein of Thrombin (1U/  $\mu$ l) and 5 mM CaCl<sub>2</sub> was prepared. The column was removed from the FPLC apparatus and the thrombin mixture slowly injected into the column. The column was sealed and incubated at 14° C for 30 hours. Elution of proteins cleaved from thrombin was carried out in 1 ml fractions on the FPLC apparatus with the lysis buffer, until the UV absorbance at 280 nm reached a stable baseline. The purified protein pool was concentrated up to the desired concentration. When GST fusion was not required the not pecifically bound proteins were washed from the column with lysis buffer, and the GST-fused protein was eluted with the lysis buffer, supplied with 10-mM glutathione solution and adjusted to the required pH.



#### 18.4.4 Size Exclusion Chromatography (Gel Filtration)

The proteins were concentrated up to 4 mg/ml using a Centriprep centrifugal filtration device (10 KDa cutoff) and then centrifuged for 10 min at 14000 rpm. A maximum of 4 ml per run of concentrated protein were injected onto the gel filtration column.

Pharmacia HiLoad 16/60 Superdex 75 gel filtration column was equilibrated with 2 CV of gel filtration buffer. The concentrated protein was injected into a 5 ml loop and then automatically loaded on the column. The purification was carried out with 2 CV of gel filtration buffer at flow rate of 1 ml/min. The eluted protein was collected in 2 ml fractions.

Aliquots corresponding to 1 µg of protein were loaded on 12.5% SDS-PAGE to determine the purity of the eluted protein fractions. The purified protein was concentrated to the required concentration using a Centriprep or Centricon centrifugal filtration device.

#### 18.5 Buffer exchange of RAR and RXR

Buffer exchange was performed through dialysis using a membrane with 10 KDa cut off against the required buffer under constant stirring at 4° C. Two changes of the dialysis buffer were carried out at 6 hours interval.

Small volumes samples of concentrated protein were buffer exchanged using Illustra NAP-5 Columns filled with Sephadex G-25. The column was equilibrated with 10 ml buffer. A maximum volume of 500 µl of concentrated protein was added to the column. If the protein volume was less than 500 µl a required volume of buffer was added up to 500 µl upon protein absorption into the column. The buffer exchanged protein was eluted with 1 ml buffer and collected in 3 x 300 µl fractions.

### 18.5.1 Buffers for : His-RAR $\gamma$ S371E and RAR $\gamma$ WT

Lysis: T20N500I5 (mM) pH8.0

Binding: T20N500I5 (mM) pH8.0

Elution: T20N500I1M (mM) pH8.0

Gel filtration: T20N500Chaps2C12M0.15 TCEP1 (mM) pH 8.0

### 18.5.2 Buffers for : GST-RAR $\alpha$ S371E and RAR $\alpha$ WT

Lysis: T20N5005I (mM) pH8.0

Elution: T20N500Chaps2 TCEP1 (mM) pH 8.0

Gel filtration: T20N500Chaps1TCEP1 (mM) pH 6.5

### 18.5.3 Buffers for : His-PKA

Lysis: T20N5005I (mM) pH8.0

Elution: T20N500I1000 (mM) pH8.0

### 18.5.4 Buffers for : His-CycH

Lysis: T50N500 $\beta$ 2.5 (mM) pH8.0

Binding: T50N500 $\beta$ 2.5 (mM) pH8.0

Elution : T20 mM N500 $\beta$ 2.5 I1000 (mM) pH8.0

Gel filtration: T50N500 $\beta$ 2.5 (mM) pH8.0

**Abbreviations:** T= Tris HCl, N= NaCl,  $\beta$ =  $\beta$ -Mercaptoethanol, I= Imidazole, Chaps = 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate, TCEP= tris(2-carboxyethyl)phosphine, DTT= Dithiothreitol

### 18.5.5 Phosphorylation buffer for RAR $\alpha$ WT

Molar ratio (protein:PKA)=200:1

[ATP]=2mM

Buffer: T50N250(MgCl<sub>2</sub>)20 DTT1 (mM) pH 6.5

### 18.6 SDS-PAGE Denaturing Gel Electrophoresis

One dimension electrophoresis was performed on denaturing SDS-PAGEs to check protein purity. All the constituents found in the table 10 were mixed according to the desired concentration of acrylamide in the gel.

The gel was cast, and the solution for the resolving gel loaded between the two gel plates. Water was loaded on top to avoid formation of air bubbles within the resolving gel. The stacking gel was added upon polymerisation of the stacking gel and combs containing the suitable number of wells were inserted between the gel plates. The protein samples were loaded and electrophoresis was carried out at 300 Volts for 40 minutes.

	<b>10%</b>	<b>12%</b>	<b>15%</b>	<b>17%</b>	<b>Stacking gel</b>
Water	6.1 ml	5.4 ml	4.1 ml	3.9 ml	3.8 ml
Tris pH 8.8	3.15 ml	3.15 ml	3.15 ml	3.15 ml	
Tris pH 8.8					1.6 ml
Acrylamide 40%	3.1 ml	3.9 ml	4.7 ml	5.3 ml	0.8 ml
APS 10%	125 $\mu$ l	125 $\mu$ l	125 $\mu$ l	125 $\mu$ l	125 $\mu$ l
Temed	12 $\mu$ l	12 $\mu$ l	12 $\mu$ l	12 $\mu$ l	8 $\mu$ l

Table 10- Composition of SDS PAGES at different concentrations.

## 19. Methods

### 19.1 X-ray crystallography

#### 19.1.1 Basics

The resolution limit (LR) of any optical method is defined by the equation:

$$\text{Eq: 1} \qquad \text{LR} = \lambda/2$$

and is a function of the wavelength ( $\lambda$ ) of the incident radiation<sup>47, 172</sup>. According to this equation common microscopes that use visible light (whose wavelength range between 400 and 700 nm) cannot produce an image of individual atoms for which inter-atomic distances are only about 0.15 or 1.5Å.

X-rays are in the correct range to visualize atoms, and for this reason they are used to solve the atomic structures of proteins. However it is not possible to achieve a focused image of a single molecule for two reasons.

First, the X-rays cannot be focused with lenses because they have refraction index close to one. This problem is overcome by crystallographers by measuring the position and intensities of the diffracted X-ray and then using a computer to simulate an image-reconstructing lens.

Secondly, a single molecule would scatter very weakly and the diffracted beams would be too weak to be analysed. Using protein crystals provides a good solution to the problem. In a protein crystal single molecules display a high degree of order and are arranged in an identical orientation. Therefore, in a crystal, scattered waves can add up in phase producing strong X-ray beams.

A crystal behaves like a three-dimensional grate which diffracts the X-ray beam into discrete beams and gives rise to both constructive and destructive interference. Beams that interact constructively sum up and produce distinct spots on the detector, called reflections in the diffraction pattern. Each reflection contains information on all atoms in the molecule.

W. L. Bragg showed that the beams scattered from a crystal can be treated as if they were originated by the reflection from equivalent and parallel planes of atoms in the crystal, each plane identified by the indices (h, k, l). According to Bragg's description a set of parallel planes with indices (h, k, l) and interplanar spacing  $d_{hkl}$  produces a diffracted beam when X-rays, of wavelength  $\lambda$ , enters and is reflected from the surface of the crystal with an angle  $\theta$  that meets the condition:

$$\text{Eq:2} \qquad 2d_{hkl} \sin\theta = n\lambda$$

where n is an integer. Thus for two waves to be in phase with the wave which reflects from the surface, one wave have to travel a whole number of wavelengths while inside the crystal<sup>48</sup>. For other angles of incidence  $\theta'$ , for which Eq. 2 does not hold, waves reflected by successive planes are out of phase and thus interact destructively, reflecting no beam at that angle.

The space constituted by the crystal reflections is in a reciprocal relationship to respect to the real space constituted by the crystal and is therefore called "reciprocal space". Thus, the dimensions of the reciprocal space ( $\text{\AA}^{-1}$ ) are inversely proportional to dimensions of the real space ( $\text{\AA}$ ).

Each reflection in the reciprocal space the sum of electro-magnetic radiations and is therefore characterized by an amplitude and a phase. The calculation of the electronic density through Fourier transform requires the knowledge both of these parameters for each reflection<sup>147</sup>. Unfortunately, only information about amplitudes is accessible from

experimental measurements while no information is achievable about the phases. This is known as the “phase problem”.

If the structure of the protein target displays sequence homology higher than 30% to that of a known protein the two structures are similar and the known protein can be used as a phasing model (molecular replacement).

If no phasing model is available then the phases have to be obtained from alternative methods such as isomorphous replacement and anomalous scattering (MAD or SAD phasing). Both these methods use protein heavy atoms derivatives<sup>99, 167</sup>.

In isomorphous replacement a heavy atom as mercury, lead or gold soaked into the crystal binds to specific sites on the protein. In most cases heavy atoms derivatives are isomorphic with crystals of the native protein and displaying same unit cell dimensions and symmetry and same protein conformation. Nevertheless the introduction of the heavy atoms originates reflections that are slightly different from that of the native crystal and can be used to obtain estimates of the phases.

Anomalous scattering is based on heavy atoms ability to absorb X-ray at specific wavelengths. The X-ray wavelength is scanned past an absorption edge characteristic for each heavy atom, which changes the scattering in a known way<sup>167</sup>.

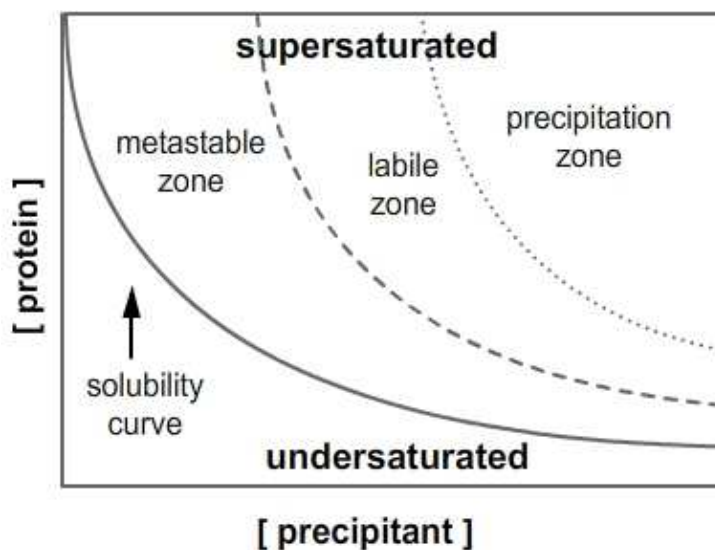
Once the amplitudes and the phases are known, the electron density map is calculated by Fourier transform. With a good quality electronic density map manual construction of the three-dimensional model can be carried on basing on the knowledge of the protein sequence. The obtained model is then submitted to iterative cycles of structure refinement.

### 19.1.2 Protein crystallization

Crystal formation occurs in two stages including (1) nucleation and (2) growth.

The phase diagram for crystallization (Fig.27) shows the solubility of a protein in solution as a function of the concentration of the precipitant and can be divided in four main zones<sup>6</sup>.

- 1) Precipitation zone. This region is characterized by a very high protein supersaturation. In this region the protein will precipitate.
- 2) Metastable zone. This region is characterised by very low protein supersaturation. In this region very slow spontaneous nucleation can occur.
- 3) Labile zone. This is the region of large supersaturation. Here spontaneous observable nucleation.
- 4) Region of protein unsaturation. In this region the protein is completely dissolved and neither nucleation nor crystal grow can occur.



**Fig. 27- The phase diagram.**

The phase diagram in protein crystallization schematically represents how the protein and precipitant concentrations are related. The phase diagram is divided into 4 different zones. In the Undersaturated zone protein crystallization can not occur because the protein is completely solubilized. The Precipitation zone is highly saturated in protein concentration. In this region the protein will precipitate and would not form crystals. The Labile zone (or nucleation zone) is the region where crystal nucleation can occur. As the crystals start to form, the protein concentration decreases leading the system from the labile to Metastable zone, where crystal growth can occur.

### 19.1.2.1 Nucleation

Nucleation arises from a solution supersaturated with protein.

Large supersaturation (labile zone) is required to overcome the activation barrier constituted by the free energy required to create microscopic clusters of proteins, the nuclei, from which the crystal will grow<sup>87</sup>. Very low supersaturation (metastable zone) will require an unreasonable nucleation time whereas too much high supersaturation (precipitation zone) will lead to protein precipitation.

Once nuclei have formed, the concentration of protein in the solute decreases leading the system to the metastable zone where crystal growth can occur. Persistence of nucleation conditions leads to the formation of many small crystals with poor diffracting quality<sup>36, 147</sup>.

### 19.1.2.2 Growth

Once nuclei have formed, the concentration of protein in the solute decreases leading the system to the metastable zone where crystal growth can occur. Persistence of nucleation conditions leads to the formation of many small crystals with poor diffracting quality<sup>36, 67, 147</sup>. Crystal growth requires different conditions than those of nucleation and it occurs in the metastable zone. Once a crystal has become larger than the initial nucleus, its growth becomes thermodynamically favoured, and the crystal can grow to macroscopic dimensions. Crystals grow by the attachment of molecules in layers to a growing surface and it occurs through protein supply either via diffusion or convection.



### 19.1.3 Parameters that influence Crystallization

The difficulty of obtaining well diffracting macromolecular crystals is due to the fact that protein crystallization depends on a large number of different variables. These parameters include physico-chemical variables such as pH, temperature, precipitating agents and also macromolecular variables such as protein concentration, stability, and homogeneity. Therefore it is almost impossible to predict the condition in which the protein will crystallize<sup>48</sup>. Recently developed nanodrop crystallization robots and commercial available crystallization screenings allow screening of hundred of variables with a limited amount of conditions and provide crystallographers with precious help in the difficult task of finding the best conditions for protein crystallization.

### 19.1.4 Crystallization Techniques

Several methods are employed to crystallize proteins including batch, micro-batch<sup>35</sup> and dialysis<sup>170</sup>. One widely used crystallization technique is vapour diffusion<sup>35</sup>. What follows is a brief introduction to the vapour diffusion method, the one that I used to crystallize RAR $\gamma$ S371E.

#### 19.1.4.1 The Vapour Diffusion Method

In the vapour diffusion method equilibrium is reached between two separated liquids. The protein drop is mixed with the crystallization buffer and suspended above a reservoir containing crystallization buffer. As the reservoir solution contains a higher precipitant concentration to respect to the protein drop, water slowly evaporates from the protein drop to achieve the same precipitant concentration as in the reservoir<sup>47</sup>. As water leaves the drop, the sample undergoes an increase in relative protein supersaturation.

The two main crystallisation techniques based on the vapour diffusion method are the sitting drop and the hanging drop. In the first one the protein drop sits in a concave depression on the top region of the rod placed in the sealed reservoir. In the hanging drop technique the protein drop is placed under a glass cover slip sealed on top of the reservoir.

#### 19.1.5 Crystals Cryo-protection.

Crystals often suffer from radiation damage caused by X-rays. X-ray radiation produce thermal vibration of the molecules, which create enough energy to break bonds between atoms. Further damages are caused by free active radicals, produced by X-ray damage of the polypeptide chain or production of H<sup>·</sup> or OH<sup>·</sup> from disruption of H<sub>2</sub>O molecules.

Reducing the temperature to -196°C, would allow decrease of the thermal vibrations therefore limiting the extent of the radiation damage<sup>65, 66</sup>. In order to freeze the protein, chemicals known as cryo-protectants such as sucrose, MPD, PEGs, ethylene glycol and glycerol have to be added to the crystal. Cryoprotectant act disrupting the local order of water and reducing ice crystal formation<sup>65</sup>. In presence cryoprotectants water can be frozen rapidly and non diffracting vitreous ice is formed with no damage for the crystal.

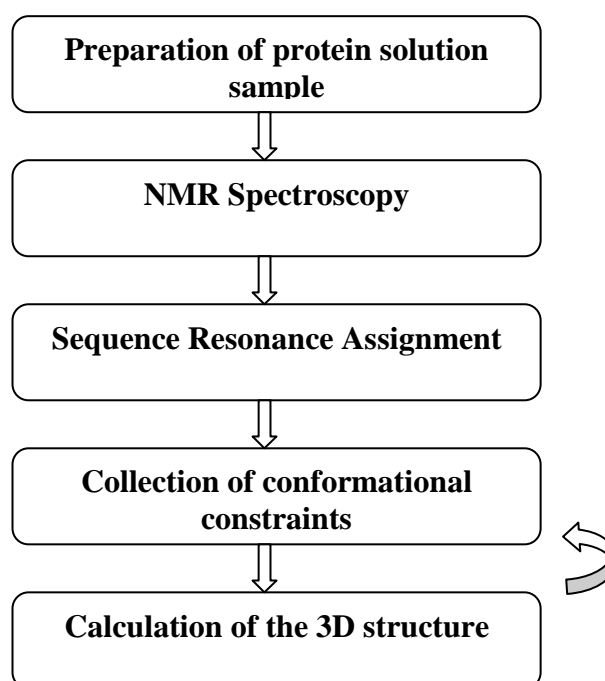
#### 19.1.6 Nuclear Magnetic Resonance (NMR)

NMR assesses the response of nuclear spins immersed in an intense and homogeneous magnetic field, to perturbations generated by the irradiation of sequence of radio pulses. The response to such pulse sequence is a sum of radio-frequencies emitted by the NMR active nuclei.

The signal emitted by the nuclei decays exponentially with a characteristic time constant, called transverse relaxation time T<sub>2</sub>. Fourier-transformation converts the signal that is originally in the time domain into the frequency domain. The result of this transformation is a spectrum containing resonances lines representing the radio-frequencies emitted by the nuclei.

The width of the resonance lines is inversely proportional to T2 which depends in turn on the molecular size of the molecule. Large molecules are characterized by fast relaxation rates (low values of T2) and determine therefore broad spectral lines representing the major limit of nuclear magnetic resonance spectroscopy.

NMR studies leading to structure determination of macromolecules is based onto a process that entails the following steps:



The first task to accomplish is the preparation of highly purified protein samples. Inhomogeneous preparations may hinder spectra acquisition or severely affect the quality of the data. Optimization of the experimental conditions such as pH, ionic strength, and temperature is therefore essential step to ensure the quality of the NMR data. In an ideal case the protein in the sample must be stable at room temperature for several weeks. NMR samples range to 0.3 to 0.5 ml of protein in H<sub>2</sub>O/D<sub>2</sub>O (90/10). Typical protein concentrations range from 0.5 to 2.0 mM.

NMR conventional one-dimensional spectra (1D) of biological macromolecules are far too complex for interpretation due to extensive signal overlapping. The invention of multidimensional spectra simplified one dimensional spectra by introducing additional spectral dimensions and therefore rendered accessible a lot extra information.

Structural studies and three dimensional structure determination of small proteins with molecular weight below 10 kDa are usually performed through the combination of 2D homonuclear  $^1\text{H}$  NMR experiments such as [ $^1\text{H},^1\text{H}$ ]-TOCSY (Total Correlation Spectroscopy), and [ $^1\text{H},^1\text{H}$ ]-NOESY (Nuclear Overhauser Effect) and allows the sequential assignment of most proton signals to individual protons. TOCSY experiments provide correlation between different nuclei via a system of  $J$  coupling<sup>7, 182</sup> allowing through-bond correlations across the peptide bond in proteins isotopically enriched with  $^{15}\text{N}$  and  $^{13}\text{C}$ . NOESY experiments measure through-space correlations via the nuclear Overhauser effect (NOE) and provide correlations between pairs of hydrogen atoms which might be far apart in the protein sequence but are close in space, separated by a distance within  $5 \text{ \AA}$ <sup>183</sup>.

Proteins with a molecular weight higher than 10 kDa give rise to extensive signal overlapping. This prevents complete assignments of all  $^1\text{H}$  signals in proton spectra. In this case heteronuclear 2D experiments using  $^{15}\text{N}$  and  $^{13}\text{C}$  are usually performed. In heteronuclear experiment the chemical shifts of  $^{15}\text{N}$  and  $^{13}\text{C}$  isotopes are used to spread the resonances. Besides being NMR active nuclei,  $^{15}\text{N}$  and  $^{13}\text{C}$  isotopes possess low natural abundance and a gyromagnetic ratio which is considerably lower than that of protons. As a direct consequence,  $^{15}\text{N}$  and  $^{13}\text{C}$  display a remarkably lower sensibility than that of hydrogen. Two strategies are used to increase the sensitivity of these nuclei: (1) sample isotopic enrichment and (2) inverse NMR experiments in which the magnetization is transferred from proton to the heteronucleus, which results in the enhancement of the signal to noise ratio. A typical n-dimensional (nD) NMR experiment consists in the combination of (n-1) two-dimensional experiments which contains only one excitation and one detection period but repeats the evolution and mixing times (n-1) times. A typical nD NMR experiment thus follows the scheme:

excitation - (evolution - mixing)<sub>n-1</sub> - detection

During the excitation time the spins are perturbed and then let to precess freely during the evolution period ( $t_1$ ). In the mixing period ( $t_m$ ) the information on the chemical shift is transferred between correlated spins and only during the detection time the signal is physically recovered. Detection time ( $t_2$ ) is usually referred to as the “direct dimension” while the evolution times are often referred to as “indirect dimensions”.

The proton offers the highest sensitivity to NMR to respect to the other active nuclei such as  $^{15}\text{N}$  and  $^{13}\text{C}$ . For this reason  $^1\text{H}$  constitutes the preferred nucleus for the detection of the NMR signal while the other nuclei are usually stimulated during evolution periods of multidimensional NMR experiments and their information is transferred to protons for detection.

The most important inverse NMR experiments is the HSQC (heteronuclear single quantum correlation). Each signal in a HSQC spectrum represents a  $\text{H}^{\text{N}}$  proton in the protein backbone bound to a nitrogen atom and therefore represents one single amino acid. A HSQC has no diagonal like the homonuclear spectrum because different nuclei are observed in the  $t_1$  and  $t_2$ . An analogous experiment ( $^{13}\text{C}$  HSQC) can be performed for  $^{13}\text{C}$  and  $^1\text{H}$ .

Two-dimension spectra (2D) of proteins with a molecular weight higher than 30 kDa are often crowded with signals. In a 3D spectra the signals are spread out in a third dimension (usually  $^{13}\text{C}$  or  $^{15}\text{N}$ ) by the combination of a TOCSY / NOESY with an HSQC experiment.

The 3D triple resonance experiments exclusively correlate the resonances of the peptide backbone ( $\text{HN}(i)$ ,  $\text{N}(i)$ ,  $\text{C}\alpha(i)$ ,  $\text{H}\alpha(i)$ ,  $\text{C}\alpha(i-1)$ ,  $\text{CO}(i)$  and  $\text{CO}(i-1)$ ). The 3D experiments used to identify the backbone resonances are, usually, HNCA or HNCACB, HN(CO)CA or HN(CO)CACB, HNCOC, HN(CA)CO and HNHA. The HNCACB for example, correlates each  $\text{H}-^{15}\text{N}$  group with both the intra- and the neighbouring inter-residue  $\text{C}\alpha$  and  $\text{C}\beta$ .

As previously discussed, large molecules relax fast. This causes severe line broadening, poor spectral resolution and low signal to noise ratio.

Great improvement of NMR spectra of biological macromolecules with molecular weights above 30 KDa has been achieved with deuterium labelling. Further improvement in the spectra resolution of such macromolecules is represented by the introduction of transverse relaxation-optimized spectroscopy (TROSY)<sup>133</sup>. TROSY uses constructive interference between different relaxation mechanisms and allows spectra acquisition for macromolecules with molecular weights above 100 KDa.

# **Public Communications**

## Poster

- X-ray Crystal Structure Determination of Macromolecules, EMBO COURSE; Synchrotron Soleil, Saint Aubin (France), September 2008

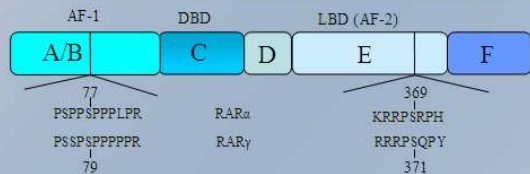
“Structural study of nuclear transcription regulation through phosphorylation”

Authors: Serena Sirigu, Sato Yoshiteru, Arnaud Poterszman, Natacha Rochel, Dino Moras

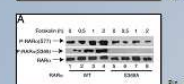
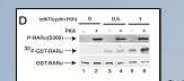
## Publications

- Editing

- Nuclear retinoic acid receptors (RARs) work as ligand dependent heterodimeric RAR/retinoid X receptor transcription activators
- Transcriptional activity of nuclear receptor is regulated by ligand binding and post-translational modifications such as phosphorylation
- Phosphorylation by PKA propagates cyclic AMP cAMP signalling from the AF-2 domain to the AF-1 domain. The first step involves the phosphorylation of the residue S369 (S371) of RAR $\alpha$  ( $\gamma$ ) located in the AF-2 domain. This signal is transferred to the CyclinH (CycH) binding domain located in the AF-1 domain resulting in enhanced CycH interaction and higher amounts of RAR  $\alpha$  ( $\gamma$ ) phosphorylated in the AF-1 domain at the residue S77 (S79) by the cdk7/TFIIH complex



- S369 (S371) RAR  $\alpha$  ( $\gamma$ ) phosphorylation increases RAR interaction with cyclin H
- S369 (S371) RAR  $\alpha$  ( $\gamma$ ) phosphorylation increases phosphorylation at S77 through cyclin H binding

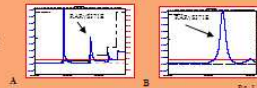
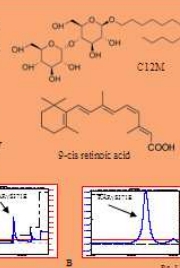


Gaillard et al. FASEB, 2005

**Aim** Provide structural evidences for an allosteric control of the molecular communication pathway between a general transcription factor and components of the activated transcription machinery

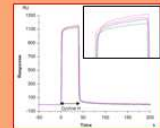
## Cloning and Purification

- Cloning of the mutant His-hRAR $\gamma$ S371E in which the mutation is well known to mimic a phosphorylation event
- The purification of His-hRAR $\gamma$ S371E consists in a metal affinity chromatography followed by a size exclusion chromatography (Fig 3 a, b) and is performed in presence of detergents (C12M)
- The purified protein is incubated with 9-cis retinoic acid



## Surface Plasmonic Resonance

- GST-hRXR $\alpha$ , hRAR $\alpha$ WT, hRAR $\alpha$ S369E were immobilized on the Biacore chip to test their interaction with hCycH
- The SPR signal increases as the concentration of hCycH is increased (Fig 4)
- The affinity constant is in the  $\mu$ M range
- No drastic difference in the affinity between hCycH and hRAR $\alpha$ WT or hRAR $\alpha$  S369E is observed.



Binding profiles of the interaction between the immobilized hRAR $\alpha$ S369E (purple), hRAR $\alpha$ WT (blue), hRXR $\alpha$  (green), LBDs or GST (red) and hCycH (115 nM)

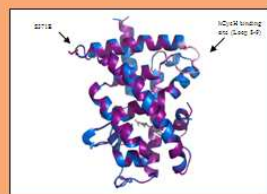
## Crystallization

- We solved the crystal structure of HisRAR $\gamma$ S371E in complex with the ligand 9-cis retinoic acid

STRUCTURE PARAMETERS	VALUES
Resolution (Å)	1.7
Rfactor (%)	19.4%
Rfree (%)	23.4%
rms bonds	<0.01
rms angles	1.12
rms chiral	0.07



Crystal space group P4<sub>1</sub>2<sub>1</sub>2



Crystal structures of HisRAR $\gamma$ S371E acid (blue) and hRAR $\alpha$ WT (purple) both bound to 9-cis retinoic acid (red). The mutation S371E is in green and the hCycH binding site is in rose

## Results and perspectives

- No drastic structural conformational changes between hRAR $\gamma$ WT or hRAR $\gamma$ S371E in their crystalline state
- Solution studies by NMR and Fluorescence Anisotropy will be performed to investigate the differences in the dynamics of hRAR $\gamma$ WT and hRAR $\gamma$ S371E



# Bibliography

## Reference List

1. Allenby,G. *et al.* Retinoic acid receptors and retinoid X receptors: interactions with endogenous retinoic acids. *Proc. Natl. Acad. Sci. U. S. A* **90**, 30-34 (1993).
2. Almlof,T., Wallberg,A.E., Gustafsson,J.A., & Wright,A.P. Role of important hydrophobic amino acids in the interaction between the glucocorticoid receptor tau 1-core activation domain and target factors. *Biochemistry* **37**, 9586-9594 (1998).
3. Altucci,L. & Gronemeyer,H. The promise of retinoids to fight against cancer. *Nat. Rev. Cancer* **1**, 181-193 (2001).
4. Anzano,M.A. *et al.* Prevention of breast cancer in the rat with 9-cis-retinoic acid as a single agent and in combination with tamoxifen. *Cancer Res.* **54**, 4614-4617 (1994).
5. Aranda,A. & Pascual,A. Nuclear hormone receptors and gene expression. *Physiol Rev.* **81**, 1269-1304 (2001).
6. Asherie,N. Protein crystallization and phase diagrams.(2004).
7. Aue,W.P. Two-dimensional spectroscopy. Application to nuclear magnetic resonance.(1976).
8. Bain,D.L., Heneghan,A.F., Connaghan-Jones,K.D., & Miura,M.T. Nuclear receptor structure: implications for function. *Annu. Rev. Physiol* **69**, 201-220 (2007).
9. Baniahmad,A. The tau 4 activation domain of the thyroid hormone receptor is required for release of a putative corepressor(s) necessary for transcriptional silencing.(1995).
10. Baretino,D., Vivanco Ruiz,M.M., & Stunnenberg,H.G. Characterization of the ligand-dependent transactivation domain of thyroid hormone receptor. *EMBO J.* **13**, 3039-3049 (1994).
11. Barrero,M.J. & Malik,S. Two functional modes of a nuclear receptor-recruited arginine methyltransferase in transcriptional activation. *Mol. Cell* **24**, 233-243 (2006).
12. Bas,D.C., Rogers,D.M., & Jensen,J.H. Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins* **73**, 765-783 (2008).
13. Bastien,J. *et al.* TFIIF interacts with the retinoic acid receptor gamma and phosphorylates its AF-1-activating domain through cdk7. *J. Biol. Chem.* **275**, 21896-21904 (2000).

14. Benecke,A., Chambon,P., & Gronemeyer,H. Synergy between estrogen receptor alpha activation functions AF1 and AF2 mediated by transcription intermediary factor TIF2. *EMBO Rep.* **1**, 151-157 (2000).
15. Berry,M., Metzger,D., & Chambon,P. Role of the two activating domains of the oestrogen receptor in the cell-type and promoter-context dependent agonistic activity of the anti-oestrogen 4-hydroxytamoxifen. *EMBO J.* **9**, 2811-2818 (1990).
16. Bevan,C.L., Hoare,S., Claessens,F., Heery,D.M., & Parker,M.G. The AF1 and AF2 domains of the androgen receptor interact with distinct regions of SRC1. *Mol. Cell Biol.* **19**, 8383-8392 (1999).
17. Bivona,T.G. *et al.* PKC regulates a farnesyl-electrostatic switch on K-Ras that promotes its association with Bcl-XL on mitochondria and induces apoptosis. *Mol. Cell* **21**, 481-493 (2006).
18. Blanquart,C., Barbier,O., Fruchart,J.C., Staels,B., & Glineur,C. Peroxisome proliferator-activated receptor alpha (PPARalpha ) turnover by the ubiquitin-proteasome system controls the ligand-induced expression level of its target genes. *J. Biol. Chem.* **277**, 37254-37259 (2002).
19. Blomhoff,R. & Blomhoff,H.K. Overview of retinoid metabolism and function. *J. Neurobiol.* **66**, 606-630 (2006).
20. Blomhoff,R. & Blomhoff,H.K. Overview of retinoid metabolism and function. *J. Neurobiol.* **66**, 606-630 (2006).
21. Blomhoff,R., Green,M.H., Berg,T., & Norum,K.R. Transport and storage of vitamin A. *Science* **250**, 399-404 (1990).
22. Boudjelal,M., Wang,Z., Voorhees,J.J., & Fisher,G.J. Ubiquitin/proteasome pathway regulates levels of retinoic acid receptor gamma and retinoid X receptor alpha in human keratinocytes. *Cancer Res.* **60**, 2247-2252 (2000).
23. Bour,G. Protein kinases and the proteasome join in the combinatorial control of transcription by nuclear retinoic acid receptors.(2007).
24. Bour,G. *et al.* Cyclin H binding to the RARalpha activation function (AF)-2 domain directs phosphorylation of the AF-1 domain by cyclin-dependent kinase 7. *Proc. Natl. Acad. Sci. U. S. A* **102**, 16608-16613 (2005).
25. Bourguet,W. *et al.* Purification, functional characterization, and crystallization of the ligand binding domain of the retinoid X receptor. *Protein Expr. Purif.* **6**, 604-608 (1995).
26. Bourguet,W. *et al.* Crystal structure of a heterodimeric complex of RAR and RXR ligand-binding domains. *Mol. Cell* **5**, 289-298 (2000).
27. Brazil,D.P. & Hemmings,B.A. Ten years of protein kinase B signalling: a hard Akt to follow. *Trends Biochem. Sci.* **26**, 657-664 (2001).

28. Brelivet,Y., Kammerer,S., Rochel,N., Poch,O., & Moras,D. Signature of the oligomeric behaviour of nuclear receptors at the sequence and structural level. *EMBO Rep.* **5**, 423-429 (2004).
29. Brooks,B.R. *et al.* CHARMM : A program for macromolecular energy, minimzation, and dynamics calculations. *J Comput Chem*187-217 (1983).
30. Bruck,N. *et al.* A coordinated phosphorylation cascade initiated by p38MAPK/MSK1 directs RARalpha to target promoters. *EMBO J.* **28**, 34-47 (2009).
31. Brzozowski,A.M. *et al.* Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **389**, 753-758 (1997).
32. Bunone,G., Briand,P.A., Miksicek,R.J., & Picard,D. Activation of the unliganded estrogen receptor by EGF involves the MAP kinase pathway and direct phosphorylation. *EMBO J.* **15**, 2174-2183 (1996).
33. Burgdorf,S., Leister,P., & Scheidtmann,K.H. TSG101 interacts with apoptosis-antagonizing transcription factor and enhances androgen receptor-mediated transcription by promoting its monoubiquitination. *J. Biol. Chem.* **279**, 17524-17534 (2004).
34. Casanova,J. Functional evidence for ligand-dependent dissociation of thyroid hormone and retinoic acid receptors from an inhibitory cellular factor.(1994).
35. Chayen,N.E. Comparative studies of protein crystallization by vapour-diffusion and microbatch techniques.(1998).
36. Chayen,N.E. Methods for separating nucleation and growth in protein crystallisation.(2005).
37. Chen,D. *et al.* Activation of estrogen receptor alpha by S118 phosphorylation involves a ligand-dependent interaction with TFIID and participation of CDK7. *Mol. Cell* **6**, 127-137 (2000).
38. Chen,H., Lin,R.J., Xie,W., Wilpitz,D., & Evans,R.M. Regulation of hormone-induced histone hyperacetylation and gene activation via acetylation of an acetylase. *Cell* **98**, 675-686 (1999).
39. Chen,J.D. & Evans,R.M. A transcriptional co-repressor that interacts with nuclear hormone receptors. *Nature* **377**, 454-457 (1995).
40. Clarke,N., Germain,P., Altucci,L., & Gronemeyer,H. Retinoids: potential in cancer prevention and therapy. *Expert. Rev. Mol. Med.* **6**, 1-23 (2004).
41. Collins,S.J. The role of retinoids and retinoic acid receptors in normal hematopoiesis. *Leukemia* **16**, 1896-1905 (2002).

42. Couture,J.F., Collazo,E., Hauk,G., & Trievel,R.C. Structural basis for the methylation site specificity of SET7/9. *Nat. Struct. Mol. Biol.* **13**, 140-146 (2006).
43. Dace,A. *et al.* Hormone binding induces rapid proteasome-mediated degradation of thyroid hormone receptors. *Proc. Natl. Acad. Sci. U. S. A* **97**, 8985-8990 (2000).
44. Danielian,P.S., White,R., Lees,J.A., & Parker,M.G. Identification of a conserved region required for hormone dependent transcriptional activation by steroid hormone receptors. *EMBO J.* **11**, 1025-1033 (1992).
45. Darimont,B.D. *et al.* Structure and specificity of nuclear receptor-coactivator interactions. *Genes Dev.* **12**, 3343-3356 (1998).
46. Davies,P. & Lippman,S. Biologic basis of retinoid pharmacology:implications for cancer prevention and therapy. *Adv. Oncol.* **2-10** (1996).
47. Drenth,J. Principles of protein X-ray crystallography. 1994.  
Ref Type: Serial (Book,Monograph)
48. Ducruix,A. *Crystallization of Nucleic Acid and Protein: A Practical Approach*(1999).
49. Durand,B. *et al.* Activation function 2 (AF-2) of retinoic acid receptor and 9-cis retinoic acid receptor: presence of a conserved autonomous constitutive activating domain and influence of the nature of the response element on AF-2 activity. *EMBO J.* **13**, 5370-5382 (1994).
50. Durand,B., Saunders,M., Leroy,P., Leid,M., & Chambon,P. All-trans and 9-cis retinoic acid induction of CRABP II transcription is mediated by RAR-RXR heterodimers bound to DR1 and DR2 repeated motifs. *Cell* **71**, 73-85 (1992).
51. Egea,P.F., Mitschler,A., & Moras,D. Molecular recognition of agonist ligands by RXRs. *Mol. Endocrinol.* **16**, 987-997 (2002).
52. Emsley,P. & Cowtan,K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126-2132 (2004).
53. Evans,T.R. & Kaye,S.B. Retinoids: present role and future potential. *Br. J. Cancer* **80**, 1-8 (1999).
54. Farrow,N.A. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by <sup>15</sup>N NMR relaxation.(1994).
55. Faus,H. & Haendler,B. Post-translational modifications of steroid receptors. *Biomed. Pharmacother.* **60**, 520-528 (2006).

56. Finzi,E., Blake,M.J., Celano,P., Skouge,J., & Diwan,R. Cellular localization of retinoic acid receptor-gamma expression in normal and neoplastic skin. *Am. J. Pathol.* **140**, 1463-1471 (1992).
57. Forman,B.M. & Evans,R.M. Nuclear hormone receptors activate direct, inverted, and everted repeats. *Ann. N. Y. Acad. Sci.* **761**, 29-37 (1995).
58. Freedman,L.P. Transcriptional targets of the vitamin D3 receptor-mediating cell cycle arrest and differentiation. *J. Nutr.* **129**, 581S-586S (1999).
59. Freeman,B.C. & Yamamoto,K.R. Disassembly of transcriptional regulatory complexes by molecular chaperones. *Science* **296**, 2232-2235 (2002).
60. Fu,M. *et al.* Acetylation of androgen receptor enhances coactivator binding and promotes prostate cancer cell growth. *Mol. Cell Biol.* **23**, 8563-8575 (2003).
61. Fu,M. *et al.* Androgen receptor acetylation governs trans activation and MEKK1-induced apoptosis without affecting in vitro sumoylation and trans-repression function. *Mol. Cell Biol.* **22**, 3373-3388 (2002).
62. Fu,M., Wang,C., Zhang,X., & Pestell,R.G. Acetylation of nuclear receptors in cellular growth and apoptosis. *Biochem. Pharmacol.* **68**, 1199-1208 (2004).
63. Gaillard,E. *et al.* Phosphorylation by PKA potentiates retinoic acid receptor alpha activity by means of increasing interaction with and phosphorylation by cyclin H/cdk7. *Proc. Natl. Acad. Sci. U. S. A* **103**, 9548-9553 (2006).
64. Gampe,R. *et al.* Asymmetry in the PPARg/RXRa Crystal Structure Reveals the Molecular Basis of Heterodimerization among Nuclear Receptors. *Mol. Cell*(2000).
65. Garman,E. Macromolecular Cryocrystallography.(1997).
66. Garman,E. Cool data: quantity AND quality.(1999).
67. George,A. Predicting protein crystallization from a dilute solution property.(1994).
68. Gianni,M., Bauer,A., Garattini,E., Chambon,P., & Rochette-Egly,C. Phosphorylation by p38MAPK and recruitment of SUG-1 are required for RA-induced RAR gamma degradation and transactivation. *EMBO J.* **21**, 3760-3769 (2002).
69. Gianni,M. *et al.* Down-regulation of the phosphatidylinositol 3-kinase/Akt pathway is involved in retinoic acid-induced phosphorylation, degradation, and transcriptional activity of retinoic acid receptor gamma 2. *J. Biol. Chem.* **277**, 24859-24862 (2002).

70. Giguere,V. Orphan nuclear receptors: from gene to function. *Endocr. Rev.* **20**, 689-725 (1999).
71. Glass,C.K., Rose,D.W., & Rosenfeld,M.G. Nuclear receptor coactivators. *Curr. Opin. Cell Biol.* **9**, 222-232 (1997).
72. Glass,C.K. & Rosenfeld,M.G. The coregulator exchange in transcriptional functions of nuclear receptors. *Genes Dev.* **14**, 121-141 (2000).
73. Gottardis,M.M., Lamph,W.W., Shalinsky,D.R., Wellstein,A., & Heyman,R.A. The efficacy of 9-cis retinoic acid in experimental models of cancer. *Breast Cancer Res. Treat.* **38**, 85-96 (1996).
74. Gronemeyer,H. & Moras,D. Nuclear receptors. How to finger DNA. *Nature* **375**, 190-191 (1995).
75. Gundersen,T.E. & Blomhoff,R. Qualitative and quantitative liquid chromatographic determination of natural retinoids in biological samples. *J. Chromatogr. A* **935**, 13-43 (2001).
76. Haelens,A., Tanner,T., Denayer,S., Callewaert,L., & Claessens,F. The hinge region regulates DNA binding, nuclear translocation, and transactivation of the androgen receptor. *Cancer Res.* **67**, 4514-4523 (2007).
77. Hager,G., Lim,C., Elbi,C., & Baumann,C. Steroid receptor interactions with heat shock protein and immunophilin chaperones. *Journal of Steroid Biochemistry & Molecular Biology*(2000).
78. Henriksson,A. *et al.* Role of the Ada adaptor complex in gene activation by the glucocorticoid receptor. *Mol. Cell Biol.* **17**, 3065-3073 (1997).
79. Henttu,P.M., Kalkhoven,E., & Parker,M.G. AF-2 activity and recruitment of steroid receptor coactivator 1 to the estrogen receptor depend on a lysine residue conserved in nuclear receptors. *Mol. Cell Biol.* **17**, 1832-1839 (1997).
80. Heyman,R.A. *et al.* 9-cis retinoic acid is a high affinity ligand for the retinoid X receptor. *Cell* **68**, 397-406 (1992).
81. Hittelman,A.B., Burakov,D., Iniguez-Lluhi,J.A., Freedman,L.P., & Garabedian,M.J. Differential regulation of glucocorticoid receptor transcriptional activation via AF-1-associated proteins. *EMBO J.* **18**, 5380-5388 (1999).
82. Hong,S.H. & Privalsky,M.L. The SMRT corepressor is regulated by a MEK-1 kinase pathway: inhibition of corepressor function is associated with SMRT phosphorylation and nuclear export. *Mol. Cell Biol.* **20**, 6612-6625 (2000).

83. Horlein,A.J. *et al.* Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. *Nature* **377**, 397-404 (1995).
84. Huq,M.D., Ha,S.G., & Wei,L.N. Modulation of retinoic acid receptor alpha activity by lysine methylation in the DNA binding domain. *J. Proteome. Res.* **7**, 4538-4545 (2008).
85. Huq,M.D., Tsai,N.P., Khan,S.A., & Wei,L.N. Lysine trimethylation of retinoic acid receptor-alpha: a novel means to regulate receptor function. *Mol. Cell Proteomics.* **6**, 677-688 (2007).
86. Ito,S. *et al.* In vivo potentiation of human oestrogen receptor alpha by Cdk7-mediated phosphorylation. *Genes Cells* **9**, 983-992 (2004).
87. Kashchiev,D. Applications of the nucleation theorem 2000).
88. Katz,R.W. & Koenig,R.J. Specificity and mechanism of thyroid hormone induction from an octamer response element. *J. Biol. Chem.* **269**, 18915-18920 (1994).
89. Kaushik,R. & Ratner,L. Role of human immunodeficiency virus type 1 matrix phosphorylation in an early postentry step of virus replication. *J. Virol.* **78**, 2319-2326 (2004).
90. Kim,J.Y., Son,Y.L., & Lee,Y.C. Involvement of SMRT corepressor in transcriptional repression by the vitamin D receptor. *Mol. Endocrinol.* **23**, 251-264 (2009).
91. Klaholz,B.P. Structural role of a detergent molecule in retinoic acid nuclear receptor crystals.(2000).
92. Klaholz,B.P. *et al.* Conformational adaptation of agonists to the human nuclear receptor RAR gamma. *Nat. Struct. Biol.* **5**, 199-202 (1998).
93. Kong,G. *et al.* The retinoid X receptor-selective retinoid, LGD1069, down-regulates cyclooxygenase-2 expression in human breast cells through transcription factor crosstalk: implications for molecular-based chemoprevention. *Cancer Res.* **65**, 3462-3469 (2005).
94. Kopf,E. *et al.* Dimerization with retinoid X receptors and phosphorylation modulate the retinoic acid-induced degradation of retinoic acid receptors alpha and gamma through the ubiquitin-proteasome pathway. *J. Biol. Chem.* **275**, 33280-33288 (2000).
95. Kraemer,K.H., DiGiovanna,J.J., Moshell,A.N., Tarone,R.E., & Peck,G.L. Prevention of skin cancer in xeroderma pigmentosum with the use of oral isotretinoin. *N. Engl. J. Med.* **318**, 1633-1637 (1988).
96. Kraemer,K.H., DiGiovanna,J.J., & Peck,G.L. Chemoprevention of skin cancer in xeroderma pigmentosum. *J. Dermatol.* **19**, 715-718 (1992).



97. Kurokawa,R. *et al.* Differential orientations of the DNA-binding domain and carboxy-terminal dimerization interface regulate binding site selection by nuclear receptor heterodimers. *Genes Dev.* **7**, 1423-1435 (1993).
98. Lala,D. *et al.* Activation of the orphan nuclear receptor steroidogenic factor 1 by oxysterols. *PNAS*(1997).
99. Lamura,A. Solvent content of protein crystals from diffraction intensities by Independent Component Analysis.(2003).
100. Laudet,V. & Gronemeyer,H. *The Nuclear Receptor FactsBook*2002).
101. Laudet,V. Evolution of the nuclear receptor superfamily: early diversification from an ancestral orphan receptor.(1997).
102. Leder,A., Kuo,A., Cardiff,R.D., Sinn,E., & Leder,P. v-Ha-ras transgene abrogates the initiation step in mouse skin tumorigenesis: effects of phorbol esters and retinoic acid. *Proc. Natl. Acad. Sci. U. S. A* **87**, 9178-9182 (1990).
103. Lee,M.S., Kliewer,S.A., Provencal,J., Wright,P.E., & Evans,R.M. Structure of the retinoid X receptor alpha DNA binding domain: a helix required for homodimeric DNA binding. *Science* **260**, 1117-1121 (1993).
104. Lefstin,J.A. & Yamamoto,K.R. Allosteric effects of DNA on transcriptional regulators. *Nature* **392**, 885-888 (1998).
105. Leid,M., Kastner,P., & Chambon,P. Multiplicity generates diversity in the retinoic acid signalling pathways. *Trends Biochem. Sci.* **17**, 427-433 (1992).
106. Levin,A.A. *et al.* 9-cis retinoic acid stereoisomer binds and activates the nuclear receptor RXR alpha. *Nature* **355**, 359-361 (1992).
107. Li,E. & Tso,P. Vitamin A uptake from foods. *Curr. Opin. Lipidol.* **14**, 241-247 (2003).
108. Li,H., Robertson,A.D., & Jensen,J.H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins* **61**, 704-721 (2005).
109. Lin,H.K. *et al.* Proteasome activity is required for androgen receptor transcriptional activity via regulation of androgen receptor nuclear translocation and interaction with coregulators in prostate cancer cells. *J. Biol. Chem.* **277**, 36570-36576 (2002).
110. Lu,W. & Ou,J.H. Phosphorylation of hepatitis C virus core protein by protein kinase A and protein kinase C. *Virology* **300**, 20-30 (2002).
111. MacKerell,A.D. *et al.* All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J Phys Chem B*3586-3616 (1998).

112. MacKerell,A.D., Jr., Feig,M., & Brooks,C.L., III Improved treatment of the protein backbone in empirical force fields. *J Am. Chem Soc.* **126**, 698-699 (2004).
113. Maden,M. Retinoid signalling in the development of the central nervous system. *Nat. Rev. Neurosci.* **3**, 843-853 (2002).
114. Mader,S. *et al.* The patterns of binding of RAR, RXR and TR homo- and heterodimers to direct repeats are dictated by the binding specificities of the DNA binding domains. *EMBO J.* **12**, 5029-5041 (1993).
115. Mangelsdorf,D.J. The nuclear receptor superfamily: the second decade.(1995).
116. Mantel,A.G., Delabre,K., Lescop,P., & Milgrom,E. Nuclear localization signals also mediate the outward movement of proteins from the nucleus. *PNAS*(1994).
117. Mendonca,R.L. *et al.* Nuclear Hormone Receptors and Evolution. *Amer. Zool.*(1999).
118. Metivier,R., Penot,G., Flouriot,G., & Pakdel,F. Synergism between ERalpha transactivation function 1 (AF-1) and AF-2 mediated by steroid receptor coactivator protein-1: requirement for the AF-1 alpha-helical core and for a direct interaction between the N- and C-terminal domains. *Mol. Endocrinol.* **15**, 1953-1970 (2001).
119. Miller,V.A. *et al.* Initial clinical trial of the retinoid receptor pan agonist 9-cis retinoic acid. *Clin. Cancer Res.* **2**, 471-475 (1996).
120. Montano,M.M., Muller,V., Trobaugh,A., & Katzenellenbogen,B.S. The carboxy-terminal F domain of the human estrogen receptor: role in the transcriptional activity of the receptor and the effectiveness of antiestrogens as estrogen antagonists. *Mol. Endocrinol.* **9**, 814-825 (1995).
121. Moon,T.E., Levine,N., Cartmel,B., & Bangert,J.L. Retinoids in prevention of skin cancer. *Cancer Lett.* **114**, 203-205 (1997).
122. Moras,D. & Gronemeyer,H. The nuclear receptor ligand-binding domain: structure and function. *Curr. Opin. Cell Biol.* **10**, 384-391 (1998).
123. Murshudov,G.N., Vagin,A.A., & Dodson,E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D. Biol. Crystallogr.* **53**, 240-255 (1997).
124. Nagy,L. *et al.* Mechanism of corepressor binding and release from nuclear hormone receptors. *Genes Dev.* **13**, 3209-3216 (1999).
125. Navaza,J. *Acta Cryst.*(1994).

126. Niederreither,K., Subbarayan,V., Dolle,P., & Chambon,P. Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. *Nat. Genet.* **21**, 444-448 (1999).
127. Niles,R.M. Signaling pathways in retinoid chemoprevention and treatment of cancer. *Mutat. Res.* **555**, 81-96 (2004).
128. Nolte,R.T. *et al.* Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma. *Nature* **395**, 137-143 (1998).
129. Oppenheimer,J.H., Schwartz,H.L., & Surks,M.I. Determination of common parameters fo iodothyronine metabolism and distribution in man by noncompartmental analysis. *J. Clin. Endocrinol. Metab* **41**, 319-324 (1975).
130. Otwinowski,Z. & Minor,W. *Methods Enzymol.*(1997).
131. Pakdel,F., Reese,J.C., & Katzenellenbogen,B.S. Identification of charged residues in an N-terminal portion of the hormone-binding domain of the human estrogen receptor important in transcriptional activity of the receptor. *Mol. Endocrinol.* **7**, 1408-1417 (1993).
132. Perissi,V. *et al.* Molecular determinants of nuclear receptor-corepressor interaction. *Genes Dev.* **13**, 3198-3208 (1999).
133. Pervushin,K. Impact of tranverse relaxation optimized spectroscopy (TROSY) on NMR as a technique in structural biology.(2000).
134. Peters,G.A. & Khan,S.A. Estrogen receptor domains E and F: role in dimerization and interaction with coactivator RIP-140. *Mol. Endocrinol.* **13**, 286-296 (1999).
135. Phillips,J.C. *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802 (2005).
136. Picard,D. Chaperoning steroid hormone action. *Trends Endocrinol. Metab* **17**, 229-235 (2006).
137. Pogenberg,V. Purification and crystallization of the heterodimeric complex of RARbeta and RXRalpha ligand-binding domains in the active conformation.(2004).
138. Polly,P. *et al.* VDR-Alien: a novel, DNA-selective vitamin D(3) receptor-corepressor partnership. *FASEB J.* **14**, 1455-1463 (2000).
139. Potter,G.B. *et al.* The hairless gene mutated in congenital hair loss disorders encodes a novel nuclear receptor corepressor. *Genes Dev.* **15**, 2687-2701 (2001).
140. Qiu,M. & Lange,C.A. MAP kinases couple multiple functions of human progesterone receptors: degradation, transcriptional synergy, and nuclear association. *J. Steroid Biochem. Mol. Biol.* **85**, 147-157 (2003).

141. Rachez,C. *et al.* A novel protein complex that interacts with the vitamin D3 receptor in a ligand-dependent manner and enhances VDR transactivation in a cell-free system. *Genes Dev.* **12**, 1787-1800 (1998).
142. Radoja,N., Komine,M., Jho,S.H., Blumenberg,M., & Tomic-Canic,M. Novel mechanism of steroid action in skin through glucocorticoid receptor monomers. *Mol. Cell Biol.* **20**, 4328-4339 (2000).
143. Rastinejad,F., Perlmann,T., Evans,R.M., & Sigler,P.B. Structural determinants of nuclear receptor assembly on DNA direct repeats. *Nature* **375**, 203-211 (1995).
144. Reid,G. *et al.* Cyclic, proteasome-mediated turnover of unliganded and liganded ERalpha on responsive promoters is an integral feature of estrogen signaling. *Mol. Cell* **11**, 695-707 (2003).
145. Reijntjes,S., Blentic,A., Gale,E., & Maden,M. The control of morphogen signalling: regulation of the synthesis and catabolism of retinoic acid in the developing embryo. *Dev. Biol.* **285**, 224-237 (2005).
146. Renaud,J.P. *et al.* Crystal structure of the RAR-gamma ligand-binding domain bound to all-trans retinoic acid. *Nature* **378**, 681-689 (1995).
147. Rhodes,G. Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models.(1993).
148. Rochette-Egly,C. Phosphorylation of the retinoic acid receptor-alpha by protein kinase A.(1995).
149. Rochette-Egly,C. Nuclear receptors: integration of multiple signalling pathways through phosphorylation.(2003).
150. Rochette-Egly,C., Adam,S., Rossignol,M., Egly,J.M., & Chambon,P. Stimulation of RAR alpha activation function AF-1 through binding to the general transcription factor TFIID and phosphorylation by CDK7. *Cell* **90**, 97-107 (1997).
151. Rowan,B.G., Garrison,N., Weigel,N.L., & O'Malley,B.W. 8-Bromo-cyclic AMP induces phosphorylation of two sites in SRC-1 that facilitate ligand-independent activation of the chicken progesterone receptor and are critical for functional cooperation between SRC-1 and CREB binding protein. *Mol. Cell Biol.* **20**, 8720-8730 (2000).
152. Russell,R.M. The vitamin A spectrum: from deficiency to toxicity. *Am. J. Clin. Nutr.* **71**, 878-884 (2000).
153. Ryckaert,J., Cicotti,G., & Berendsen,H.J.C. Numerical Integration of the cartesian equation of motion of a system with constraints: molecular dynamics of N-alkanes. *J Comput Phys*327-341 (1977).

154. Schaefer,M., van Vlijmen,H.W., & Karplus,M. Electrostatic contributions to molecular free energies in solution. *Adv. Protein Chem* **51**, 1-57 (1998).
155. Schiff,R. & Osborne,C.K. Endocrinology and hormone therapy in breast cancer: new insight into estrogen receptor-alpha function and its implication for endocrine therapy resistance in breast cancer. *Breast Cancer Res.* **7**, 205-211 (2005).
156. Schuttelkopf,A.W. & van Aalten,D.M. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 1355-1363 (2004).
157. Schwabe,J.W., Chapman,L., Finch,J.T., Rhodes,D., & Neuhaus,D. DNA recognition by the oestrogen receptor: from solution to the crystal. *Structure.* **1**, 187-204 (1993).
158. Schwabe,J.W., Neuhaus,D., & Rhodes,D. Solution structure of the DNA-binding domain of the oestrogen receptor. *Nature* **348**, 458-461 (1990).
159. Shen,T., Horwitz,K.B., & Lange,C.A. Transcriptional hyperactivity of human progesterone receptors is coupled to their ligand-dependent down-regulation by mitogen-activated protein kinase-dependent phosphorylation of serine 294. *Mol. Cell Biol.* **21**, 6122-6131 (2001).
160. Shiau,A.K. *et al.* The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **95**, 927-937 (1998).
161. Sladek,F.M., Ruse,M.D., Jr., Nepomuceno,L., Huang,S.M., & Stallcup,M.R. Modulation of transcriptional activation and coactivator interaction by a splicing variation in the F domain of nuclear receptor hepatocyte nuclear factor 4alpha1. *Mol. Cell Biol.* **19**, 6509-6522 (1999).
162. Smith,M.A. *et al.* Phase I and pharmacokinetic evaluation of all-trans-retinoic acid in pediatric patients with cancer. *J. Clin. Oncol.* **10**, 1666-1673 (1992).
163. Soprano,K.J. & Soprano,D.R. Retinoic acid receptors and cancer. *J. Nutr.* **132**, 3809S-3813S (2002).
164. Spencer,T.E. *et al.* Steroid receptor coactivator-1 is a histone acetyltransferase. *Nature* **389**, 194-198 (1997).
165. Sporn,M., Roberts,A., & Goodman,D. *The Retinoids: Biology, Chemistry, and Medicine*, 2nd edition. ed. Raven Press, L. N. Y. 631-658. 1995.

Ref Type: Serial (Book,Monograph)

166. Subramanian,K. *et al.* Regulation of estrogen receptor alpha by the SET7 lysine methyltransferase. *Mol. Cell* **30**, 336-347 (2008).

167. Sun,P.D. Generating isomorphous heavy-atom derivatives by a quick-soak method. Part I: test cases.(2002).
168. Tanaka,T., Rodriguez de la Concepcion ML, & De Luca,L.M. Involvement of all-trans-retinoic acid in the breakdown of retinoic acid receptors alpha and gamma through proteasomes in MCF-7 human breast cancer cells. *Biochem. Pharmacol.* **61**, 1347-1355 (2001).
169. Taneja,R. *et al.* Phosphorylation of activation functions AF-1 and AF-2 of RAR alpha and RAR gamma is indispensable for differentiation of F9 cells upon retinoic acid and cAMP treatment. *EMBO J.* **16**, 6452-6465 (1997).
170. Thomas,D.H. A novel dialysis procedure for the crystallization of proteins.(1989).
171. Umesono,K. & Evans,R.M. Determinants of target gene specificity for steroid/thyroid hormone receptors. *Cell* **57**, 1139-1146 (1989).
172. van Holde,K.E. Principles of Physical Biochemistry.(1998).
173. van Vlijmen,H.W., Schaefer,M., & Karplus,M. Improving the accuracy of protein pKa calculations: conformational averaging versus the average structure. *Proteins* **33**, 145-158 (1998).
174. Vine,A.L., Leung,Y.M., & Bertram,J.S. Transcriptional regulation of connexin 43 expression by retinoids and carotenoids: similarities and differences. *Mol. Carcinog.* **43**, 75-85 (2005).
175. Vivat,V. *et al.* Sequences in the ligand-binding domains of the human androgen and progesterone receptors which determine their distinct ligand identities. *J. Mol. Endocrinol.* **18**, 147-160 (1997).
176. Voegel,J.J. *et al.* The coactivator TIF2 contains three nuclear receptor-binding motifs and mediates transactivation through CBP binding-dependent and -independent pathways. *EMBO J.* **17**, 507-519 (1998).
177. Wagner,R.L. *et al.* A structural role for hormone in the thyroid hormone receptor. *Nature* **378**, 690-697 (1995).
178. Wallace,A.D. & Cidlowski,J.A. Proteasome-mediated glucocorticoid receptor degradation restricts transcriptional signaling by glucocorticoids. *J. Biol. Chem.* **276**, 42714-42721 (2001).
179. Wang,C. *et al.* Direct acetylation of the estrogen receptor alpha hinge region by p300 regulates transactivation and hormone sensitivity. *J. Biol. Chem.* **276**, 18375-18383 (2001).
180. Wang,Y. *et al.* Prevention of lung cancer progression by bexarotene in mouse models. *Oncogene* **25**, 1320-1329 (2006).

181. Warnmark,A. Activation functions 1 and 2 of nuclear receptors: molecular strategies for transcriptional activation.(2003).
182. Wider,G. Homonuclear Two-Dimensional <sup>1</sup>H NMR of Proteins. Experimental Procedures.(1984).
183. Wider,G. Technical aspects of NMR spectroscopy with biological macromolecules and studies of hydration in solution.(1998).
184. Williams,S.P. Atomic structure of progesterone complexed with its receptor.(1998).
185. Wilson,T.E., Fahrner,T.J., & Milbrandt,J. The orphan receptors NGFI-B and steroidogenic factor 1 establish monomer binding as a third paradigm of nuclear receptor-DNA interaction. *Mol. Cell Biol.* **13**, 5794-5804 (1993).
186. Wu,S.C. & Zhang,Y. Minireview: role of protein methylation and demethylation in nuclear hormone signaling. *Mol. Endocrinol.* **23**, 1323-1334 (2009).
187. Wurtz,J.M. *et al.* A canonical structure for the ligand-binding domain of nuclear receptors. *Nat. Struct. Biol.* **3**, 206 (1996).
188. Xu,H.E. *et al.* Structural basis for antagonist-mediated recruitment of nuclear co-repressors by PPARalpha. *Nature* **415**, 813-817 (2002).
189. Yamauchi,P.S., Rizk,D., & Lowe,N.J. Retinoid therapy for psoriasis. *Dermatol. Clin.* **22**, 467-76, x (2004).
190. Yokota,K. *et al.* Proteasome-mediated mineralocorticoid receptor degradation attenuates transcriptional response to aldosterone. *Endocr. Res.* **30**, 611-616 (2004).
191. Zamir,I. *et al.* A nuclear hormone receptor corepressor mediates transcriptional silencing by receptors with distinct repression domains. *Mol. Cell Biol.* **16**, 5458-5465 (1996).
192. Zechel,C., Shen,X.Q., Chambon,P., & Gronemeyer,H. Dimerization interfaces formed between the DNA binding domains determine the cooperative binding of RXR/RAR and RXR/TR heterodimers to DR5 and DR4 elements. *EMBO J.* **13**, 1414-1424 (1994).
193. Zechel,C. *et al.* The dimerization interfaces formed between the DNA binding domains of RXR, RAR and TR determine the binding specificity and polarity of the full-length receptors to direct repeats. *EMBO J.* **13**, 1425-1433 (1994).
194. Zhao,Q. *et al.* Structural basis of RXR-DNA interactions. *J. Mol. Biol.* **296**, 509-520 (2000).

195. Zhu, J. *et al.* Retinoic acid induces proteasome-dependent degradation of retinoic acid receptor alpha (RARalpha) and oncogenic RARalpha fusion proteins. *Proc. Natl. Acad. Sci. U. S. A* **96**, 14807-14812 (1999).



## **Annex:**

### hRAR $\gamma$ (LBD) (178-423):

DSYELSPQLEELITKVSKAHQETFPSPCLQGLGKYTTNNSADHRVQLDLGLWDFSELATKCIKIVEFAKRLPGFTGLSIADQI  
TLLKAAACLDILMLRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFGPLTDLVFAFAGQLPLEMDDTETGLLSAICLICGDR  
MDLEEPEKVDKLEPLLEALRLYARRRRPSQPYMFPRMLMKITDLRGISTKGAERAITLKMEIPGPMPLIREMLENPEMFE

### hRAR $\gamma$ S371E (LBD) (178-423)

DSYELSPQLEELITKVSKAHQETFPSPCLQGLGKYTTNNSADHRVQLDLGLWDFSELATKCIKIVEFAKRLPGFTGLSIADQI  
TLLKAAACLDILMLRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFGPLTDLVFAFAGQLPLEMDDTETGLLSAICLICGDR  
MDLEEPEKVDKLEPLLEALRLYARRRRPEQPYMFPRMLMKITDLRGISTKGAERAITLKMEIPGPMPLIREMLENPEMFE

### hRAR $\alpha$ (LBD) (176-421)

ESYTLTPEVHELIEKVRKAHQETFPALCQLGKYTTNNSSEQRVSLDIDLWDFSELSTKCIKIVEFAKQLPGFTTLTIADQIT  
LLKAAACLDILMLRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFGPLTDLVFAFANQLPLEMDDAETGLLSAICLICGDRQD  
LEQPDRVDMLQEPLLEALKVYVRKRRPSRPHMFPKMLMKITDLRSISAKGAERVITLKMEIPGSMPLIQEMLENSEGLD

### hRAR $\alpha$ S369E (LBD) (176-421)

ESYTLTPEVHELIEKVRKAHQETFPALCQLGKYTTNNSSEQRVSLDIDLWDFSELSTKCIKIVEFAKQLPGFTTLTIADQIT  
LLKAAACLDILMLRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFGPLTDLVFAFANQLPLEMDDAETGLLSAICLICGDRQD  
LEQPDRVDMLQEPLLEALKVYVRKRRPERPHMFPKMLMKITDLRSISAKGAERVITLKMEIPGSMPLIQEMLENSEGLD

### hXR $\alpha$ (LBD) (223-462)

TSSANEDMPVERILEAEVAVEPKTETYVEANMGLNPPSPNDPVTNICQAADKQLFTLVEWAKRPHFSELPLDDQVILLRAG  
WNELLIASFSHRISIAVKDILLATGLHVHRNSAHSAGVGAIFDRVLTVELVSKMRDMQMDKTELGCLRAIVLFPDPSKGLSN  
PAEVEALREKVYASLEAYCKHKYPEQGRFAKLLLRPALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEAPHQMT

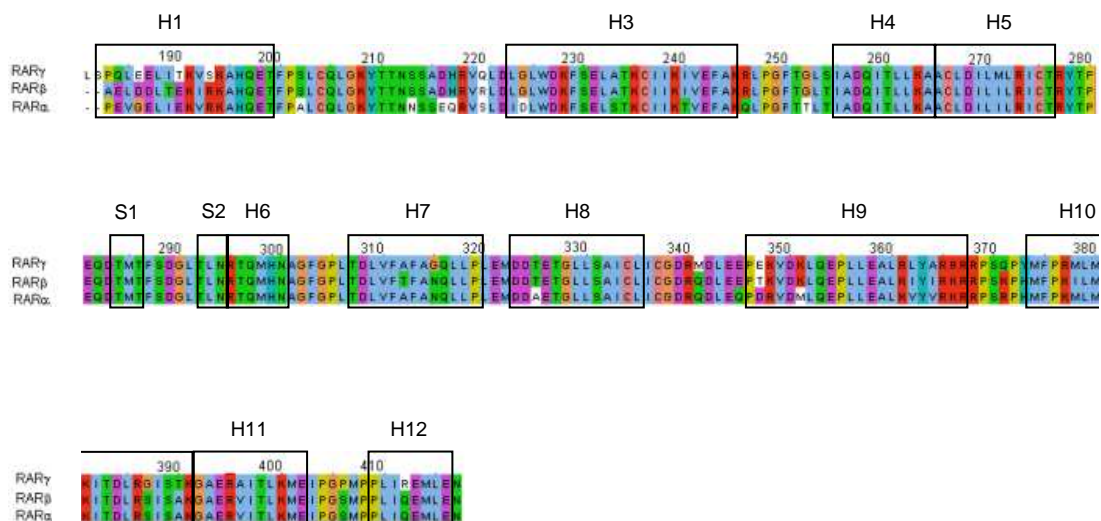
### hCycH (1-323)

MYHNSSQKRHWTFSSSEQLARLRADANRFRCKAVANGKVLNPNPVPFLEPHEEMTLCKYIEKRLLEFCSVFKPAMPRSV  
VGTACMYFKRFYLNNSVMEYHPRIMLTCAFLACKVDEFNVSSPQFVGNLRESPLGQEKALEQILEYELLIQQLNFHLIVH  
NPYRPFEGFLIDLKTRYPILNPEILRKTADDFLNRIALTDAYLLYTPSQIALTAILSSASRAGITMESYLSSEMLKENRTCLS  
QLDIMKSMRNLVKKYEPPESEVAVLKQKLERCHSAELALNVITKRRKGYEDDDYVSKKSKHEEEWTDDDLVESL

### hPKA (2-351)

GNAATAKKGSEVESVKEFLAKAKEDFLKKWENPTQNNAGLEDFERKKTTLGTGSFGRVMLVKHKATEQYYAMKILDQKQ  
VVKLLQIEHTLNEKRILQAVNFPFLVRLEYAFKDNSNLYMVMMEYVPGGEMFSLRRIGRFSEPHARFYAAQIVLTFEYLHS  
LDLIYRDLKPENLLIDHQGYIQVDFGFAKRVKGRTWLTCGTPEYLAPEIILSKGYNKAVDWWALGVLIYEMAAAGYPPFF  
ADQPIQIYEKIVSGKVRFPSPHFSSDLKDLLRNLLQVDLTKRFGNLKNGVSDIKTHKWFATTDWIAIYQRKVEAPFIPKFRGS  
GDTSNFDDYEEEDIRVSITEKCAKEFGEF

**Annex1.** Sequences of hRAR $\gamma$ , hRAR $\gamma$  S371E, hRAR $\alpha$ , hRAR $\alpha$ S369E, hXR $\alpha$  (LBDs), hCycH, hPKA.



**Annex 2.** Sequence alignment of LBDs of human RAR  $\alpha$ ,  $\beta$ ,  $\gamma$ . The elements of the secondary structure are boxed (H for helices and S for  $\beta$ -strands). Amino-acid numbering is referred to RAR $\gamma$ .