# Chapter I

# Introduction

## 1.1 Biological Background

### 1.1.1 Thyroid hormone function and diseases

Thyroid hormones (THs) play essential roles in numerous critical biological processes, such as differentiation, development and metabolism. THs are also found to be indispensable for the normal function of nearly every tissue in the body because of their impact on regulating oxygen consumption to generate the energy without which no biological activity can occur [1]. Abnormal increases or decreases in the amount of TH results in hyper- or hypothyroidism, which are among the most common endocrine maladies. The crucial function of THs is exerted through binding to the thyroid hormone receptor (TR), which is a member of the nuclear receptor protein family.

The decreases in TH or malfunction of its receptor can lead to hypothyroidism, which will cause mental retardation, referred as cretinism, if occurs during early development. During the past century, there has been a prevalence of endemic cretinism causing mental and physical impairment due to hypothyroidism linked with iodine deficiency, since iodine is one special resource that is required for the body to synthesize the thyroid hormones. In several countries in Europe where the soil is poor in iodine, cases of cretinism were very abundant and were considered to be genetically caused in the early

nineteenth century when TH function had not yet been resolved. The Clinical Society of London revealed for the first time the link between cretinism and hypothyroidism due to the destruction of the thyroid gland. After that, thyroid extracted from sheep began to be used as treatment for hypothyroidism. To date, hypothyroidism due to iodine deficiency is still a significant public health issue in some developing countries, which cause various levels of impairments during the mental and physical development in the population depending on individual physiological conditions. In the United States, reports of congenital hypothyroidism have also been increasing for the past decades, which may due to some complicate factors other than iodine intake [2].

When the body is exposed to elevated levels of THs, it can result in hyperthyroidism. Clinically, this state often manifests itself by symptoms such as weight loss, lowering of serum lipid levels, cardiac arrhythmias, heart failure, muscle weakness, bone loss in postmenopausal women, and even psychological symptoms as well [3, 4]. At present, treatment of hyperthyroidism is mostly directed to reducing overproduction of THs by inhibiting their synthesis or release or by ablating thyroid tissue with surgery or radioiodine. TR antagonists may have significant clinical use such as for treating hormone excess states, as it might quickly restore the euthyroid state and consequently improve the clinical manifestations mentioned above.

Given the critical role of TRs in mediating TH function, it is not surprising that mutations of TRs also have a deleterious effect on the human body. A tight linkage between mutations in TR and the syndrome of resistance to thyroid hormone (RTH) has been discovered since the last century [5]. To date, more than 100 mutations in the TRβ gene have been reported in hundreds of patients affected by the diseases [6]. So far,

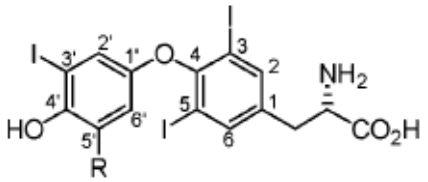
mutations in the TRα gene have not been reported in human, but abnormalities due to the TRα mutations have been observed in mice [7]. The two subtypes of the TR will be discussed in section 1.1.3 in more detail.

### 1.1.2 Thyroid hormone synthesis

Endogenous THs, 3,5,3',5'-tetraiodo-L-thyronine (T4, 1a), 3,5,3'-triiodo-L-thyronine (T3, 1b) (Figure 1), and their metabolites, regulate crucial genes throughout the organism and influence basal and adaptive metabolism, lipid levels, and bone/muscle function. In the early twentieth century, Kendall isolated T4 from thyroid tissue extract and proved its chemical composition [8]. In the middle of the same century, Gross and Pitt-Rivers successfully synthesized T3 and also proved its natural existence and function in the human body [9]. In following years, the regulatory roles of THs in metabolism and oxygen consumption, and amphibian development were studied more extensively [10]. For the synthesis of THs, iodide is actively transported into the thyroid gland cells by the $Na^+/I^-$ symporter (NIS). Afterward, iodide is oxidized by the thyroid peroxidase (TPO) in the presence of hydrogen peroxide and is incorporated into specific tyrosine residues in the glycoprotein thyroglobulin (Tg). These iodinated tyrosines can yield monoiodinated or diiodinated residues, which are enzymatically coupled to the synthesis of T4 and T3. Defects in the synthetic pathway of THs have been found to be the major cause of hereditary hypothyroidism in the regions with sufficient iodine uptake source [11].

Secretion of THs depends on endocytosis of iodinated Tg and carrier proteins. The unit structure of the thyroid gland is the follicle, which is basically like a sphere – a sphere of thyroid epithelial cells with a central acellular core that is filled with colloid. The iodinated Tg is stored extracelluarly in colloid. Tg undergoes endocytosis and

(A)




(B)

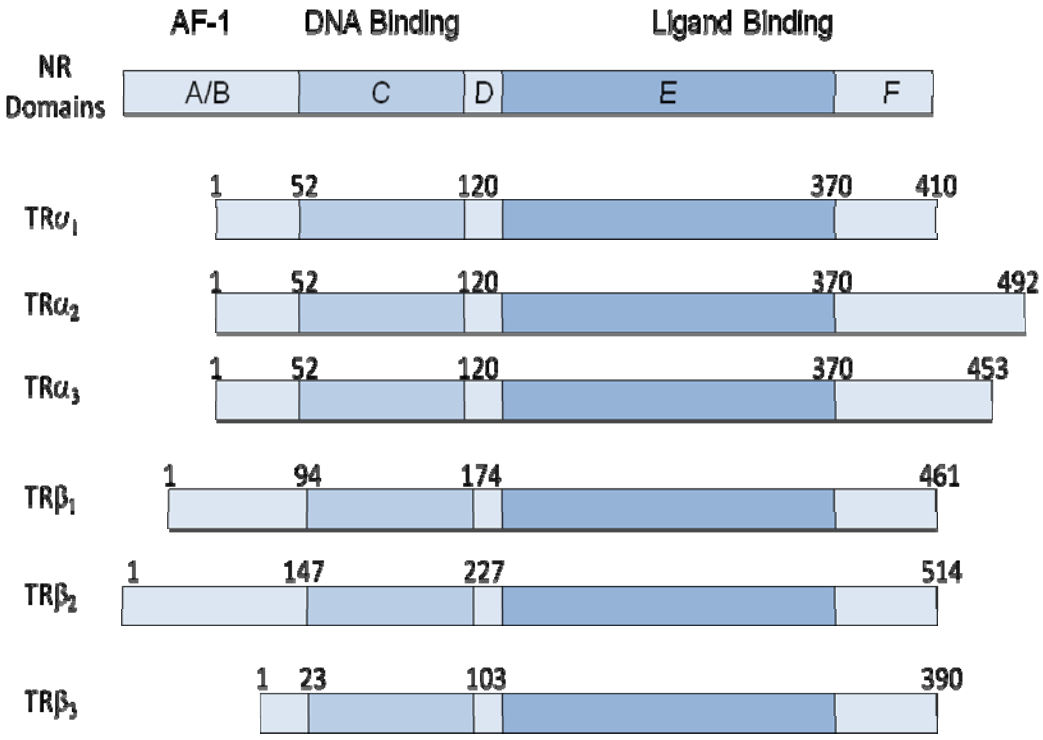



Figure 1.1: (A) Chemical structures of thyroid hormones. The R group equals an iodine atom in T4 and a hydrogen atom in T3. (B) Domain organization diagrams of the TR isoforms.

becomes internalized into phagolysosomes, where it is proteolytically digested releasing T4 and T3 into the circulation. The majority of released TH is in the form of T4, which is almost all bound to carrier proteins in serum such as thyroxine binding protein (TBP), albumin, and thyroid binding prealbumin. T3, likewise, is mostly found in a bound form with TBG and albumin. The total TH concentration in serum is approximately 90nM and 2nM for T4 and T3, respectively [12]. Despite the very small portion of free THs in serum, which is about 0.03% of T4 and 0.3% of T3 in humans, they are the active form of compounds generating thyroid hormone-related biological responses. T3, the more potent form of the THs, is actually produced mostly via deiodination of the less potent but more abundant T4 [13].

### 1.1.3 Thyroid hormone receptors: isoforms, domain organization

Shortly after the successful cloning of two different TR genes from chicken embryo and human placental cDNA libraries, amino acid sequence analysis indicated that TRs are homologs of the viral oncogene product v-erbA and are also related to the steroid hormone receptors. Later, it was confirmed that TR belongs to the large superfamily of nuclear hormone receptors. The nuclear receptor (NR) protein family consists of transcription factors that regulate gene expression in a ligand-dependent manner. Members of the NR superfamily comprise receptors for small ligands including the steroid hormones and non-steroid hormones. The subgroup of the steroid receptors includes the receptor for the male and female sex hormones, such as androgen receptor (AR), progesterone receptor (PR), and estrogen receptor (ER). The non-steroid receptors includes members such as thyroid hormone receptor (TR), retinoic acid receptor (RAR and RXR), vitamin D receptor (VDR), peroxisome proliferator-activated receptor

(PPAR), as well as receptors that bind diverse products of lipid metabolism, such as fatty acids, cholesterol metabolites and prostaglandins [14-16]. Within the NR superfamily, there are also a large number of so-called orphan receptors for which regulatory ligands have not yet been completely identified [17]. Nevertheless, the activities of many orphan receptors are probably also regulated by small-molecular-weight ligands in a similar manner as the better-known conventional nuclear receptors are. Nevertheless, some other mechanisms of functional regulation in NRs, such as phosphorylation in the activation function region (AF-1) in the N-terminal region, have also proven to be play an important role in regulating the coregulator recruitment and transcriptional activity status in a ligand-independent way [18, 19].

At the end of the last century, the sequence of the *Caenorhabditis elegans* genome revealed the presence of over 200 members of the NR family, suggesting a critical role of these proteins in evolution or environmental adaptation even at a very primitive stage [20]. On the other hand, mammalian genomes, with their more complicated, evolved and specialized components, are less likely to maintain such a large portion of the nuclear factors in the genome. So far, 24 distinct classes of NRs have been identified in humans. Nevertheless, these nuclear receptors still exert diverse roles in the regulation of growth, development, and homeostasis. Based on their importance in biology and medicine, as well as the relatively simple mechanism of regulation, the NR protein family represents one of the most intensively studied and best-understood classes of transcription factors at the molecular level.

The nuclear receptors share a common domain organization, which contains a variable A/B N-terminal region, a central conserved DNA binding domain (DBD), a less

conserved ligand binding domain (LBD) and a linker region between the DBD and LBD [21, 22].

Like nearly all the NRs, TRs also have the three major functional domains. Two different genes located on the chromosome 17 and chromosome 3 express two TR subtypes: TRα and TRβ, respectively [23, 24]. Furthermore, each RNA transcript of the TR genes can be alternatively spliced generating different isoforms: $TR\alpha_1$, $TR\alpha_2$, TR $\alpha_3$, $TR\beta_1$, $TR\beta_2$, and $TR\beta_3$. As shown in Figure 1.1B, the TR isoforms varied slightly in the DBD and LBD domains, but varied more in the length and sequence composition of the A/B domain and the far c-terminal region of the F domain starting from residue 370 [23, 25, 26]. Although the expression of all the TR isoforms is detectable in almost all the tissues, the expression levels have distinctive isoform-specific patterns in different tissues and are also dependent on developmental stages [27]. For example, $TR\beta_1$ is the most widely distributed isoform, which is relatively more abundant in the brain, liver, and kidney. $TR\beta_2$, on the other hand, is almost exclusively expressed in the pituitary and hypothalamus within adult human. Considering the restricted localization of its expression, TR-β2 may play a unique role in negative regulation by the thyroid hormone of centrally located genes, but its specific role in negative regulation of transcription has not been extensively studied yet [23, 27].

The transcription of many genes is affected by TRs. The most prevalent are positively regulated genes, on which TR causes transcription repression in the ligand-free state and $T_3$ activates transcription [28]. Other genes are negatively regulated by thyroid hormones, such as the thyrotropin releasing hormone (TRH), thyroid stimulating hormone (TSH) α [29] and β subunit [30], epidermal growth factor , β-myosin heavy chain, and keratin

genes. Despite less extensive studies than of the common positively regulated genes, negatively regulated gene transcription exhibits ligand-independent activation and ligand-dependent repression by TRs. In addition, many of the negatively regulated genes are involved in the establishing the feed-back modulating system, which are the primary regulatory factors of the TH level [31].

### 1.1.4 Functional domains

**Amino-terminal A/B domain**

The N-terminal A/B domain of TR incorporates a activation function-1 (AF-1) domain, which acts as a ligand-independent transactivation factor when fused with a heterologous DBD (e.g. Gal4) [32]. However, in the context of full-length TR, the activity of AF-1 appears to be ligand-dependent as a result of the corepressor release induced by ligand binding. In contrast to the direct ligand-dependent activity of AF-2 domain in the ligand binding domain, AF-1 action is less well understood. Nevertheless, recent studies have suggested that some of AF-2 interacting coactivators are also capable of mediating AF-1 function in NRs. For instance, transcription intermediary factor-2 (TIF-2) transactivates the AF-1 of RAR [33], and cAMP response element-binding protein (CBP) can associate with the A/B domain of PPAR to initiate transactivation function [34]. Similar AF-1 mediated coactivator functions are even more common in steroid receptors such as ER [19], PR and AR [35, 36]. In the TR A/B domain, which is the location of the major variation of the length and composition of amino acid sequence in the difference isoforms, it is not unexpected that the AF-1 function also varies in individual TR isoforms.

**DNA binding domain**

The DNA binding domain (DBD) is located at the central section of the TR and maintains the highest degree of sequence conservation through different species or among members of the NR protein family. The DBD has two zinc fingers, which are composed of four cysteines coordinated with a zinc ion, as a signature of DNA binding proteins. Substitution of any cysteine residue in the zinc finger abrogates DNA binding as well as transcription regulation by TR and other NR proteins [37-40]. The previously designated "P-box" region between the third and fourth cysteine in the N-terminal zinc-finger region has been proven to be crucial in DNA sequence recognition. Accordingly, the similarity in this "P-box" region, which has been observed in ER, RAR, RXR, and VDR as well as in TR, generates the same preferred sequence pattern in their DNA response elements [41-44]. Most genomic DNA response elements located at the promoter region of target genes contain two repeats of the consensus hexameric half-sites. The two half-sites can be arranged in different ways in the response element: the 'head to tail' assembled direct repeats (DR), the 'head to head' assembled palindrome (Pal), or the 'tail to tail' assembled inverted palindrome (IP or F2). In the previously determined structure of TR-RXR DBD heterodimer in complex with a direct repeat (DR4) TRE, the directionality of assembly of the two DBDs on the DR4 and the half-sites spacing are determined by the dimerization interactions between the TR DBD and RXR DBD [45]. The heterodimerization interface in the TR DBD is composed of residues upstream of the first zinc finger, inside the first zinc finger and the residues following the second zinc finger. The residues following the second zinc finger are designated as "T-box", which is also the linker connecting the core region to the C-

terminal extension region of the DBD. The dimerization interface in the RXR DBD, on the other hand, is located within the second zinc finger, which has been referred to as "D-box". This D-box has been demonstrated to be crucial for the discrimination of half-sites spacing in DRs by the different RXR-involved heterodimers [46, 47].

**Hinge region**

The hinge regions linking the DBD to the T3 ligand binding domain of TR has an amino acid sequence with possible nuclear localization function [48]. The highly positively charged region is also quite conserved in the NR protein family and shows homology similar to the simian virus 40T antigen nuclear localization sequence. However, unlike some steroid receptors, which are found to associate with heat shock proteins in the cytoplasm until the ligand-induced nuclear translocation, the TRs are more likely to be imported into the nucleus after their cytoplasmic synthesis, even in the absence of ligands [49]. Studies using the green fluorescent fusion proteins of wild-type TR or TR mutants disrupting the hinge region have demonstrated the importance of this region in determining the nuclear localization of TRs [50].

Other studies on coregulator recruitment by TR and other NRs have indicated the possible involvement of the hinge region in corepressor interaction. Mutations in the hinge region of TR or v-erbA have been shown to abrogate corepressor recruitment and basal transcription repression in the absence of ligands [51]. However, more recent work suggests that helices 3, 5 and 6 of the TR LBD are the main regions that contribute to the corepressor binding interface [52, 53]. In addition, others have also proposed that the hinge region may also play a part in TR dimerization in solution. In their TR LBD homodimer model built from small angle x-ray scattering (SAXS) data, the hinge region

of one TR LBD partner interacting with the corepressor binding interface of the other partner in the TR homodimer. They argued that this asymmetric dimerization is due to the partial sequence similarity of hinge region compared to the NR interaction domain of corepressors [54]. Nevertheless, the exact role of the hinge region in the corepressor interaction or dimerization is not yet fully understood.

**Ligand binding domain**

The ligand binding properties have been studied quite extensively for the TR LBD. TR LBDs are located at the C-terminal end of receptor proteins. With advances in structural studies of the TR LBD and other related NR LBDs in the past decade, we now know that the TR LBD consists of 12 α-helices and a short β-turn [55, 56]. The hydrophobic ligand binding pocket formed in the middle of the helices is able to incorporate the T3 efficiently with nanomolar affinity and TRα binds T3 with slightly higher affinity than the TRβ-1 [57]. Both agonists and antagonists of TR have been developed and tested with help from the structure directed drug design [58, 59]. With a better understanding of sub-type specific interactions and transcriptional activities induced by drug binding, novel therapeutics targeting TR-related disease may be developed with minimized side effects.

The ligand binding domain is not only necessary for TH binding but also fulfills other functional responsibilities. NR LBDs are found to be responsible for the DNA-independent oligomerization of nuclear receptors in solution [60-62]. The AF-2 region within NR LBDs is also proved to be directly involved in coregulator recruiting. The distinct binding interfaces of NR coactivators and corepressors can be exclusively established by alterations in the AF2 conformation [63-65]. In addition, biochemical and

transcriptional evidence suggests that structural changes induced by ligand binding influence both the recruitment of co-regulatory proteins as well as the oligomerization properties of the LBDs, providing a link between ligand binding, dimerization and transcriptional activity of the NRs [66-68].

### 1.1.5 Molecular mechanisms of thyroid hormone receptor action

Figure 1.2 illustrates a simplified molecular mechanism of the nuclear receptor action. The non-steroid NRs can form either homodimers or heterodimers in complex with RXR. Both kinds of NR dimers can bind specific DNA response elements in the promoter regions of target genes. In the presence of ligand, they can recruit coactivator proteins to activate transcription. In the absence of ligand, NR dimers can recruit corepressor proteins and inhibit transcription.

In contrast to steroid hormone receptors that are transcriptionally inactive without the ligand present, *apo* TRs can still enter the cell nucleus and bind to TREs even in the absence of ligand. Co-transfection studies have shown that unliganded TRs can repress basal transcription of positively regulated TR genes in vitro [69, 70]. In addition, early studies also demonstrated that T3 decreases TR homodimer binding to TREs, which led to the hypothesis that *apo* TR may repress basal transcription through homodimerization [71]. The unliganded TR has also been shown to interact directly with transcription factor IIB (TFIIB), a central component of the basal transcription machinery, and to potentially interfere with the assembly of the general transcription complex [37, 72, 73]. On the other hand, corepressor proteins are also critical components mediating transcriptional repression by unliganded TRs [74, 75].

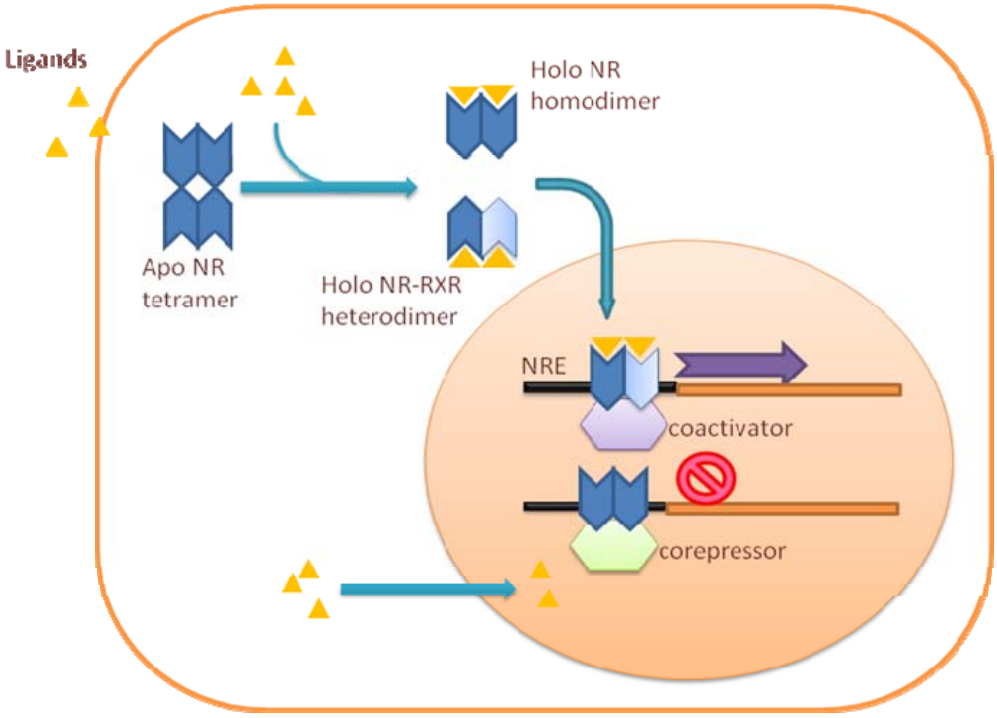



Figure 1.2: Molecular mechanism diagram of the nuclear receptor function.

Two major nuclear corepressor proteins associated with TR and RAR function are the NCoR (nuclear corepressor protein) and SMRT (silent mediator for retinoid and thyroid hormone receptors) [76, 77]. Recent studies have proved the interaction between the NR corepressors and other transcription repressors, Sin3 and histone deacetylases 1 (HDAC1), which are the mammalian homologs of the well-characterized transcriptional repressors RPD1 and RPD3 in yeast. As a result, the corepressors are able to directly or indirectly modify chromatin structure by the deacetylation of the histones, and thus play substantial roles in TR regulation of gene expression.

On the other hand, previous studies using Far-Western and coimmunoprecipitation have proved that liganded TR can interact with multiple nuclear proteins, potentially

forming a transcriptionally active complex [78]. Subsequently, some of the putative coactivator proteins identified from these assays are proved to interact with the AF-2 region, which has high homology among the NR proteins, in a ligand-dependent manner [79]. The steroid receptor coactivator-1 (SRC-1), which was originally identified to enhance transcription activation by steroid hormone receptors, can also associate with TR and enhance its ligand-dependent transactivation [80, 81]. Later, another 160kDa protein, TIFII/GRIP-1/SRC-2, with partial sequence homology to SRC-1, has also been identified as a coactivator for liganded TR and some other NR proteins [82]. A family of coactivators related to SRC-1 was then discovered based on these findings. In addition, a signature "LXXLL" sequence motif was revealed within multiple NR interaction sites located in the central part of coactivators. More recently, another group of coactivators was found as a key component in transactivation complexes (DRIPs, VDR interacting proteins; TRAPs TR-associated proteins) associating with liganded VDR and TR, but distinct from the complex assembled through SRCs. The "LXXLL" motif is also found in some critical protein components, such as DRIP205 and TRAP220, within the DRIP/TRAP coactivator complexes [83]. Other evidence also suggests there are different classes of coactivators in addition to SRC and DRIP/TRAP complexes [84], but the functional character of NR-related coactivators needs to be further elucidated.

There are still several aspects of the molecular mechanisms of TR function that remain to be fully understood. We are especially interested in the following questions: (1) how TR DBDs recognize and assemble on the different kinds of DNA response elements, what contributes to the specificity and affinity of DNA binding, and how dimerization of the DBDs communicates to dimerization of the LBDs; (2) what the TR LBD looks like in

the ligand-free state, and what conformational or dynamic switches may contribute to ligand-dependent transactivation functions; (3) what structural components determine the specificity and affinity of corepressor interactions with the TR and how ligand binding can modify the affinity for different coregulator proteins.

## 1.2 Experimental and Computational Techniques

### 1.2.1 Fluorescence anisotropy

We have used fluorescence anisotropy assays in measurements of the affinity of TR DBD binding to DNA. Fluorescence anisotropy is a method for measuring the binding between two molecules and can be used to quantify the binding constant for the interaction. The basic theory is that a fluorophore excited by polarized light will also emit polarized light. However, if the molecule is moving, it will tend to "scramble" the polarization of the light by radiating from a different orientation from which it was exited. The "scrambling" effect is largest when fluorophores freely tumble in solution and decreases in response to slower tumbling rates or rotational diffusion constants. As a result, protein interactions can be detected when one of the interacting partners is fused to a fluorophore and upon binding to the partner molecule a larger complex is formed which will tumble more slowly. Thus, the polarization of the emitted light will be increased while the "scrambling" effect is reduced. This technique works best if a small molecule is fused to a fluorophore and binds to a larger partner, which can maximize the difference in signal between the bound and the unbound states [85].

The affinity calculation is based on the measurements of the steady state emission-light polarization of the sample fluorescence during a series titration of the partner molecules into a sample containing the fluorophore fusion molecule. A linearly polarized

excitation light source excites fluorescent target molecules with transition moments aligned parallel to the incident polarization vector. The resultant fluorescence is collected and directed into two channels that measure the intensity of the fluorescence polarized both parallel and perpendicular to that of the excitation beam. With these two measurements, the fluorescence anisotropy, r, can be found from the following equation:

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}}$$

The I terms are the intensities measurements parallel ($\parallel$) and perpendicular ($\perp$) to the incident polarization.

### 1.2.2 Homology modeling

In order to take full advantage of the currently solved crystal structures of NR proteins, we have used the homology modeling technique to study the structure that can be adopt by the TR LBDs in the ligand-free state and the corepressor-bound state based on other related NR protein structures. The program Modeller v8.2 [86] is used for homology modeling of protein three-dimensional structures in this thesis. With an alignment of the target sequence to be modeled with known related structures as the template, Modeller can generate homology models containing all non-hydrogen atoms. Modeller implements comparative protein structure modeling by satisfaction of spatial restraints (3,4) and can perform many additional tasks including *de novo* modeling of loops in protein structures, optimization of various models of protein structure with respect to a flexibly defined objective function, multiple alignment of protein sequences and/or structures, clustering, searching of sequence databases, comparison of protein structures, etc.

### 1.2.3 Molecular dynamic simulation

Molecular dynamic simulations have been used in our study to evaluate the stability of the homology models and to demonstrate conformational dynamic properties correlated with TR functions as well. The essential concept of molecular dynamics simulation is quite straightforward: by defining a system of atoms and the forces between them, the solutions of the simultaneous equations of classical motions can present us a view of the time evolution of the system. However, due to large number of atoms and complexities in biomolecules, applications of this method have been very limited until the 21st century. With dramatic advances in the computer power, software and more reliable and accurate MD methodology, the MD technique has now been expanded to study a wide range of biomolecule systems.

The empirical force field implemented in Amber, which is one of the major MD program suite, has the following basic form. The equation represents the simplest functional form describing the interactions of atoms in amino acids that preserves the essential nature of molecules in condensed phases:

$$V(r) = \sum_{bonds} K_b(b - b_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2$$

$$+ \sum_{dihedrals} \frac{V_n}{2}(1 + \cos[n\varphi - \delta])$$

$$+ \sum_{nonbondij} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{r_{ij}}$$

The equation calculates the force potentials ($V_{(r)}$) of a certain atom in the system. The first three terms represent the interactions of the bonded atoms corresponding to the bond,

angle, and dihedral terms, respectively. The $K_b$ and $K_\theta$ are the associated force constants for the bond and angle terms, respectively. The b and θ are the current bond and angle values, respectively, and the values with a subscript of zero are the equilibrium values. The dihedral term has a periodicity of n and an offset of δ. The Vn and φ represent the periodic potential and the dihedral angle, respectively. The last term of summation is over all the nonbonded interactions, specifically the Lennard-Jones and Coulombic interactions between each pair of atoms. $A_{ij}$ and $B_{ij}$ are constants describing the Lennard-Jones interactions and $r_{ij}$ are the distance between the atom pairs, while $q_i$ and $q_j$ represent the electrostatic values of the atoms.

### 1.2.4 Free energy calculation

Free energy calculation was used in the analysis of the stability of my homology models of TR LBDs, and is also used to quantify the binding affinities between the TR LBD and corepressor peptides, which is not included in this thesis. We have used the MM_PBSA module of AMBER9 to evaluate the free energies of binding or absolute energies of the protein solute in various states during MD simulations. Representative structure snapshots were extracted at certain time intervals and used as input for the energy evaluation. The MM_PBSA methods help to combine the molecular mechanical energies of the solutes with the solvent-related energy components. The molecular mechanical energies are determined with the *sander* program from AMBER and represent the internal energy (bond, angle and dihedral) and *van der Waals* and electrostatic interactions. The electrostatic contribution to the solvation free energy is calculated with a numerical solver for the Poisson-Boltzmann (PB) method, for example, as implemented in the MM_PBSA program performed with a numerical solver in the

widely used DelPhi program [87, 88] or by generalized Born (GB) methods implemented in *sander*. On the other hand, the nonpolar contribution to the solvation free energy is determined with solvent-accessible-surface-area-dependent terms. The surface area is computed with *molsurf* program, which is based on analytical ideas primarily developed by Connolly [89].

As all biochemical processes such as the protein binding or protein-DNA binding take place in the cell are all in a liquid environment, the solvent effect will therefore need to be taken into account. Among solvent effect estimations, free energies of solvation and the solvation component of the free energies of coregulator protein binding are very challenging to calculate. Water molecules affect solvation in two aspects: they substantially screen the electrostatic interactions and result in hydrophobic interactions. Continuum solvent models are computationally tractable and reasonably accurately developed for free energy calculations [91, 92]. In these continuum models, solute molecules are treated as charged atoms inside a low dielectric volume (i.e., the volume of the solute) surrounded by a high dielectric medium (e.g., water). Two types of continuum approaches are commonly used to describe such a system, the Poisson-Boltzmann (PB) approaches [88, 92, 93] and the generalized Born (GB) approaches [93].

In summary, the free energy calculation protocol used in our study includes gas phase energy components and the solvation free energy contribution depicted as the following equation:

$$Etot = Eelec + Evdw + Eintra + Eelec,desolv + Enp,desolv - T\Delta S$$

*Eelec* and *Evdw* are the electrostatic and van der Waals energies. *Eintra* is the internal energy of the molecule including the covalent bond, angle and torsion energies.

*Eelec*,*desolv* and *Enp,desolv* are the electrostatic and non-polar desolvation energies, respectively. The *Eelec,desolv* energies can be evaluated using either the generalized Born (GB) or the Poisson-Boltzmann (PB) method. The non-polar desolvation energy is generally assumed to be proportional to the solvent-accessible surface area of the molecule with the constant of 0.0072kcal/(mol $Å^2$) [94], while the entropy term $T\Delta S$ is neglected in some of the energy evaluations. The entropy contribution can otherwise be included by turning on the Nmode module in MM_PBSA. The Nmode program is designed for molecular mechanics calculations on proteins and nucleic acids, using first and second derivative information to find local minima, transition states, and to perform the vibrational analyses [95,96].

## 1.3 Specific Aims

This dissertation aims to advance our understanding of the thyroid hormone receptor functions as they relate to static and dynamic structural properties of TR DBDs and LBDs in various functional states.

In Chapter 2, I aimed at study the interaction between TR DNA binding domain homodimer and the inverted palindromic response elements (F2). Previous biochemical assays have established several distinct features in the TR interaction with the F2 TRE, compared to the interaction with DR TREs. The DNA binding properties also communicate with the ligand binding, dimerization, subtype selection, and transcription regulation function of TRs. Previous structural studies on the NR-DNA interaction all focus on the DR TRE, which is the most abundant type of TRE found in the genome. However, as the second most abundant type of TRE that mediates the transcriptional regulation of TRs, the structural studies have been very limited in this area. Our structure

reveals a nearly symmetric structure of the TR DBD assembled on the F2 TRE where the base recognition contacts in the homodimer DNA complex are conserved relative to the previously published structure of a TR-RXR heterodimer DNA complex. Additionally, the T-box region of the DBD is discovered as a structural hinge that enables a large degree of flexibility in the position of the C-terminal extension helix that connects the DBD to the ligand binding domain. Although the isolated TR DBDs exist as monomers in solution, we have measured highly cooperative binding of the two TR DBD subunits on the inverted repeat DNA sequence. This suggests that elements of the DBD can influence the specific TR oligomerization on target genes and it is not just interactions between the ligand binding domains that are responsible for TR oligomerization at target genes. Mutational analysis shows that inter-subunit contacts at the DBD c-terminus account for some, but not all, of the cooperative homodimer TR binding to the inverted repeat class TRE.

In Chapter 3, we have utilized computational homology modeling and MD simulation techniques to study the *apo* TR LBD structure, which is not yet resolved by the experimental studies. Two closely related *apo* LBD structures that have been determined belong to the retinoid X receptor (RXR) and PPAR. However, the helix 12 regions, which are crucial in establishing the ligand-dependent transactivation function of TRs, display two distinct conformations in the two resolved *apo* NR LBDs. Therefore, it is still uncertain which model better represents the *apo* TR LBD conformation. We have then applied MD simulation and free energy calculations on both homology models to determine their relative stability as well as their dynamic properties that may contribute to TR functions. Our data support the PPAR-based *apo* TR LBD model in which helix 12

adopts a relatively stable conformation mimicking the active *holo* TR LBD structure. Consequently, the binding of corepressors to *apo* TR LBD will comprise the adjustment of the AF2 conformation and exposure of the hydrophobic interface only in the presence of corepressors.

In Chapter 4, in order to understand mechanisms underlying the specific binding of corepressor to TR, and ligand-dependent modification of the corepressor binding affinity, I have also used computation homology modeling and molecular dynamics (MD) simulations to study the corepressor interactions with the TR ligand binding domain in more detail. The homology modeling protocol for this study has been optimized during the process. In addition, an MD simulation was carried out on each complex model to further study its structural dynamics. All of the homology models of the TR LBD with different corepressor peptides from the NR interaction domains were stable enough to stay bound to the TR LBD during the MD simulation. From the MD simulations, we predict that the amino acid compositions flanking the CoRNR box are as important as the ones inside the box for directing nuclear receptor specificity and structural dynamics in binding. Specifically, data from the MD studies have also demonstrated that the unique ID domain present solely in NCoR but not in SMRT is crucial in determining the preferential binding by TR, which is consistent with some of the previous experimental results. In addition, during the simulation, it has also been observed that incorporation of the T3 ligand into the TR-corepressor complex will alter the mobility of several regions in the TR LBD and cause the corepressor peptide to drift away from the TR LBD.

# BIBLIOGRAPHY


1. Videla, L.A., et al., *Thyroid hormone calorigenesis and mitochondrial redox signaling: upregulation of gene expression.* Front Biosci, 2007. **12**: p. 1220-8.
2. Parks, J.S., et al., *The impact of transient hypothyroidism on the increasing rate of congenital hypothyroidism in the United States.* Pediatrics. **125 Suppl 2**: p. S54-63.
3. Bunevicius, R. and A.J. Prange, Jr., *Thyroid disease and mental disorders: cause and effect or only comorbidity?* Curr Opin Psychiatry.
4. Kendall-Taylor, P., *Hyperthyroidism.* Br Med J, 1972. **2**(5809): p. 337-41.
5. Usala, S.J., et al., *Tight linkage between the syndrome of generalized thyroid hormone resistance and the human c-erbA beta gene.* Mol Endocrinol, 1988. **2**(12): p. 1217-20.
6. Refetoff, S. and A.M. Dumitrescu, *Syndromes of reduced sensitivity to thyroid hormone: genetic defects in hormone receptors, cell transporters and deiodination.* Best Pract Res Clin Endocrinol Metab, 2007. **21**(2): p. 277-305.
7. Cheng, S.Y., *Thyroid hormone receptor mutations and disease: beyond thyroid hormone resistance.* Trends Endocrinol Metab, 2005. **16**(4): p. 176-82.
8. Kendall, E.C., *The isolation in crystalline form of the compound containing iosine which occurs in the thyroid: its chemical nature and physiological activity.* Trans Assoc Am Physicians, 1915(30): p. 30.
9. Gross, J., Pitt-Rivers, R., *Tiiodothyronine in relation to thyroid physiology.* Rec. Prog. Horm. Res., 1954(10): p. 20.

10. Shi, Y.B., et al., *Tadpole competence and tissue-specific temporal regulation of amphibian metamorphosis: roles of thyroid hormone and its receptors.* Bioessays, 1996. **18**(5): p. 391-9.
11. Kopp, P., *Pendred's syndrome and genetic defects in thyroid hormone synthesis.* Rev Endocr Metab Disord, 2000. **1**(1-2): p. 109-21.
12. Taugod, A., *Hormone Synthesis*. Werner and Ingbar's The Thyroid, ed. L. Braverman, Utiger, R. 1996, Philandelphia: Lippincott-Raven. 33.
13. Kohrle, J., *The selenoenzyme family of deiodinase isozymes controls local thyroid hormone availability.* Rev Endocr Metab Disord, 2000. **1**(1-2): p. 49-58.
14. Beato, M., P. Herrlich, and G. Schutz, *Steroid hormone receptors: many actors in search of a plot.* Cell, 1995. **83**(6): p. 851-7.
15. Chambon, P., *The molecular and genetic dissection of the retinoid signaling pathway.* Recent Prog Horm Res, 1995. **50**: p. 317-32.
16. Mangelsdorf, D.J., et al., *The nuclear receptor superfamily: the second decade.* Cell, 1995. **83**(6): p. 835-9.
17. Mangelsdorf, D.J. and R.M. Evans, *The RXR heterodimers and orphan receptors.* Cell, 1995. **83**(6): p. 841-50.
18. Hammer, G.D., et al., *Phosphorylation of the nuclear receptor SF-1 modulates cofactor recruitment: integration of hormone signaling in reproduction and stress.* Mol Cell, 1999. **3**(4): p. 521-6.
19. Tremblay, A., et al., *Ligand-independent recruitment of SRC-1 to estrogen receptor beta through phosphorylation of activation function AF-1.* Mol Cell, 1999. **3**(4): p. 513-9.

20. Sluder, A.E., et al., *The nuclear receptor superfamily has undergone extensive proliferation and diversification in nematodes.* Genome Res, 1999. **9**(2): p. 103-20.

21. Bain, D.L., et al., *Nuclear receptor structure: implications for function.* Annu Rev Physiol, 2007. **69**: p. 201-20.

22. Wurtz, J.M., et al., *A canonical structure for the ligand-binding domain of nuclear receptors.* Nat Struct Biol, 1996. **3**(2): p. 206.

23. Williams, G.R., *Cloning and characterization of two novel thyroid hormone receptor beta isoforms.* Mol Cell Biol, 2000. **20**(22): p. 8329-42.

24. Mitsuhashi, T., G. Tennyson, and V. Nikodem, *Nucleotide sequence of novel cDNAs generated by alternative splicing of a rat thyroid hormone receptor gene transcript.* Nucleic Acids Res, 1988. **16**(12): p. 5697.

25. Lazar, M.A., *Thyroid hormone receptors: multiple forms, multiple possibilities.* Endocr Rev, 1993. **14**(2): p. 184-93.

26. Mitsuhashi, T., G.E. Tennyson, and V.M. Nikodem, *Alternative splicing generates messages encoding rat c-erbA proteins that do not bind thyroid hormone.* Proc Natl Acad Sci U S A, 1988. **85**(16): p. 5804-8.

27. Cheng, S.Y., *Multiple mechanisms for regulation of the transcriptional activity of thyroid hormone receptors.* Rev Endocr Metab Disord, 2000. **1**(1-2): p. 9-18.

28. Glass, C.K., et al., *A c-erb-A binding site in rat growth hormone gene mediates trans-activation by thyroid hormone.* Nature, 1987. **329**(6141): p. 738-41.

29. Burnside, J., et al., *Thyroid hormone regulation of the rat glycoprotein hormone alpha-subunit gene promoter activity.* J Biol Chem, 1989. **264**(12): p. 6886-91.

30. Bodenner, D.L., et al., *A detailed functional and structural analysis of a major thyroid hormone inhibitory element in the human thyrotropin beta-subunit gene.* J Biol Chem, 1991. **266**(32): p. 21666-73.

31. Shupnik, M.A., E.C. Ridgway, and W.W. Chin, *Molecular biology of thyrotropin.* Endocr Rev, 1989. **10**(4): p. 459-75.

32. Metzger, D., et al., *Characterization of the amino-terminal transcriptional activation function of the human estrogen receptor in animal and yeast cells.* J Biol Chem, 1995. **270**(16): p. 9535-42.

33. Bommer, M., et al., *TIF2 mediates the synergy between RARalpha 1 activation functions AF-1 and AF-2.* J Biol Chem, 2002. **277**(40): p. 37961-6.

34. Gelman, L., et al., *p300 interacts with the N- and C-terminal part of PPARgamma2 in a ligand-independent and -dependent manner, respectively.* J Biol Chem, 1999. **274**(12): p. 7681-8.

35. Onate, S.A., et al., *The steroid receptor coactivator-1 contains multiple receptor interacting and activation domains that cooperatively enhance the activation function 1 (AF1) and AF2 domains of steroid receptors.* J Biol Chem, 1998. **273**(20): p. 12101-8.

36. He, B., et al., *Activation function 2 in the human androgen receptor ligand binding domain mediates interdomain communication with the NH(2)-terminal domain.* J Biol Chem, 1999. **274**(52): p. 37219-25.

37. Yen, P.M., et al., *Studies on the repression of basal transcription (silencing) by artificial and natural human thyroid hormone receptor-beta mutants.* Endocrinology, 1995. **136**(7): p. 2845-51.

38. Green, S., et al., *The N-terminal DNA-binding 'zinc finger' of the oestrogen and glucocorticoid receptors determines target gene specificity.* EMBO J, 1988. **7**(10): p. 3037-44.

39. Nagaya, T., L.D. Madison, and J.L. Jameson, *Thyroid hormone receptor mutants that cause resistance to thyroid hormone. Evidence for receptor competition for DNA sequences in target genes.* J Biol Chem, 1992. **267**(18): p. 13014-9.

40. O'Donnell, A.L., et al., *Thyroid hormone receptor mutations that interfere with transcriptional activation also interfere with receptor interaction with a nuclear protein.* Mol Endocrinol, 1991. **5**(1): p. 94-9.

41. Mader, S., et al., *Three amino acids of the oestrogen receptor are essential to its ability to distinguish an oestrogen from a glucocorticoid-responsive element.* Nature, 1989. **338**(6212): p. 271-4.

42. Nelson, C.C., S.C. Hendy, and P.J. Romaniuk, *Relationship between P-box amino acid sequence and DNA binding specificity of the thyroid hormone receptor. The effects of sequences flanking half-sites in thyroid hormone response elements.* J Biol Chem, 1995. **270**(28): p. 16988-94.

43. Rastinejad, F., et al., *Structural determinants of nuclear receptor assembly on DNA direct repeats.* Nature, 1995. **375**(6528): p. 203-11.

44. Umesono, K. and R.M. Evans, *Determinants of target gene specificity for steroid/thyroid hormone receptors.* Cell, 1989. **57**(7): p. 1139-46.

45. Zilliacus, J., et al., *Structural determinants of DNA-binding specificity by steroid receptors.* Mol Endocrinol, 1995. **9**(4): p. 389-400.

46. Luisi, B.F., J.W. Schwabe, and L.P. Freedman, *The steroid/nuclear receptors: from three-dimensional structure to complex function.* Vitam Horm, 1994. **49**: p. 1-47.

47. Umesono, K., et al., *Direct repeats as selective response elements for the thyroid hormone, retinoic acid, and vitamin D3 receptors.* Cell, 1991. **65**(7): p. 1255-66.

48. Evans, R.M., *The steroid and thyroid hormone receptor superfamily.* Science, 1988. **240**(4854): p. 889-95.

49. Dalman, F.C., et al., *In contrast to the glucocorticoid receptor, the thyroid hormone receptor is translated in the DNA binding state and is not associated with hsp90.* J Biol Chem, 1990. **265**(7): p. 3615-8.

50. Zhu, X.G., et al., *Hormone-induced translocation of thyroid hormone receptors in living cells visualized using a receptor green fluorescent protein chimera.* J Biol Chem, 1998. **273**(42): p. 27058-63.

51. Li, H., et al., *Characterization of receptor interaction and transcriptional repression by the corepressor SMRT.* Mol Endocrinol, 1997. **11**(13): p. 2025-37.

52. Hu, X. and M.A. Lazar, *The CoRNR motif controls the recruitment of corepressors by nuclear hormone receptors.* Nature, 1999. **402**(6757): p. 93-6.

53. Nagy, L., et al., *Mechanism of corepressor binding and release from nuclear hormone receptors.* Genes Dev, 1999. **13**(24): p. 3209-16.

54. Nascimento, A.S., et al., *Structural rearrangements in the thyroid hormone receptor hinge domain and their putative role in the receptor function.* J Mol Biol, 2006. **360**(3): p. 586-98.

55. Bourguet, W., et al., *Crystal structure of the ligand-binding domain of the human nuclear receptor RXR-alpha.* Nature, 1995. **375**(6530): p. 377-82.

56. Sandler, B., et al., *Thyroxine-thyroid hormone receptor interactions.* J Biol Chem, 2004. **279**(53): p. 55801-8.

57. Schueler, P.A., et al., *Binding of 3,5,3'-triiodothyronine (T3) and its analogs to the in vitro translational products of c-erbA protooncogenes: differences in the affinity of the alpha- and beta-forms for the acetic acid analog and failure of the human testis and kidney alpha-2 products to bind T3.* Mol Endocrinol, 1990. **4**(2): p. 227-34.

58. Chiellini, G., et al., *A high-affinity subtype-selective agonist ligand for the thyroid hormone receptor.* Chem Biol, 1998. **5**(6): p. 299-306.

59. Koehler, K., et al., *Thyroid receptor ligands. 6. A high affinity "direct antagonist" selective for the thyroid hormone receptor.* J Med Chem, 2006. **49**(23): p. 6635-7.

60. Rosen, E.D., E.G. Beninghof, and R.J. Koenig, *Dimerization interfaces of thyroid hormone, retinoic acid, vitamin D, and retinoid X receptors.* J Biol Chem, 1993. **268**(16): p. 11534-41.

61. McKenna, N.J. and B.W. O'Malley, *Combinatorial control of gene expression by nuclear receptors and coregulators.* Cell, 2002. **108**(4): p. 465-74.

62. Figueira, A.C., et al., *Low-resolution structures of thyroid hormone receptor dimers and tetramers in solution.* Biochemistry, 2007. **46**(5): p. 1273-83.

63. Darimont, B.D., et al., *Structure and specificity of nuclear receptor-coactivator interactions.* Genes Dev, 1998. **12**(21): p. 3343-56.

64. Nolte, R.T., et al., *Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma.* Nature, 1998. **395**(6698): p. 137-43.

65. Xu, H.E., et al., *Structural basis for antagonist-mediated recruitment of nuclear co-repressors by PPARalpha.* Nature, 2002. **415**(6873): p. 813-7.

66. Figueira, A.C., et al., *Human thyroid receptor forms tetramers in solution, which dissociate into dimers upon ligand binding.* Cell Biochem Biophys, 2006. **44**(3): p. 453-62.

67. Jeyakumar, M., et al., *Quantification of ligand-regulated nuclear receptor corepressor and coactivator binding, key interactions determining ligand potency and efficacy for the thyroid hormone receptor.* Biochemistry, 2008. **47**(28): p. 7465-76.

68. Yen, P.M., *Physiological and molecular basis of thyroid hormone action.* Physiol Rev, 2001. **81**(3): p. 1097-142.

69. Baniahmad, A., A.C. Kohne, and R. Renkawitz, *A transferable silencing domain is present in the thyroid hormone receptor, in the v-erbA oncogene product and in the retinoic acid receptor.* EMBO J, 1992. **11**(3): p. 1015-23.

70. Zhang, X.K., et al., *Ligand-binding domain of thyroid hormone receptors modulates DNA binding and determines their bifunctional roles.* New Biol, 1991. **3**(2): p. 169-81.

71. Yen, P.M., et al., *Triiodothyronine (T3) decreases binding to DNA by T3-receptor homodimers but not receptor-auxiliary protein heterodimers.* J Biol Chem, 1992. **267**(6): p. 3565-8.

72. Hadzic, E., et al., *A 10-amino-acid sequence in the N-terminal A/B domain of thyroid hormone receptor alpha is essential for transcriptional activation and interaction with the general transcription factor TFIIB.* Mol Cell Biol, 1995. **15**(8): p. 4507-17.

73. Tong, G.X., M.R. Tanen, and M.K. Bagchi, *Ligand modulates the interaction of thyroid hormone receptor beta with the basal transcription machinery.* J Biol Chem, 1995. **270**(18): p. 10601-11.

74. Qi, J.S., et al., *The ligand-binding domains of the thyroid hormone/retinoid receptor gene subfamily function in vivo to mediate heterodimerization, gene silencing, and transactivation.* Mol Cell Biol, 1995. **15**(3): p. 1817-25.

75. Tong, G.X., et al., *Transcriptional silencing by unliganded thyroid hormone receptor beta requires a soluble corepressor that interacts with the ligand-binding domain of the receptor.* Mol Cell Biol, 1996. **16**(5): p. 1909-20.

76. Fondell, J.D., H. Ge, and R.G. Roeder, *Ligand induction of a transcriptionally active thyroid hormone receptor coactivator complex.* Proc Natl Acad Sci U S A, 1996. **93**(16): p. 8329-33.

77. Barettino, D., M.M. Vivanco Ruiz, and H.G. Stunnenberg, *Characterization of the ligand-dependent transactivation domain of thyroid hormone receptor.* EMBO J, 1994. **13**(13): p. 3039-49.

78. Takeshita, A., et al., *Molecular cloning and properties of a full-length putative thyroid hormone receptor coactivator.* Endocrinology, 1996. **137**(8): p. 3594-7.

79. Kamei, Y., et al., *A CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors.* Cell, 1996. **85**(3): p. 403-14.

80. Hong, H., et al., *GRIP1, a transcriptional coactivator for the AF-2 transactivation domain of steroid, thyroid, retinoid, and vitamin D receptors.* Mol Cell Biol, 1997. **17**(5): p. 2735-44.

81. Rachez, C., et al., *The DRIP complex and SRC-1/p160 coactivators share similar nuclear receptor binding determinants but constitute functionally distinct complexes.* Mol Cell Biol, 2000. **20**(8): p. 2718-26.

82. McKenna, N.J., R.B. Lanz, and B.W. O'Malley, *Nuclear receptor coregulators: cellular and molecular biology.* Endocr Rev, 1999. **20**(3): p. 321-44.

83. Anderson, B.J., et al., *Using fluorophore-labeled oligonucleotides to measure affinities of protein-DNA interactions.* Methods Enzymol, 2008. **450**: p. 253-72.

84. Marti-Renom, M.A., et al., *Comparative protein structure modeling of genes and genomes.* Annu Rev Biophys Biomol Struct, 2000. **29**: p. 291-325.

85. Sigalov, G., A. Fenley, and A. Onufriev, *Analytical electrostatics for biomolecules: beyond the generalized Born approximation.* J Chem Phys, 2006. **124**(12): p. 124902.

86. Rocchia, W., et al., *Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects.* J Comput Chem, 2002. **23**(1): p. 128-37.

87. Connolly, M.L., *Solvent-accessible surfaces of proteins and nucleic acids.* Science, 1983. **221**(4612): p. 709-13.

88. Gallicchio, E.K., M.M.; Levy, R.M., *Enthalpy-entropy and cavity decomposition of alkane hydration free energies: Numerical results and implications for theories of hydrophobic solvation.* J. Phys. Chem., 2000(104): p. 15.

89. Cramer, C.J. and D.G. Truhlar, *Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics.* Chem Rev, 1999. **99**(8): p. 2161-2200.

90. Simonson, T., *Macromolecular electrostatics: continuum models and their growing pains.* Curr Opin Struct Biol, 2001. **11**(2): p. 243-52.

91. Baker, N.A., et al., *Electrostatics of nanosystems: application to microtubules and the ribosome.* Proc Natl Acad Sci U S A, 2001. **98**(18): p. 10037-41.

92. Gilson, M.K. and B.H. Honig, *Calculation of electrostatic potentials in an enzyme active site.* Nature, 1987. **330**(6143): p. 84-6.

93. Bashford, D. and D.A. Case, *Generalized born models of macromolecular solvation effects.* Annu Rev Phys Chem, 2000. **51**: p. 129-52.

94. Gohlke, H., C. Kiel, and D.A. Case, *Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes.* J Mol Biol, 2003. **330**(4): p. 891-913.

95. D.A. Pearlman., et al., *AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules.* Comp. Phys. Commun. 1995, **91**: p.1-41.

96. D.A. Case, et al., *The Amber biomolecular simulation programs*. J. Computat. Chem. 2005, **26**: p.1668-88.

# Chapter II

# Structure of the Thyroid Hormone Receptor DNA Binding Domain Homodimer on the Inverted Palindrome Response Elements

## 2.1 Abstract

Thyroid hormone receptor (TR), as a member of the nuclear hormone receptor family, can recognize and bind different classes of DNA response element targets as either a monomer, a homo-oligomer, or a hetero-oligomer. We report here the first crystal structure of a homodimer TR DNA binding domain (DBD) in complex with an inverted repeat class of thyroid response element (F2 TRE). The structure reveals a nearly symmetric structure of the TR DBD assembled on the F2 TRE where the base recognition contacts in the homodimer DNA complex are conserved relative to the previously published structure of a TR-RXR heterodimer DNA complex [1]. The new structure reveals that the T-box region of the DBD can function as a structural hinge that enables a large degree of flexibility in the position of the C-terminal extension helix that connects the DBD to the ligand binding domain. Although the isolated TR DBDs exist as

monomers in solution, we have measured highly cooperative binding of the two TR DBD subunits on the inverted repeat DNA sequence. This suggests that elements of the DBD can influence the specific TR oligomerization on target genes and it is not just interactions between the ligand binding domains that are responsible for TR oligomerization at target genes. Mutational analysis shows that inter-subunit contacts at the DBD c-terminus account for some, but not all, of the cooperative homodimer TR binding to the inverted repeat class TRE.

## 2.2 Introduction

The thyroid hormone receptor is a member of the nuclear receptor (NR) protein family. The nuclear hormone receptors are able modulate gene transcription levels in response to hormone ligands by virtue of their ability to recognize specific DNA sequences and recruit a variety of co-activator and co-repressor proteins [2]. The specific DNA sequences recognized by NR proteins are referred to as nuclear response elements (NREs), which are located at the promoter regions of target genes [3, 4]. The binding of NR ligands to the receptors regulates their function by inducing changes to the conformation and dynamics of the receptors that can influence receptor oligomerization, nuclear localization, DNA binding, and co-regulator protein recruitment [5-9].

In general, one or more nuclear receptors bind to a specific DNA response element and regulate the expression of proximal genes by recruiting coregulator proteins that often possess chromatin remodeling functionality. A consensus hexameric DNA sequence 5′-AGGTCA-3′ is the core recognition element or half-site for TR. This sequence is recognized by and will bind to a single DBD subunit. The same AGGTCA consensus sequence is conserved across all the other non-steroid nuclear receptors,

including the 9-*cis* retinoic acid receptor (RXR), the 9-*cis*/all-*trans* retinoic acid receptor (RAR), vitamin D receptor (VDR) and the peroxisome proliferator-activated receptor (PPAR). Most genomic DNA response elements contain two repeats of the consensus hexameric half-sites. As shown in Figure 2.1A, the two half-sites can be arranged in different ways in the response element: the 'head to tail' assembled direct repeats (DR), the 'head to head' assembled palindrome (Pal) or the 'tail to tail' assembled inverted palindrome (IP or F2). In addition, the two half-sites can be separated by various numbers of base pairs in between, often indicated by a number at the end of the NRE classification [10-12]. For example, the IP6 class of NRE is composed of two inverted-palindrome consensus sequences separated from one another by 6 base-pairs. So, while the consensus half-sites are conserved across NR family members, it is believed that specific interactions between the NR LBD and DBD subunits are able to define the preferred oligomerization partners on different classes of NREs. For example, in the case of direct repeats, distinctive affinities for different RXR heterodimers are observed when the inter-half-sites spacing varies between 1 and 5 base pairs [4, 11].

The detailed nature of the NR oligomerization is of considerable interest because the ability of the NRs to bind to specific DNA response elements as either homo or hetero-oligomers, and with a specific directionality on the DNA, has a major impact on NR mediated regulation of transcription of specific genes [13, 14]. Despite the known oligomerization tendencies of the NR LBDs, specific features of the DBDs can also play an important role in defining the preferred NR oligomerization on target DNA response elements. Although the isolated DBDs exist only as monomers in solution, their assembly on target DNA response elements has been reported to be highly cooperative [15, 16]. It

has also been suggested that the isolated DBDs of the nuclear receptors can generate similar patterns of DNA sequence selectivity and homo or hetero-oligomerization as the full-length receptor [15-17]. The structure of the RXR-TR heterodimer bound to a DR4 NRE illustrated how inter-subunit steric constraints between the DBDs dictate that only one of the two possible polarities of the RXR-TR binding is allowed [1].

Within the nuclear receptor superfamily, the DNA binding domains have the highest degree of conservation with respect to both the sequence and the structural architecture. The core region of the DBDs contains about 66 amino acids, with over 40% sequence identity within the NR superfamily [18]. This core sequence comprises two α-helices and two zinc-binding motifs, each coordinated by four conserved cysteine residues. Following the core region is the less conserved and highly charged C-terminal extension (CTE) region, which has been observed to adopt an α-helical conformation in TR [1]. *In vitro* binding assays have found that the CTE region can be essential in mediating protein-protein and protein-DNA interactions required for the cooperative DNA binding by homodimeric NR DBDs [19, 20]. The T-box region, located at the beginning of the CTE (Figure 2.1B), incorporates the residues contributing to the TR-RXR DBD heterodimer interface when bound to DR4 that establishes the downstream positioning of the TR DBD in the heterodimer-DNA complex [1]. The CTE region is also capable of forming extensive contacts with the phosphate backbone of the spacer nucleotides and thus specifies the proper spacing requirements between the half-sites recognized by the DBDs [21]. Additionally, conformational changes in some CTEs have been observed in response to DNA binding. The CTE of RXR, which exists as α-helix in solution, becomes

an extended loop when bound to DNA and the CTEs of some orphan receptors exist in an unstructured form in solution becoming structured when bound to DNA [22-24].

To date, structural and biochemical studies of sequence specificity have mostly focused on the DNA binding mechanism of non-steroid NR DBDs forming RXR heterodimers in complex with DR response elements. A structure of a NR DBD homodimer in complex with an inverted palindrome class of DNA response element has not yet been reported. While the DRs are the most abundant class of TRE, the F2 is the second most prevalent class of TRE identified within natural promoters [25, 26]. It has also been shown that homodimers of the TRβ isoform preferentially bind and regulate genes possessing the F2 TRE over genes with the DR class TRE [27-29]. As a consequence of the preferred homodimeric TR binding, *in vitro* experiments have demonstrated that F2 containing genes are strongly responsive to T3. T3 hormone induced transcriptional activation at F2 TRE containing genes is also preserved in the absence of the heterodimer partner RXR [30-32]. An analogous RXR-independent pathway has also been reported in the case of VDR [12], and these distinct combinatorial modes of NR activation add a level of diversity to transcriptional regulation by limited numbers of NRs. In addition, homodimer occupied TREs are believed to more efficiently recruit corepressors compared to the TR-RXR heterodimer complexes in the absence of T3, which results in the most efficient transcriptional repression in the F2 TRE regulated genes [30, 31]. While the addition of T3 ligand appears to destabilize the homodimeric DNA binding of full length TR [33], evidence also suggest that the TR homodimer F2 complexes can become restabilized by recruitment of coactivator proteins [30, 34-36]. Some genes such as SERCa2 and GH, with promoters containing mixed types of TREs,

may obtain specific T3 responses through the combinatorial effects of varied TR dimer recruitment and T3 responsiveness to each TRE [37, 38].

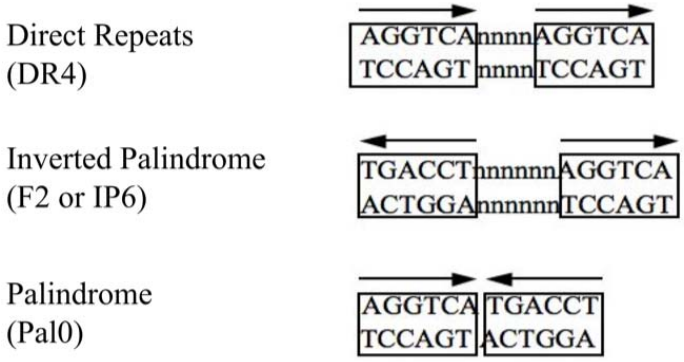
To better understand the detailed features of the TR homodimer bound to the F2 TRE, we have determined the crystal structure of the TRβDBD homodimer assembled on the idealized F2 TRE. The complex structure maintains the protein-DNA contacts responsible for the specific DNA sequence recognition observed in the heterodimeric RXR/TR complex. The structure also illustrates how the 6 base pair inter-half-site spacing of the F2 TRE orients the two subunits of TRβDBD in a position such that the C-terminal tails are able to contact each other. We have also established that the homodimeric binding of two TRβDBD subunits to F2 TRE is highly cooperative and the interaction at the c-terminal tails of the DBDs is one component that enhances the cooperative F2 DNA binding. Finally, we observe two distinct orientations of the CTE helix in the crystal structure, a structural change that is facilitated by apparent flexibility of the T-box region which lies between the core DBD region and the CTE helix. This structural flexibility could allow the LBDs to adopt a range of relative conformations in the TR homodimer while still remaining bound to the F2 TRE.

## 2.3 Results

### 2.3.1 Overall architecture

The DNA used in the structure is a 22 base-pair F2 sequence, with 2 base-pair blunt overhangs at both ends. The TR protein construct boundaries were designed to include the complete unit required for the formation of TR homodimers on an F2 TRE, including residues 106-207 of the rat TRβDBD, extending from the start of the DBD core region

(A)




(B)

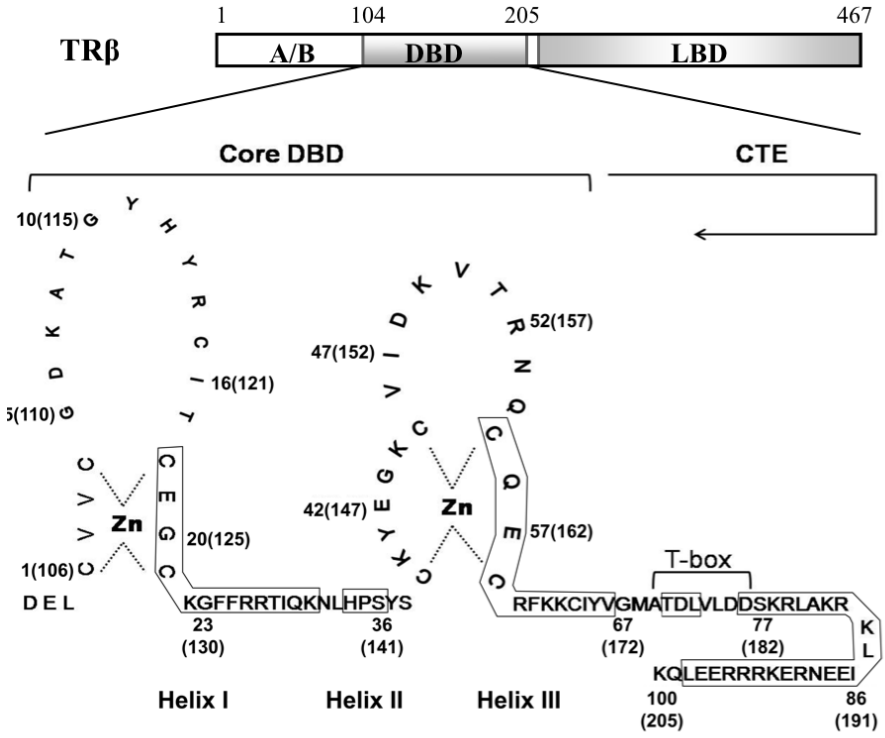



Figure 2.1: The three major kinds of TREs and the protein sequence have been used in this study. (A) A simplified overview of the three types of half-site organizations in TREs. The preferred inter-half-site spacing is 4bp, 6bp and 0bp for the DR, F2 and Pal, respectively. Previous gel mobility shift assay studies have shown that TR homodimers bind preferentially to the F2 TRE (B) Schematic diagram of domain organization of full length TR and the TRβDBD construct used in crystallization. The residue numbering scheme is referenced to the first Zn-coordinating cysteine residue. The authentic numbers appear in the parentheses. The DBD contains the conserved core DBD (residue 1-66) and a C-terminal extension including the T-box region. The helix region in the construct is boxed and the three helices in the core DBD is labeled at the bottom.

through the end of the CTE. The protein construct diagram is shown in Figure 2.1B. To be consistent with the convention used in the heterodimer TR/RXR structure, the original residue numbers from the full-length rat protein are indicated in parenthesis and the numbers outside the parenthesis use a numbering scheme beginning from the first cysteine of the N-terminal zinc-finger.

The asymmetric until of the crystal is comprised of two distinct conformations of the homodimer/DNA complex, i.e. four copies of the TRβDBD chain and two copies of the double helical TREs. The F2 TRE in the crystal structure maintains a regular B-form conformation and lacks significant distortion in base packing or backbone conformation relative to canonical B-form DNA. The two DNA duplexes in the asymmetric unit are arranged in an anti-parallel way with a rotation of approximately 140 degrees along the axis with respect to each other. The two subunits of TRβDBD are assembled nearly symmetrically about the central base-pair of each DNA duplex in the expected tail-to-tail conformation, as predicted from the IP6 consensus site sequence. The description that follows describes the common features of the multiple conformations found in the asymmetric unit and will highlight the significant differences between the four subunits.

Through crystallographic symmetry, the DNA fragments form contacts with their symmetrical mates in the neighboring unit cells creating a pseudo-continuous double helix within the crystal (Figure 2.3D). As is well known, B-form DNA exhibits a double-helical periodicity of ~10.5 base pairs per helical turn in the canonical conformation [39]. Consequently, a DNA sequence of 21bp will make approximately two complete 360 degree turns of the double helix. Because the sequence of the oligonucleotides used in the crystal structure is 22bp, the packing of the DNA oligonucleotides compensates for the

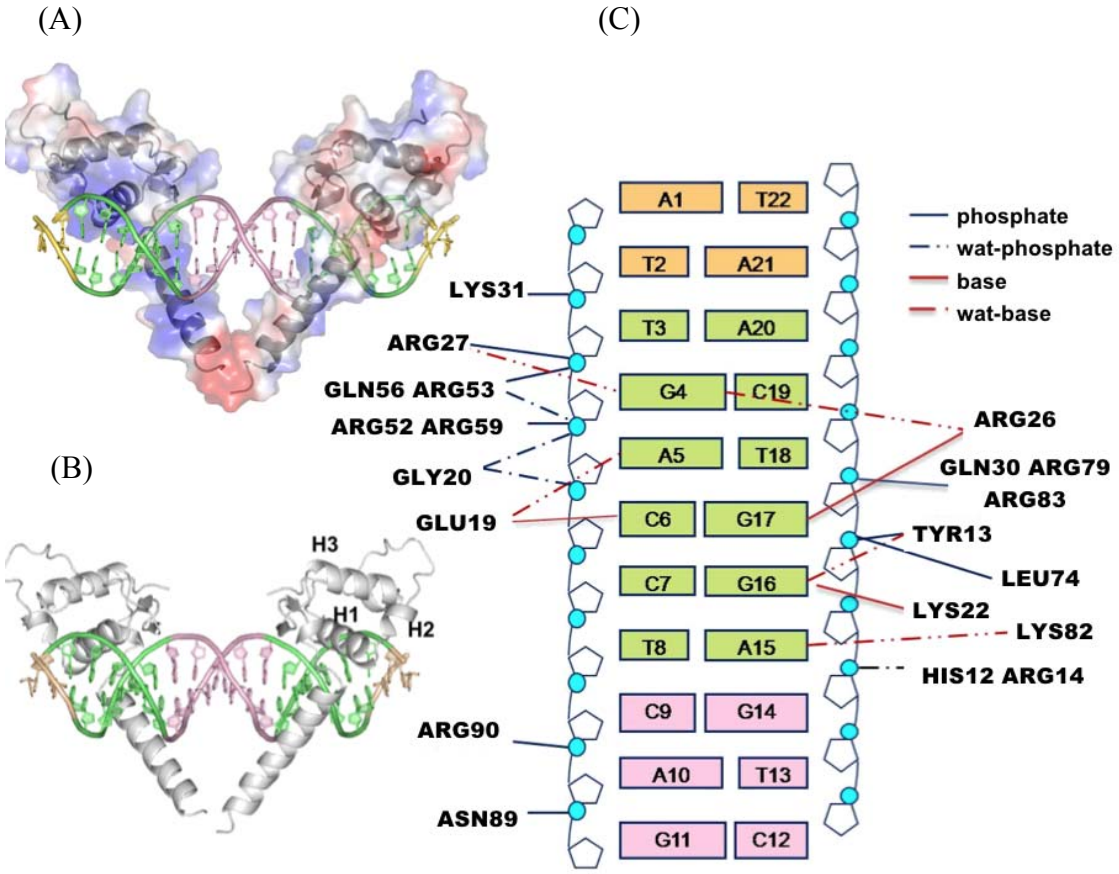



Figure 2.2: (A) The overall structure of TRβDBD homodimer on the inverted palindrome sequence of AGGTCA with 6-bp separation. The consensus half-site base pairs are colored in green, while the spacing base pairs are colored in pink. The electrostatic potential is drawn as colored surface: blue is positive and red is negative. The figure is made in Pymol. (B) The same structure representation of the DBD dimer and F2 TRE with the three helices of the core DBD marked as H1, H2 and H3. (C) Summary of the protein-DNA interactions. Base-specific contacts are colored in red and the backbone contacts are colored in navy blue. And the dashed lines represent water mediated interactions, while the solid lines represent direct contacts.

one extra helical step by incorporating a negative 34 degree (left-handed) helical stack at the DNA-DNA crystallographic interface, just enough to maintain the helical periodicity of the DNA within the crystal. Although shorter DNA sequences were screened in crystallographic trails, the extra base pair in the 3' flanking sequence of the downstream F2 consensus site turns out to be required in order to create enough distance along the DNA for two DBD monomers to avoid a steric clash when they bind side by side along the pseudo-continuous DNA helix created at the crystallographic junction of adjacent DNA duplexes. The net result is that the TR DBD homodimers are assembled in an identical orientation repeated along a pseudo-continuous double helical DNA that incorporates a single left-handed helical step at the DNA:DNA interface. The conformations of the four DBD subunits are largely conserved except for the fact that one subunit adopts a different rotational orientation of C-terminal extended helix, described in more detail below.

### 2.3.2 The Homodimer TR DBD assembly on F2 TRE

This crystal structure of the TR-DNA complex is the first one elucidating the organization of a non-steroid NR DBD homodimer on an inverted palindrome class of response element. All previously published non-steroid NR structures are DBD complexes with direct repeat response elements. The synthetic double-stranded DNA used in crystallization corresponds to an idealized F2 TRE (Figure 2.1A). The core regions of the two TRβDBD molecules in the homodimer-F2 complex bind to opposite facing half-sites of the DNA, as along a twofold rotational pseudo-symmetry axis at the center of the DNA (Figure 2.2A). Most non-steroid NRs recognize one or two of the consensus hexameric sequence 5`-AGGTCA-3` or a variation of it [40]. The TR

| Table 1. Data collection and Refinement Statistics | |
|---|---|
| Diffraction Data | |
| Space group | $P2_1$ |
| Unit cell dimension (Å) | 73.7 x 83.5 x 75.7 |
| Resolution (Å) | 2.40 |
| Unique reflections | 35442 |
| Redundancy (last shell) | 6.7 |
| Completeness (%) | 98.1 (99.2)[a] |
| Average I/σ | 27.1 (1.6) [a] |
| FOM | 0.78 |
| Crystallographic Refinement | |
| Resolution range (Å) | 42-2.40 |
| Reflections | 34766 |
| Total number of atoms (excluding H) | |
| Protein | 3285 |
| DNA | 1788 |
| Zinc | 8 |
| Water | 243 |
| RMS bond lengths (Å) | 0.007 |
| RMS bond angles (˚) | 1.312 |
| R value (%) | 18.3 b |
| Rfree (%) | 23.0 c |

Table1: Data collection and Refinement Statistics

[a] Number in parentheses is for the highest resolution shell.

[b] $R_{work} = \sum |F_{obs} - F_{calc}| \Big/ \sum F_{obs}$, where the summation is over all data used in refinement.

[c] $R_{free}$ was calculated using the 5% of reflection data that were excluded during the crystallographic refinement.

recognizes target genes with direct repeats, palindromic or inverted palindromic hexamer half-sites. The present 2.4Å structure provides a good picture of the TRβDBD arrangement on the F2 TRE and the interaction details between DNA and the protein side chains (Figure 2.2A and B). The surface representation in Figure 2.2A is colored by electrostatic potential, in which the blue represents positively charged regions and red represents negatively charged regions. The protein interface wrapping around the DNA is highly positively charged, complementing the DNA backbone phosphates. On the F2 TRE, the two DBDs are separated from one another, with possible inter-subunit contacts occurring at the C-terminal end. Figure 2.3C shows a close up view at the major groove of the consensus DNA half-site (colored in green), where the proteins make contacts with both the bases and backbone phosphates of the DNA. A schematic representation of the protein-DNA interactions on half of the F2 TRE is summarized in Figure 2.2C. The interactions of the four TRβDBD subunits are identical with the exception of the TR-DNA backbone interactions between the TR CTE and the inter-half-site DNA that are lost in the single subunit where the CTE adopts the alternative conformation.

The core regions of DBD are the most conserved region among the NRs and consist of two non-equivalent zinc finger motifs. Each zinc ion is coordinated by four highly conserved cysteine residues. The conserved sequences result in a highly conserved structural organization shared by all the NR DBDs. The tertiary structure of TR DBD contains three main helices in the core region. The N-terminal helix (helix I, Cys18–Lys31), also known as the recognition helix, directly interacts with the DNA half-site at the major groove forming the most extensive protein-DNA contacts (Figure 2.3C). As shown in Figure 2.2B, helix III (Cys55–Val66) overlays the N-terminal helix I almost

perpendicularly and contributes to the stability of the DBD by forming a hydrophobic core region with residues from both helix I and II (Leu33-Ser36). The hydrophobic core of TRβDBD is generated by a group of conserved aromatic residues, Phe24, Phe25, and Phe50, aided by several hydrophobic residues, Ile29, Ile64 and Val73. The overall fold of the DBD subunit is preserved in both homodimer and heterodimer TR DBD assemblies.

Previous studies have demonstrated that the T-box region, corresponding to residues 69-76 at the c-terminal of the core DBD, is crucial for the formation of homodimers on both DR4 and F2 TREs. Deletion of the T-box will abolish the formation of full length TRβ dimers on either class of TRE and dramatically reduces transcription activation on F2 TRE [29]. Similarly, the essential role of T-box in cooperative dimer assembly has also been revealed in the homo- and certain hetero-dimers formed by RXR DBDs on direct repeats kind of DNA response elements [14, 24, 29, 41, 42]. As shown in our TRβDBD structure, the T-box region, which forms a loop region interrupted by a helical turn (Thr70-Leu72), is responsible connecting the core region of TR DBD and the CTE region. The CTE region, which forms a distinctive long helix in TR DBD, is either much shorter or more disordered in the other NR DBD structures that have been determined [1, 19, 42, 43]. Unlike the RXR/TR heterodimer structure, the increased spacing between the two DNA consensus half-sites in the F2 TRE results in no inter-subunit contacts within the T-box region.

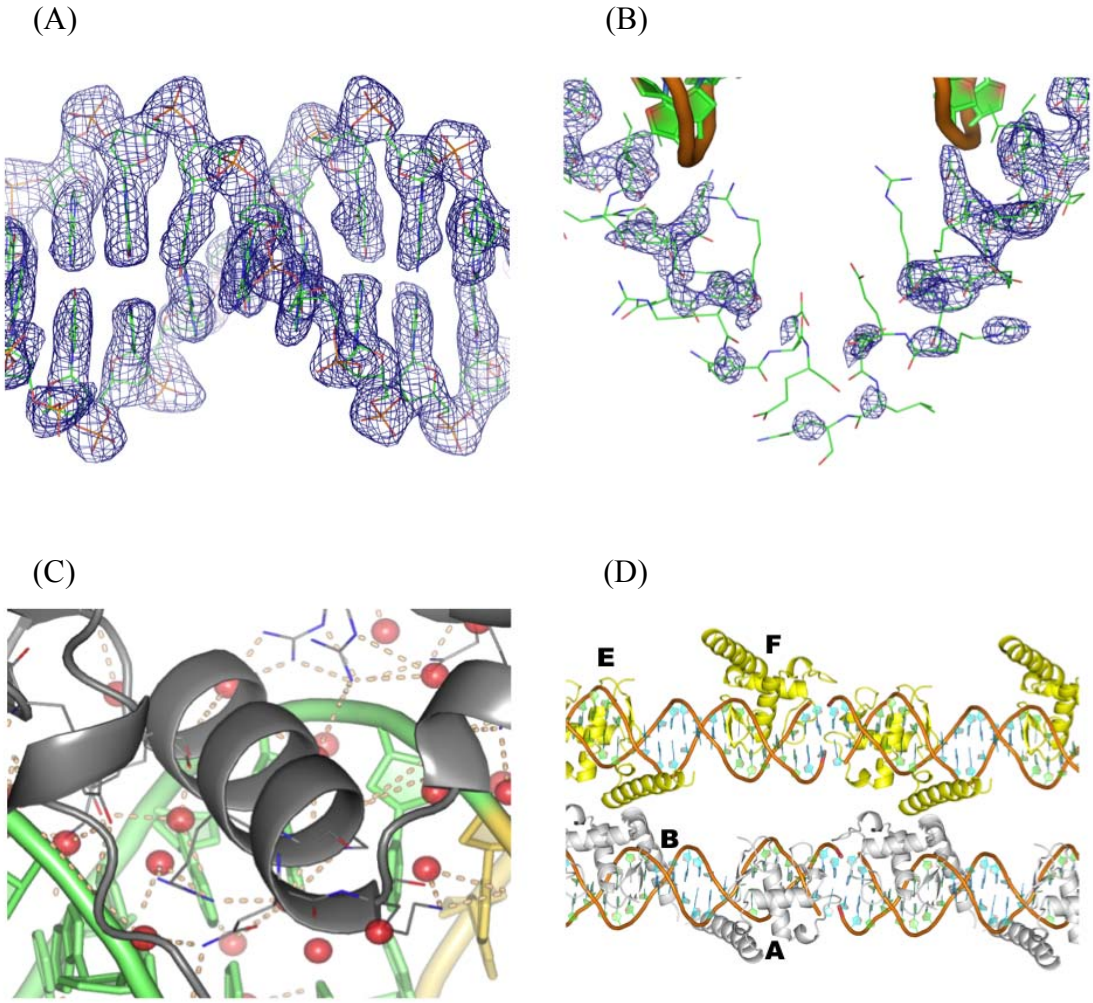



Figure 2.3: Detailed view of electron density at the (A) DNA double helix region and (B) the c-terminal region of the TR. Both electron density maps (2F0-Fc) are contoured at 1.2 sigma level. (C) The zoomed in view of the protein-DNA contacts at the major groove of the consensus half-site. The red spheres represent water molecules. The side chains of protein residues involved in the DNA interaction are shown as sticks. The direct and water mediated interactions are all drawn as dotted lines. (D) The molecule packing in the crystal. The four protein chains in an asymmetric unit is marked A, B, E, F, with respect to the pdb structure file. The neighboring asymmetric units lie in parallel revealing the slight condensed packing at the connecting region of the DNA in the center of the view.

### 2.3.3 Recognition of the DNA sequence

The TRβDBD homodimer recognizes the consensus half sites of the F2 TRE roughly in the same manner as in the complex of TR-RXR heterodimer and the DR4 TRE. The backbone rmsds between the four TRβDBD subunits in the homodimer complex (chain designations A/B and E/F in the pdb file) and the single TR subunit in the heterodimer structure (PDB ID: 2NLL) are 1.3 Å, 1.1 Å, 1.5 Å and 5.3 Å, respectively. The core regions of all the subunits, which solely contribute to the sequence specific DNA interactions, have a very low degree of variation. The interaction between the protein and DNA involves both the major and minor grooves of the DNA. The recognition helix engages the hexameric half site at the major groove by forming the sequence-specific interactions with the bases of the half-site, which are further stabilized with numerous water mediated backbone interactions (Figure 2.2C and Figure 2.3C). The specific interactions with the bases determining the half-sites are well preserved among all the subunits, involving protein residues such as Tyr13, Glu19, Lys22, Arg26, Arg27 and Lys82. The RXR DBD, with a lysine in the equivalent position of Arg26 in TR, abolishes the interaction with both of the adjacent bases G4 and G17. Previous studies have demonstrated that the substitution of Arg to Lys at this position results in more limited base recognition and a subsequent reduction in DNA binding affinity in RXR compared to TR [24, 44]. The bases involved in direct protein contacts include G4, A5 and C6 of the forward strand, G16 and G17 of the reverse strand. Of particular note, G16 is involved in multiple direct or water mediated base interactions with residue Lys13 and Lys22 as well as backbone phosphate interactions with Lys13 and Leu74. Outside of the consensus DNA, the bases of the spacer sequence are not involved in any direct protein

contacts. In addition to these base specific contacts, numerous direct or indirect hydrogen bonds with the backbone phosphates further reinforce the consensus half site binding at the major groove. Several cases of multiple protein residues contacting one base or one residue contacting multiple backbone phosphate are observed in the four subunits. Although the ordered water molecules observed have slight variations among the four subunits in the asymmetric unit, the interactions between the core regions of DBDs and the major groove DNA are almost identical, with the exception of only a couple water mediated contacts with the backbone phosphates.

The backbone interactions at the minor groove DNA are less well preserved among the four DBD copies in the asymmetric unit. The diversity is presumably due to the flexibility of the c-terminal region and varied crystal packing forces on each individual DBD subunit. A single direct interaction between Asn89 and backbone phosphate of A10 in the reverse strand is well preserved among three of the DBD subunits: chain A, B and E. The c-terminal extension helices account for all contacts with the flanking DNA half sites, which are achieved via interactions with the phosphate backbone. Because the RXR DBD lacks the long CTE extension of TR, it is not able to make any of the backbone contacts with the flanking sequence minor groove, and hence the TR/TR homodimer has more extensive DNA interactions than the RXR/TR heterodimer. A similar auxiliary binding in the inter-half-site flanking DNA is observed in some orphan receptors, which are also observed to bind to monomeric NREs, similar to TR, for example, NGFI-B [41], ROR/RZR [45, 46] and RevErb [44]. The more extended CTE loop found in structures of RevErb-DR2 complex even enables several residues be embedded into the minor groove to make direct base contacts outside the consensus hexameric half-site [44]. Unlike

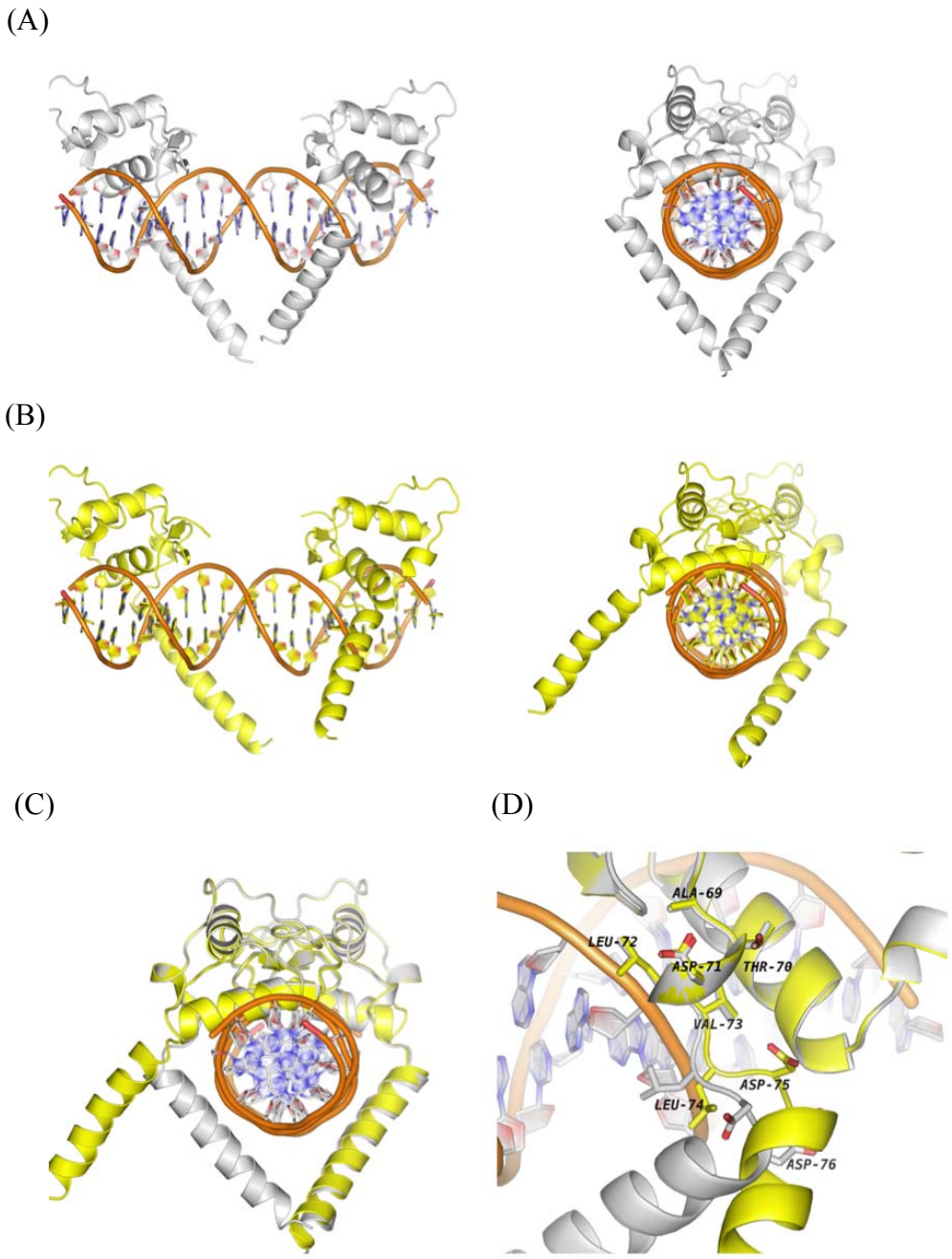



Figure 2.4: CTE helix displacement revealed in the two copies of TRβDBD-F2 TRE complexes in an asymmetric unit. (A) The protein colored in grey is in the original conformation similar to the structure in TR-RXR and DR4 complex. (B) The protein colored in yellow involves one subunit in alternative CTE orientation. From left to right shows the side-view and top-view. (C) Overlay of the two DBD-F2 complexes. (D) The zoom in view at the T-box region of the two conformations.

RevErb, the bulky and rigid helical conformation of the TR DBD CTE ensures that the minor groove contacts remain relatively superficial and involve only the backbone of DNA. Nevertheless, the major and minor groove interactions preserved by TR DBD is more extensive than those in the DBDs of RXR, RAR or ER [42, 47]. As expected, the region starting from the T-box to the C-terminal end has the highest structural deviation from the heterodimer structure, with an average 1.8Å rmsd of the four copies in the asymmetric unit.

### 2.3.4 Conformation flexibility of the CTE helix

In one of the two TRβDBD homodimer complexes in the crystal we have found that the original inter-subunit contacts at the CTE tail region are disrupted and the CTE helix of chain F is rotated approximately 70˚ relative to the canonical position seen in the other three subunits and the TR-RXR heterodimer. The angle formed between the CTE helix and helix I (the DNA recognition helix) is ~80 deg in the canonical conformation and a much more obtuse ~150 deg in the alternative conformation (Figure 2.4A and Figure 2.4B). As a result of this rotation, the two DBD subunits bound to the F2 DNA can no longer maintain any possible inter-subunit contact. Instead, the CTE of the F DBD subunits is involved in crystal symmetry related contacts with a TRβDBD subunit in a neighboring asymmetric unit. The two CTE helices of the neighboring subunits are closely aligned in an anti-parallel configuration incorporating several inter-subunit electrostatic and *van der waals* interactions. The rotational displacement of the CTE helix in this subunit is facilitated by the apparent flexibility of the T-box region linking the core region and the CTE of TRβDBD such that the T-box serves as a structural hinge between the core DBD and the CTE helix. While this large-scale displacement of the c-

terminal helix is seemly stabilized and introduced by crystal packing forces between the CTE helix of chain F and the CTE helix of chain A in the neighboring unit cell, it is likely exposing an inherent flexible propensity for this region of the DBD.

The two observed conformations of the T-box suggest that the movement of the CTE helix and the configuration of the T-box become strongly correlated when the T-box functions as a structural hinge. The specific T-box sequence "ATDLVLDD" of the TRβDBD has features that contribute stabilizing interactions to both of the observed CTE conformations. In the canonical CTE conformation of the TRβDBD (Figure 2.4A and 2.4D), Asp76, near the c-terminus of the T-box, adopts an orientation where the side chain stabilizes and wedges open the open form of the hinge by electrostatically capping the otherwise exposed c-terminal end of CTE helix dipole at the apex of the hinge. When the DBD adopts the alternative CTE conformation, the c-terminal end of the helix squeezes the back side of the hinge shut and the side chain of Asp76 is able to rotate so that the side chain points outward into the solvent to make room for the CTE helix. At the same time, because the CTE helix becomes more collinear with the helix I recognition helix, the opposing dipole of the N-terminal end of helix I is able to substitute for Asp76 to partially restabilize the C-terminal dipole of the CTE. Val73 in the middle of the T-box is the most highly conserved residue in the NR family T-box region. RXR, RAR, VDR and PPAR all have a hydrophobic residue at this position. The indispensable participation of this residue in forming the hydrophobic core of the DBD is the main reason underlying the conservation. Val73 is also able to play the additional role of serving as the hydrophobic pivot point for the TR CTE hinge. The other residues in the T-box region are much less conserved and we suggest that the resulting sequence

diversity contributes to the different preferred conformations of the CTE region. For instance, the RXR T-box replaces Ala69 with a Lysine and has a glutamine instead of a Leucine at position 74. These two changes allow RXR to associate with the DNA backbone phosphates in a much tighter way which makes it much less likely to be able to adopt the extended alternative CTE conformation of the TRβDBD [24, 42]. A similar residue swap of Leu74 to arginine, as in the PPAR and the liver receptor homologue-1 (LRH-1), also results in a more intimate association of the T-box with the DNA backbone. As a result of these additional interactions, the PPAR and LRH-1 both adopt a sharp hairpin turn at the T-box region that allows the relatively unstructured CTE to be embedded into the DNA minor groove [48, 49]. VDR, with a rather long CTE helix like TR, is also not able to insert into the DNA minor groove and its t-box lacks residues to interact with the DNA backbone [50].

The flexibility at the T-box region may be important for the accommodation of the full length TR onto various arrangements of the consensus half sites. In addition to the flexible hinge region between the DBD and LBD, the flexible T-box region adds another degree of freedom to the DBD orientation with respect to the TR LBD. Previously reported dynamic light scattering studies, have shown that full-length TR proteins exist predominately as homodimers or homotetramers in solution, presumably facilitated by the dimerization interface in the LBD. In corresponding structural models generated from small angle X-ray scattering experiments, the c-terminal tails of the DBDs were found to be positioned closely together, tied to each other by the symmetric dimers formed by the LBD dimerization interactions [51]. The LBD interactions are also expected to persist when the full-length protein binds to the TRE. Because the TRβDBD has been shown to

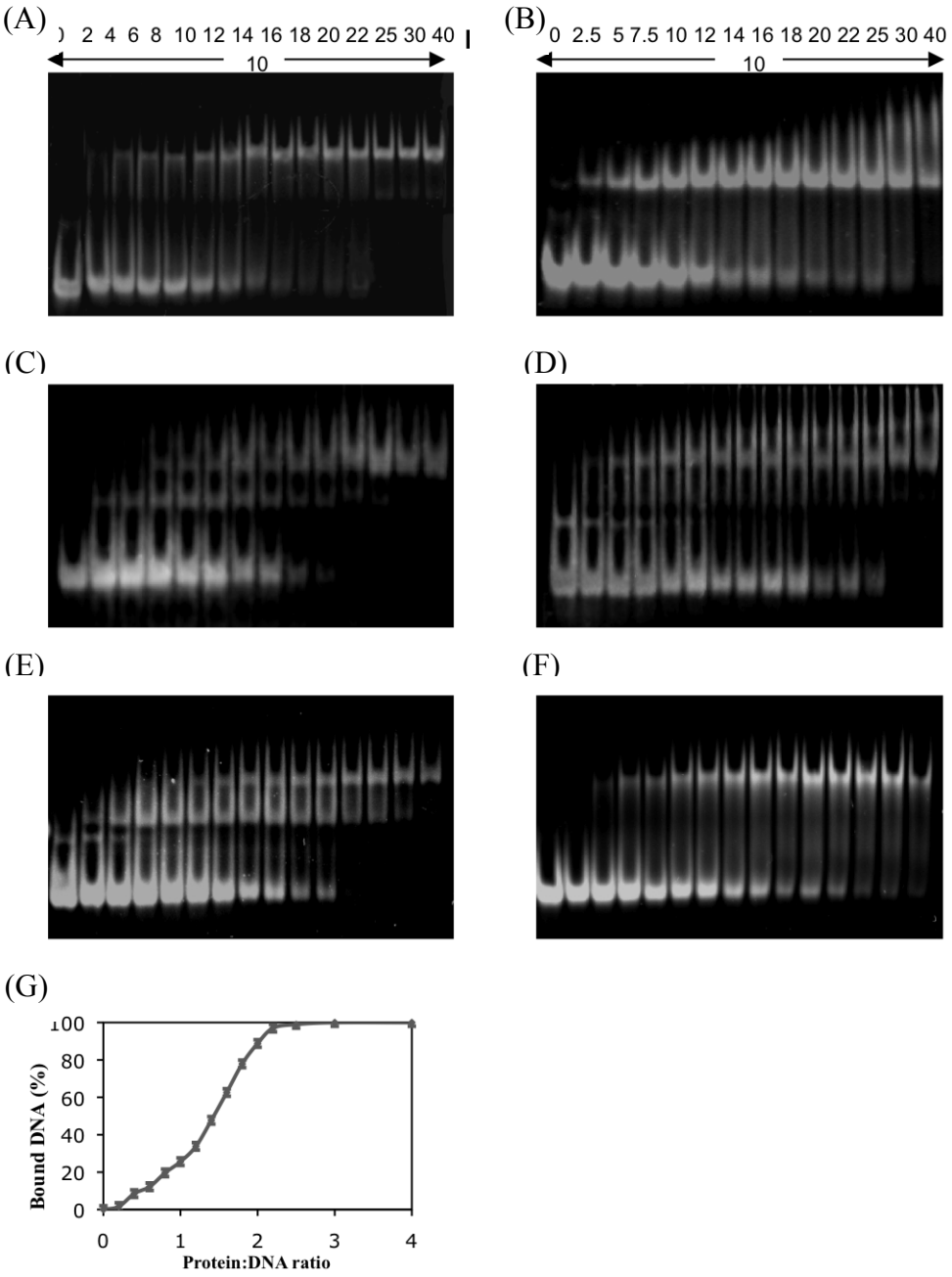



Figure 2.5: Native gel shifts assays of the wild-type TRβDBD with various TREs and the TRβDBDΔ8 deletion mutant with the F2 TRE. The TRE DNA concentration in the assay is at 10iM. While in each well of the gel, the protein concentration is increased from 0 to 40iM. The gel is stained with 1X SYBR nucleic acid gel stain (Invitrogen) to show the free and bound DNA bands. The bound DNA fraction dependence on the molecular ratio of TRβDBD:F2 TRE concentration is calculated by integrating the fluorescent signal from the wt-TRβDBD binding profile.

bind different kinds of TREs in either “head to tail” or “tail to tail” conformation depending on the half-site arrangements, both the symmetric DBD homodimer on F2TRE and the asymmetric dimer on DR4 will have to connect to the structurally symmetric LBD dimer.

### 2.3.5 Affinity and stoichiometry of the F2 TRE binding

To confirm the binding stoichiometry of the TRβDBD binding to the F2 TRE, we have used electrophoresis native gel shift assays and quantified the fraction of the bound DNA by integrating the signals from the fluorescent dye. To measure the stoichiometry, the amount of DNA used in the gel shift assay is kept constant in each well at a concentration at least 50 times the expected Kd for the binding between TRβDBD and F2 sequence. The protein concentrations in each well are increased stepwise to cover a molar ratio of protein:DNA from 0 to 4. As shown in Figure 2.5A, almost all the complex formed by TRβDBD and F2 TRE is present as a homodimeric super shift of the DNA band starting from very low concentration of proteins. The dimeric binding mode at very low protein:DNA ratio is indicative of cooperative DNA binding where binding of the second DBD subunit on the DNA is energetically favorable to binding of the first. Additionally, the quantitative binding profile of the gel, as shown in Figure 2.5G, establishes the stoichiometry of the binding is at 2:1 protein:DNA molar ratio and reinforces the cooperative binding of TRβDBD on the F2 TRE. As a negative control, we have scrambled the sequence of one half-site in the F2 TRE creating a single half-site TRE with the same base pair composition as the F2 TRE and then performed the same gel shift assay. The two gels have also been run simultaneously, so that the positions of bands are comparable. As expected, only the monomer-DNA complex band is revealed in

the gel without any band at the dimer-DNA complex position when TRβDBD binds to the scrambled F2 TRE.

To better quantify the DNA binding affinity of the TRβDBD, we have used a steady state fluorescence anisotropy assay. The F2 DNA used in the assay has a covalently coupled fluorescein probe at the 5' end of the forward strand. The resulting titration curve is shown in Figure 2.7A. After curve fitting by Hill coefficient equation, the Kd of TRβDBD binding to F2 TRE is calculated to 130±50nM with a Hill coefficient value of 1.5±0.2. The TRβDBD Kd value is less than the previously reported 350-500nM Kd of RXR DBD binding to DR1 or DR2 measured by the same method [24]. The lower Kd value implies that the TRβDBD binds with a slightly higher affinity than the RXR DBD when bind to the consensus DNA sequence. The increased binding affinity can be readily attained by the extra minor groove DNA interactions, which are allowed only by TR DBD with the long CTE helix. The calculated Hill coefficient value is also consistent with the observed cooperativity in the previous gel-shift assay.

One possible origin of the cooperative DNA binding is the inter-subunit contact of the DBDs further stabilizing the inter-subunit interaction once they become bound to the DNA. Our structure reinforces the expected result that the 6 base-pair spacing of the F2 TRE means that no inter-subunit contact can be achieved between the core DBD regions of the TRβDBD-F2 TRE complex,. The C-terminal tail of the DBD is the only region of the two subunits that could directly interact without dissociating the core DBD from the DNA. Unfortunately, due to crystallographic disorder in this part of the structure, the electron density at this part is too ambiguous to construct a definitive structure model (Figure 2.3B).

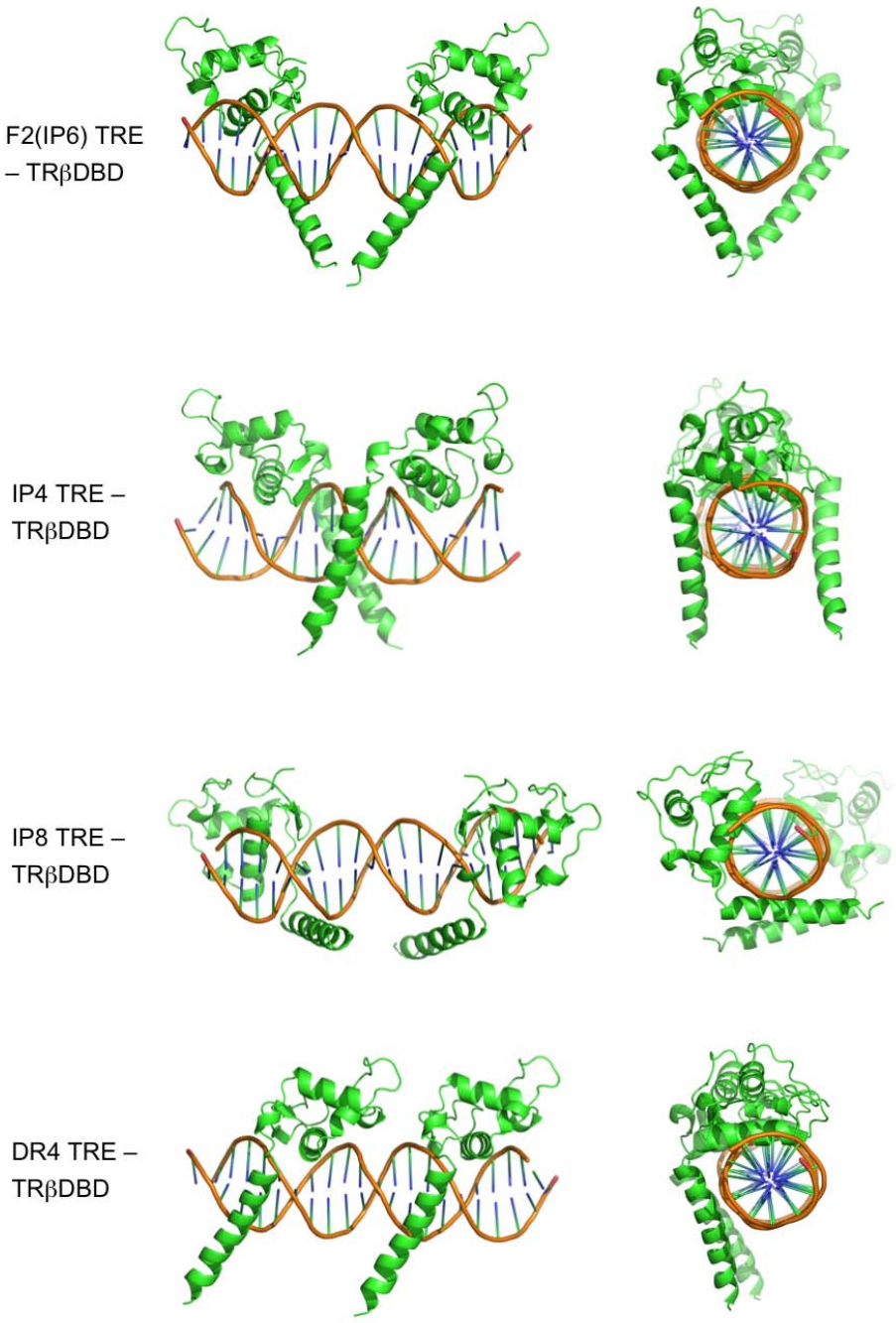



Figure 2.6: Structure models of TRβDBD with IP4, IP8 and DR4 TREs. The first one from the top is the crystal structure of TRβDBD and canonical F2 TRE with 6 base pair spacing. By rotating and translate one of the TRβDBD subunits along the DNA, we can make the models with 4 base pair spacing or 8 base pair spacing. The model on DR4 is created by overlay extra TRβDBD subunit with the original RXR DBD in the RXR-TR heterodimer structure (PDB ID: 2NLL). The side-view (left) and the top-view (right) can help us to visualize the possible orientations of the two CTE helices from the TRβDBD homodimer subunits.

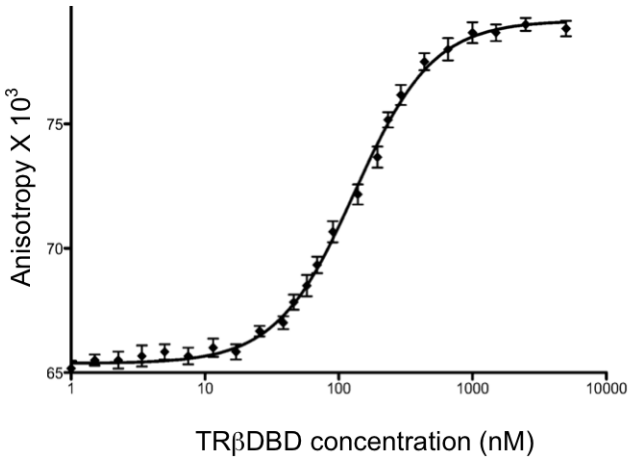
In order to evaluate whether the contacts at this tail region between two DBD monomers are responsible for the cooperative binding of TRβDBD on F2 TRE, we have created inverted palindromic TREs with different spacer length: 4bp (IP4) or 8bp (IP8) designed to alter or eliminate potential stabilizing contacts between the CTE tail regions that form when bound to DNA. As shown in Figure 2.5C, the gel shift assays of TRβDBD with IP4 and IP8 TRE both demonstrate the presence of monomer-DNA at low concentration of protein, implying some reduction in the inter-subunit cooperativity relative to IP6. For the IP4 TRE, we do not see a significant 2:1 protein:DNA preference until very high concentrations of protein while the majority of the complex formed with the IP8 DNA is still in the 2:1 protein:DNA form. The gel shift results suggest that the binding of TRβDBD with IP8 remains moderately cooperative while binding with IP4 is not cooperative, or the cooperativity is greatly reduced. The formation of TR monomers on the IP4 DNA appears to be similar to the formation of TR monomers on the DR4 TRE in the assay. This may indeed represent loss of cooperativity on the IP4 DNA because it is predicted that TR homodimer formation on DR4 TREs is more subdued because the second zinc module of the upstream TRβDBD would sterically clash with the first zinc module and T-box region of the downstream TRβDBD. The fact that this clash is replaced by salt bridges in the protein dimer interface in the previous structure of TR-RXR heterodimer-DR4 complex accounts for the observed directionality of the TR-RXR heterodimer on DR4 TRE [1].

In order to develop a structural explanation for the IP4 and IP8 cooperativity differences, computational models have been constructed for the complexes of TRβDBD dimer bound to IP4 and IP8 TREs, as shown in Figure 2.6. Shortening of the spacer to

4bp relocates the two DBD monomers on nearly opposite sides of the DNA helical axis. This orientation would completely restrict the ability of the two CTE tails to contact each other without introducing a bend into the helices to clear the DNA. On the other hand, elongation of the spacer to 8 base-pairs in the IP8 TRE moves the two monomers farther apart from one another, in which the two CTE helices might still be able to interact if we assume the exact orientation of the CTE helix is somewhat flexible. It is still possible therefore, that inter-subunit interactions account for the observed cooperativity in the binding of TRβDBD to IP8 TRE despite the increased spacing between the half-sites relative to IP6. The IP4 and IP8 structural models are thus consistent with the hypothesis that the CTE tails is important for the cooperative binding but we still cannot rule out the possibility that other factors could be contributing to the differences in cooperativity.

We have carried out two additional experiments to test whether interactions between the CTE tail regions account for the observed homodimer cooperativity on IP6 TREs. The sum of all these experiments have provided an intermediate set of results, and we suggest that inter-subunit CTE interactions must account for some, but not all, of the observed TR homodimer cooperativity when binding to the F2 TRE. In the first experiment, we found it curious that the CTE contains only a single hydrophobic patch within an otherwise extremely highly charged (mostly acidic) stretch of residues: a single leucine in the TRβDBD, which is replaced by a methionine and isoleucine in the TRαDBD. We suspected that a hydrophobic residue at this position might be able to form a hydrophobic contact between the two DBD subunits that would facilitate the CTE interaction. To test this hypothesis, we mutated leucine-98 to glutamate (L98E),

(A)




(B)

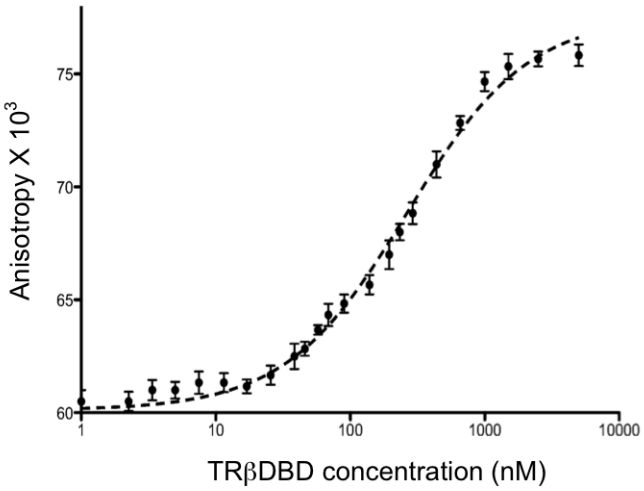



Figure 2.7: Fluorescence anisotropy measurements showing the binding of (A) the wild-type TRβDBD to F2 TRE; (B) the Leu99Glu mutant of TRβDBD to F2 TRE. The final F2 TRE concentration is 1nM in the binding mixture and the TRβDBD is titrated in from 0 to 5,000nM. The data is fit according to hill coefficient based formula.

reasoning that the highly acidic L98E CTEs would no longer favorably interact, and re-evaluated the cooperativity of binding to F2. The gel shift assay (data not shown) found that the L98E CTE mutant is still able to form 2:1 complexes on the F2 DNA, consistent with the expectation that the L98E mutation will not perturb any of the DBD-DNA interactions but not supporting a loss of CTE interaction.

Because it is difficult to quantify small changes in binding affinity or inter-subunit cooperativity from this experiment we have also performed a fluorescence anisotropy experiment, results shown in Figure 2.7B. The experiments confirm that while the mutant protein still binds F2 as a dimer, the L98E mutant has both a roughly two-fold reduction in the apparent affinity for the F2 DNA relative to the wild type DBD and a change in Hill coefficient from 1.5 to 1.1. Both of these results support the hypothesis that the inter-subunit CTE contact contributes to the cooperative binding. In a second set of experiments, a short TRβDBD construct was made where we deleted the final 8 residues from the c-terminal end of the original protein construct. This deletion was designed to completely abolish inter-subunit contacts by decreasing the length of the CTE while preserving all of the observed protein-DNA interactions. Somewhat contrary to our expectations, the deletion mutant TRβDBDΔ8, where the CTE is too short to be able to make inter-subunit contacts without disrupting its helical structure, still appears to exhibit inter-DBD subunit cooperativity when binding to the F2 TRE in the gel shift assay. Unfortunately, we are unable to observe a sufficient signal of the binding of TRβDBDΔ8 to F2 TRE using the fluorescence anisotropy assay to be able to accurately quantify the binding affinity (data not shown).

In conclusion, we suggest that the CTE tail interaction is an important component of the cooperative DBD binding but that it may not be the only factor responsible for this effect. Another possible effect that could be contributing to the cooperativity is that the binding of the first DBD subunit could operate on the DNA as a classic allosteric mediator that alters the affinity of the DBD binding at the second consensus site by acting at a distance [51]. One possible mechanism for an allosteric effect would be that the DNA could experience a subtle conformational or even entropic change induced by the first DBD binding, which would then enhance the affinity for the binding of the second DBD.

## 2.4 Discussion

The present structure of the TRβDBD homodimer bound to F2 DNA has reiterated the main features of the TRE half-site DNA recognition and binding observed in previous structures and has also highlighted additional structural implications of different TRE half-site arrangements. The major groove base contacts to the consensus DNA regions have been well preserved from the TR-RXR heterodimer structure, which are the determining factor of the specific DNA sequence recognition. We also find that the orientation and 6 base-pair inter-half-site spacing of the two hexameric half-sites in the F2 DNA sequence results in a symmetric arrangement of the DBD subunits on the F2 DNA that supports inter-subunit binding cooperativity to the DNA. Additional TR-DNA backbone contacts have also been identified in the minor groove that involve the TR CTE helix and the inter consensus-site spacer DNA. These backbone phosphate contacts with the spacer base pairs 5' to each half-site are dependent on the apparently flexible orientation of the TR CTE. An alternative orientation of the CTE is observed in one of the four TR subunits that eliminates the minor groove contacts. The interaction with the

spacer base-pairs is not only variable among TR homodimers, as the disordered CTE of RXR or the embedded CTE of PPAR each has an individual signature of less or more extensive interactions at the DNA minor groove.

Inter-subunit cooperativity is a signature property of many DNA regulatory complexes that has important functional implications. While both the native gel shift assay and FP experiments clearly indicate cooperative binding of two TRβDBD subunits to the F2 TRE, the mechanism that underlines this cooperativity is not completely revealed by the structure. Few nuclear receptors have high-resolution structures known for both monomer-DNA and free states in addition to the dimer-DNA complex so far. The CTE helix of the TRβDBD appears to contribute to some, but not all, of the binding cooperativity via its ability to form inter-subunit contacts at the c-terminal tail region. Varying the inter-half-site spacing can alter the cooperativity and shift the preference of monomer or dimer formation on the F2 TRE. At the same time, the elimination of any possible direct inter-subunit contacts by the deletion of a considerable portion of the CTE helix failed to completely abolish cooperative subunit binding. This leaves two possible explanations for the observed cooperativity: 1) non-bonded, i.e. electrostatic, interactions between the DBD subunits, or 2) DNA-mediated inter-subunit effects. One model for the latter is that the binding of the first subunit reduces the overall conformational entropy of the TRE DNA so that the binding of the second subunit pays less of an entropic cost and is more favorable. This hypothesis is difficult to test experimentally but will be the subject of future computational studies. More detailed analysis on subtle NRE DNA or NR conformational changes between different NR bound states may reveal factors

contributing to observed allosteric and cooperative binding effects, as was revealed in the subtle influence of the DNA on loops within bound glucocorticoid receptor DBDs [43].

The application of the florescence anisotropy has successfully measured the affinities of the cooperativity binding. The experiment setup and data analysis are quite straight forward and require as little as nano-grams of the DNA and protein. Compared to the protein and DNA amount used in crystallographic studies, it can be more extensively applied for the proteins that are difficult to express, such as the full-length TR or the TR-NCoRsh complex. As a result, the technique can also readily be employed to measure the binding of TR LBD and coregulator peptide, considering that the coregulator peptide is a smaller molecule bind on the surface of the TR LBD. However, due to the requirement on a significant signal change upon binding, the FA assays may not able to reflect the binding of the binding of two similar size molecules, as in the case we have tried to measure the TR DBD monomer binding with scramble F2 TRE. Moreover, the introduction of the fluorescent tag may alter the binding affinity itself by changing the chemical environment close to the binding sites, especially in the case that the binding is not occurred on the surface of the molecule, but at a more internal binding pocket. In this case, the binding of LBD ligands to TRs is an example of the internal binding, where the application of FA assay will be limited.

The relatively high degree flexibility of the T-box hinge region connecting the core TRβDBD and the CTE helix is illustrated by the alternative CTE conformation observed in one of four of the TRβDBD subunits. The repositioning of the CTE appears to be correlated with a conformational change in the T-box region of the DBD, such that the T-box region functions as a semi-flexible hinge for the CTE. Although the T-box in the

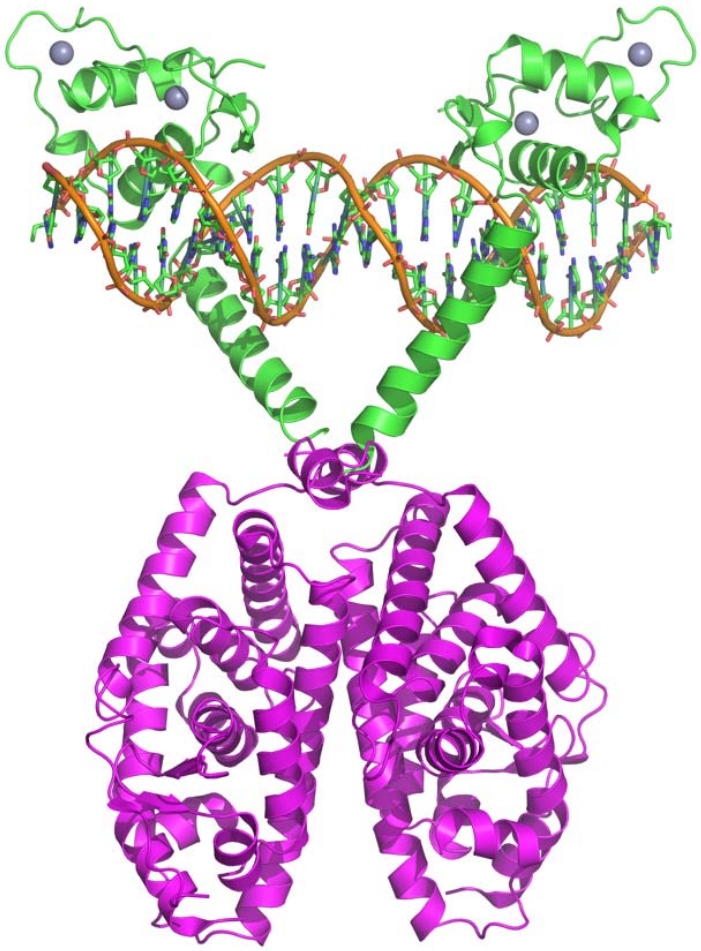


Figure 2.8: Model of the DBD-LBD TR homodimer on the F2 TRE. The TR DBD-F2 TRE complex is taken from the crystal structure. The TR LBD homodimer is created by overlay two copies of the previous reported TRβLBD structure (PDB ID: 1XZX) onto the PPAR-RXR LBD heterodimer reported in the intact PPAR-RXR and DR1 complex (PDB ID: 3DZU). And the c-terminal of the TR DBD and n-terminal hinge from the TR LBD are aligned together.

TRβDBD homodimer-F2 TRE complex is liberated from its role in the inter-subunit contact found in the TR-RXR heterodimer, the configuration of the T-box in the homodimer DBD subunit still adopts the same conformation observed in the TR-RXR heterodimer. There appears to be a balance between the favorable steric and electrostatic features of the T-box conformation in both the heterodimer and the homodimer subunit and the structurally linked forces stabilizing the conventional positioning of the CTE, including additional DNA backbone contacts with the 5' flanking DNA. The T-box flexibility may help to establish the discrimination of NRs by response element sequences, especially the identity of the 2 base pairs at the 5' flanking region of the hexameric half-site. As a result, the base preferences outside the consensus half-site region, as demonstrated in previous biochemical studies , may be determined by the linked energetics of the different T-box compositions and the distinct CTE conformations in different NRs.

To better understand the relevance of the F2-DBD homodimer structure within the context of full-length TR, we constructed a model of a full-length TR/DNA complex. We found that our structure of the DBD homodimer with F2 TRE places the first residues of the TR LBD in a position that is compatible with the formation of LBD homodimers. Evidence from dynamic light scattering and small angle x-ray experiments suggest that a symmetric homodimer is likely to form between the TR LBDs, as demonstrated by the homodimer or homotetramer crystal structure of RXRαLBD in the ligand-free state [52, 53]. Using the structures the TRβLBD monomer [54], the PPAR/RXR LBD dimer [55], and the DBD-LBD RXR/PPAR [56] heterodimer dimer as templates for the TR LBD dimer, we constructed a model of a plausible TR LBD-DBD homodimer bound to the F2

TRE, shown in Figure 2.8. While this is only one possible model for the TR LBD interface, we can predict from this arrangement that the formation of the LBD homodimer could be synergistic with the energetics of DBD homodimer formation on the F2 TRE. The expected favorable homodimer formation is consistent with the binding analysis reported recently that the TR LBD-DBD construct may have an even higher affinity than the isolated DBD when binding to F2 TRE, and this enhancement with the addition of LBD is absent when binding to the DR or Pal class TREs (57). Furthermore, the somewhat flexible CTE orientation we have observed would enable variable positioning of the C-terminus of the DBD and the N-terminus hinge of the LBD and this flexibility could facilitate alternative orientations of the TR LBD domains. While the T3 induced dissociation of the LBD homodimer can result in partial disassembly of full-length TR homodimer on the F2-TRE as demonstrated previously (29, 58), the ability to accommodate various N-terminal LBD positions may still allow LBD positional rearrangement without complete dissociation of the DBDs from the DNA. The elements of cooperative DNA binding of the TRβDBD homodimer coupled with the conformational flexibility of the CTE provide additional pieces of structural information to incorporate into different binding modes of full length TR or other NRs on different kinds of TREs.

## 2.5 Methods

### 2.5.1 Protein and DNA purification

The DNA binding fragment of the rTRβ and the deletion mutant were expressed in BL21 (Rossetta 2) E.*coli* cells as a SUMO-fusion protein including a His-tag for easy purification. The Rossetta2 E.*coli* strain was proved to effectively eliminate truncated

protein expression due to its additional tRNAs genes for rare codons located within the construct [36]. The construct of the TRβDBD deletion mutant was created by introducing a stop codon ahead of the c-terminus 8 amino acids in the wild type TRβDBD plasmid using the QuikChange site-directed mutagenesis kit (Stratagene). The cells were induced with 0.2mM IPTG for 20 hours at 18°C. After harvesting, the cells were resuspended and lysed in a buffer containing 20mM potassium phosphate, pH 7.5, 150 NaCl, 10% glycerol, 5mM dithiothreitol and freshly added 1mM PMSF. The bacterial lysates were centrifuged at 30,000rpm for 45 min and the supernatants were collected. The supernatants were subjected to purification first on a SP-Sepharose column (GE Healthcare) and subsequently on a Ni-NTA column (GE Healthcare). Afterwards, the SUMO tag was removed by a SUMO protease, ULPL1, while dialyzing into: 25mM Bis-Tris, pH 4.5, 100mM KCl, 5mM DTT and 10% glycerol overnight at 4˚C. Afterwards, the rTRβ DBD protein was purified to homogeneity on a second SP-Sepharose column. The concentration was estimated based on calculated extinction coefficient values. Synthetic oligonucleotides (IDT) with idealized F2 TRE sequence were purified using denaturing PAGE. The oligonucleotides were extracted, lyophilized, and then resuspended in buffer containing 20mM Tris 8.0, 100mM NaCl, to a final concentration of 2mM. The purified complementary strands were then mixed together at equal amount, heated up to 95°C and stepwise cooled down to 25°C over 4 hours.

### 2.5.2 Crystallization and data collection

The purified rTRβ DBD protein was concentrated to 2mM using 5,000 MWC Amicon Ultra concentrators, and then mixed with the duplex TRE DNA at a 2:1 molar ratio. The protein-DNA mixture was concentrated to 9mg/ml protein and crystals were

grown using the hanging drop method against a reservoir containing 12-13% PEG 3350, 200mM $NH_4Cl$, 5mM $MgCl_2$, 0.1M Bis-Tris, pH 6.0 and 10uM $ZnCl_2$, at 4°C within a week. Crystals were mounted and exchanged into a cryo-protected buffer containing the reservoir solution supplemented with 30% glycerol and flash frozen in liquid nitrogen. Diffraction data sets were collected on the frozen crystals at the Argonne National Laboratory using the LS-CAT 21-ID-G beamline.

### 2.5.3 Structure determination

The diffraction data were integrated and scaled using HKL-2000 [59]. The structure was solved by molecular replacement using the Phenix program suite (AutoMR) [60]. The search model was constructed from the TR portion of the former reported structure of the RXR-TR DR4 complex [1] with a DNA helix matching the F2 TRE sequence. The best solution placed two copies of the DNA and TRβ DBD homo-dimer complex in an asymmetric unit. However, an extra electron density with the shape indicating an α-helix was revealed outside the original model. It was then discovered that one c-terminal extended helix of the four TRβ DBD subunits in an asymmetric unit had an unusual orientation compared with the canonical position as in the search model. The alternative CTE helix conformation is introduced and stabilized by crystal packing forces between the chain A and chain F in neighboring asymmetric units. This rotated helix was then built in residue by residue using the Coot program according to the electron density map as a separate piece of the structural model [61]. The solved structure was then subjected to several cycles of refinements with rigid body, simulated annealing, non-crystallography symmetry (NCS) and individual sites methods using the phenix.refine program [62, 63].

TLS refinement was included in the later rounds of refinement cycles with the optimum TLS groups calculated by the online TLSMD server [64, 65]. Base on the calculation of the TLSMD server, a total of 12 TLS groups were set for each asymmetric units. And the boundaries for TLS groups were slightly modified manually in consideration of the actually biological relevant units such as the two complementary strands of the DNA or the core DBD globular region and the CTE region. Rfree was generated by setting aside 5% of the reflections through the refinement process. The program Coot was used for manual electron density map examination, residue positioning, rotamer refinements and validations [61].

### 2.5.4 Native gel-shift assay

Native gel-shift assay were performed to determine the binding cooperativity and stoichiometry. 0.6μg of the oligonucleotides were incubated with an increasing amount of the TRβ DBD protein in each reaction mix to cover the protein-DNA molar ration from 0:1 to 4:1. Binding reactions were carried out on ice for 2-3 hours in binding buffer containing 20mM Tris, pH8.0, 100mM NaCl, 1mM DTT, 10% glycerol, 5mM $MgCl_2$ and 0.1mg/ml BSA. After incubation, the reaction mixtures were analyzed on 7% non-denaturing acrylamide gels using 0.5× Tris borate/EDTA buffer at 4˚C. The gels were stained in 1× SYBR nucleic acid gel stain (Invitrogen) for 20 minutes and exposed under UV light. The band signal from the gel image was quantified by the ImageJ program[66]. The background was subtracted and then the intensity distributions of each lane were plotted. Afterwards, baselines were set manually and the peak intensity areas were integrated. The error bars in the intensity quantification chart came from different adjustments during the calculation process.

### 2.5.5 Fluorescence anisotropy assay

The affinities of TR DBDs and TREs were measured by titration assays using fluorescence anisotropy. Stock of F2 and scrambled F2 TRE at 20μM were made by annealing the 5'-fluorescein labeled oligonucleotides (Invitrogen) with its complementary sequence to form duplex DNA. The DNA was diluted to 2nM in binding buffer: 20mM Tris, pH=8.0, 0.2mg/ml BSA, 5% glycerol, 5mM $MgCl_2$ and 1mM DTT. Protein stocks were diluted in the same buffer to an initial concentration of 10μM, and then stepwise diluted to 2nM. 80ul of protein and DNA were mixed together in a 96-well plate and incubated 3 hours at 4˚C. The fluorescence anisotropy values were measured using a TECAN Safire plate reader at 30˚C using the wavelength setting for fluorescein. Parallel measurements were performed six to eight times and the average data were calculated and fit by the Prism software to determine the $K_D$ and Hill coefficient values. The equation used for the fitted is:

$$A_{obs} = A_i + (A_f - A_i)\{[(TR + n)/(k^n)]/[1 + (TR^n/k^n)]\}$$

where Aobs is the observed anisotropy at a certain TRβDBD concentration TR, and Ai and Af are the lower and upper asymptotic limits, respectively. N and k are the hill coefficient and dissociation constant, respectively. As the Kd calculated is more than 100 fold of the DNA concentration in the sample, the depletion effect of proteins is neglected in the fitting process.

*This work will be published in the August issue (2010) of Molecular Endocrinology [67].*

# BIBLIOGRAPHY


1. Rastinejad, F., et al., *Structural determinants of nuclear receptor assembly on DNA direct repeats.* Nature, 1995. 375(6528): p. 203-11.
2. Nagy, L. and J.W. Schwabe, *Mechanism of the nuclear receptor molecular switch.* Trends Biochem Sci, 2004. 29(6): p. 317-24.
3. Mangelsdorf, D.J., et al., *The nuclear receptor superfamily: the second decade.* Cell, 1995. 83(6): p. 835-9.
4. Mangelsdorf, D.J. and R.M. Evans, *The RXR heterodimers and orphan receptors.* Cell, 1995. 83(6): p. 841-50.
5. Lazar, M.A., T.J. Berrodin, and H.P. Harding, *Differential DNA binding by monomeric, homodimeric, and potentially heteromeric forms of the thyroid hormone receptor.* Mol Cell Biol, 1991. 11(10): p. 5005-15.
6. Ribeiro, R.C., et al., *Mechanisms of thyroid hormone action: insights from X-ray crystallographic and functional studies.* Recent Prog Horm Res, 1998. 53: p. 351-92; discussion 392-4.
7. Tsai, M.J. and B.W. O'Malley, *Molecular mechanisms of action of steroid/thyroid receptor superfamily members.* Annu Rev Biochem, 1994. 63: p. 451-86.
8. Jeyakumar, M., et al., *Quantification of ligand-regulated nuclear receptor corepressor and coactivator binding, key interactions determining ligand potency and efficacy for the thyroid hormone receptor.* Biochemistry, 2008. 47(28): p. 7465-76.
9. DeFranco, D.B., *Nuclear export: DNA-binding domains find a surprising partner.* Curr Biol, 2001. 11(24): p. R1036-7.

10. Tini, M., et al., *An everted repeat mediates retinoic acid induction of the gamma F-crystallin gene: evidence of a direct role for retinoids in lens development.* Genes Dev, 1993. 7(2): p. 295-307.
11. Umesono, K., et al., *Direct repeats as selective response elements for the thyroid hormone, retinoic acid, and vitamin D3 receptors.* Cell, 1991. 65(7): p. 1255-66.
12. Carlberg, C., et al., *Two nuclear signalling pathways for vitamin D.* Nature, 1993. 361(6413): p. 657-60.
13. Zamir, I., J. Zhang, and M.A. Lazar, *Stoichiometric and steric principles governing repression by nuclear hormone receptors.* Genes Dev, 1997. 11(7): p. 835-46.
14. IJpenberg, A., et al., *Polarity and specific sequence requirements of peroxisome proliferator-activated receptor (PPAR)/retinoid X receptor heterodimer binding to DNA. A functional analysis of the malic enzyme gene PPAR response element.* J Biol Chem, 1997. 272(32): p. 20108-17.
15. 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, 1994. 13(6): p. 1425-33.
16. Perlmann, T., et al., *Determinants for selective RAR and TR recognition of direct repeat HREs.* Genes Dev, 1993. 7(7B): p. 1411-22.
17. Mader, S., et al., *The patterns of binding of RAR, RXR and TR homo- and heterodimers to direct repeats are dictated by the binding specificites of the DNA binding domains.* EMBO J, 1993. 12(13): p. 5029-41.
18. Gronemeyer, H. and D. Moras, *Nuclear receptors. How to finger DNA.* Nature, 1995. 375(6528): p. 190-1.

19. Lee, M.S., et al., *Structure of the retinoid X receptor alpha DNA binding domain: a helix required for homodimeric DNA binding.* Science, 1993. 260(5111): p. 1117-21.
20. Zilliacus, J., et al., *Structural determinants of DNA-binding specificity by steroid receptors.* Mol Endocrinol, 1995. 9(4): p. 389-400.
21. Shaffer, P.L. and D.T. Gewirth, *Structural basis of VDR-DNA interactions on direct repeat response elements.* EMBO J, 2002. 21(9): p. 2242-52.
22. Sem, D.S., et al., *NMR spectroscopic studies of the DNA-binding domain of the monomer-binding nuclear orphan receptor, human estrogen related receptor-2. The carboxyl-terminal extension to the zinc-finger region is unstructured in the free form of the protein.* J Biol Chem, 1997. 272(29): p. 18038-43.
23. Gearhart, M.D., et al., *Monomeric complex of human orphan estrogen related receptor-2 with DNA: a pseudo-dimer interface mediates extended half-site recognition.* J Mol Biol, 2003. 327(4): p. 819-32.
24. Zhao, Q., et al., *Structural basis of RXR-DNA interactions.* J Mol Biol, 2000. 296(2): p. 509-20.
25. Yen, P.M., *Physiological and molecular basis of thyroid hormone action.* Physiol Rev, 2001. 81(3): p. 1097-142.
26. Williams, G.R., Brent, GA *Thyroid hormone response elements.* Molecular Endocrinology: Basic Concepts and Clinical Correlations, 1995: p. 26.
27. 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, 1993. 7(7B): p. 1423-35.

28. Darling, D.S., et al., *Different dimerization activities of alpha and beta thyroid hormone receptor isoforms.* J Biol Chem, 1993. 268(14): p. 10221-7.
29. Velasco, L.F., et al., *Thyroid hormone response element organization dictates the composition of active receptor.* J Biol Chem, 2007. 282(17): p. 12458-66.
30. Williams, G.R., et al., *Thyroid hormone receptor binds with unique properties to response elements that contain hexamer domains in an inverted palindrome arrangement.* Endocrinology, 1994. 134(4): p. 1888-96.
31. Piedrafita, F.J., et al., *Thyroid hormone receptor homodimers can function as ligand-sensitive repressors.* Mol Endocrinol, 1995. 9(5): p. 563-78.
32. Brent, G.A., et al., *Capacity for cooperative binding of thyroid hormone (T3) receptor dimers defines wild type T3 response elements.* Mol Endocrinol, 1992. 6(4): p. 502-14.
33. Ribeiro, R.C., et al., *Thyroid hormone alters in vitro DNA binding of monomers and dimers of thyroid hormone receptors.* Mol Endocrinol, 1992. 6(7): p. 1142-52.
34. Hollenberg, A.N., et al., *Function of nuclear co-repressor protein on thyroid hormone response elements is regulated by the receptor A/B domain.* J Biol Chem, 1996. 271(45): p. 28516-20.
35. Makowski, A., et al., *Determination of nuclear receptor corepressor interactions with the thyroid hormone receptor.* Mol Endocrinol, 2003. 17(2): p. 273-86.
36. Diallo, E.M., D.L. Thompson, and R.J. Koenig, *A method for efficient production of recombinant thyroid hormone receptors reveals that receptor homodimer-DNA binding is enhanced by the coactivator TIF2.* Protein Expr Purif, 2005. 40(2): p. 292-8.

37. Hartong, R., et al., *Delineation of three different thyroid hormone-response elements in promoter of rat sarcoplasmic reticulum Ca2+ATPase gene. Demonstration that retinoid X receptor binds 5' to thyroid hormone receptor in response element 1.* J Biol Chem, 1994. 269(17): p. 13021-9.

38. Williams, G.R., et al., *Oligomeric binding of T3 receptor is required for maximal T3 response.* J Biol Chem, 1991. 266(29): p. 19636-44.

39. Benham, C.J. and S.P. Mielke, *DNA mechanics.* Annu Rev Biomed Eng, 2005. 7: p. 21-53.

40. Khorasanizadeh, S. and F. Rastinejad, *Nuclear-receptor interactions on DNA-response elements.* Trends Biochem Sci, 2001. 26(6): p. 384-90.

41. Wilson, T.E., et al., *Participation of non-zinc finger residues in DNA binding by two nuclear orphan receptors.* Science, 1992. 256(5053): p. 107-10.

42. Rastinejad, F., et al., *Structure of the RXR-RAR DNA-binding complex on the retinoic acid response element DR1.* EMBO J, 2000. 19(5): p. 1045-54.

43. Meijsing, S.H., et al., *DNA binding site sequence directs glucocorticoid receptor structure and activity.* Science, 2009. 324(5925): p. 407-10.

44. Zhao, Q., et al., *Structural elements of an orphan nuclear receptor-DNA complex.* Mol Cell, 1998. 1(6): p. 849-61.

45. Harding, H.P., et al., *Transcriptional activation and repression by RORalpha, an orphan nuclear receptor required for cerebellar development.* Mol Endocrinol, 1997. 11(11): p. 1737-46.

46. Giguere, V., L.D. McBroom, and G. Flock, *Determinants of target gene specificity for ROR alpha 1: monomeric DNA binding by an orphan nuclear receptor.* Mol Cell Biol, 1995. 15(5): p. 2517-26.

47. Schwabe, J.W., et al., *The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements.* Cell, 1993. 75(3): p. 567-78.

48. Solomon, I.H., et al., *Crystal structure of the human LRH-1 DBD-DNA complex reveals Ftz-F1 domain positioning is required for receptor activity.* J Mol Biol, 2005. 354(5): p. 1091-102.

49. Chandra, V., Huang, P., Hamuro, Y., Raghuram, S., Wang, Y., Burris, T.P., Rastinejad, F., *Structural Organization of the Intact PPARgamma-RXRalpha Nuclear Receptor Complex on DNA.* Nature, 2008. 456(7413): p. 8.

50. Shaffer, P.L. and D.T. Gewirth, *Structural analysis of RXR-VDR interactions on DR3 DNA.* J Steroid Biochem Mol Biol, 2004. 89-90(1-5): p. 215-9.

51. Figueira, A.C., et al., *Human thyroid receptor forms tetramers in solution, which dissociate into dimers upon ligand binding.* Cell Biochem Biophys, 2006. 44(3): p. 453-62.

52. Figueira, A.C., et al., *Low-resolution structures of thyroid hormone receptor dimers and tetramers in solution.* Biochemistry, 2007. 46(5): p. 1273-83.

53. Bourguet, W., et al., *Crystal structure of the ligand-binding domain of the human nuclear receptor RXR-alpha.* Nature, 1995. 375(6530): p. 377-82.

54. Sandler, B., et al., *Thyroxine-thyroid hormone receptor interactions.* J Biol Chem, 2004. 279(53): p. 55801-8.

55. Connors, R.V., et al., *Identification of a PPARdelta agonist with partial agonistic activity on PPARgamma.* Bioorg Med Chem Lett, 2009. 19(13): p. 3550-4.

56. Chandra, V., et al., *Structure of the intact PPAR-gamma-RXR-alpha nuclear receptor complex on DNA.* Nature, 2008: p. 350-356.

57. Figueira, A.C., et al., *Recognition by the thyroid hormone receptor of canonical DNA response elements.* Biochemistry, 2010. 49(5): p. 893-904.

58. Yen, P.M., et al., *Triiodothyronine (T3) decreases binding to DNA by T3-receptor homodimers but not receptor-auxiliary protein heterodimers.* J Biol Chem, 1992. 267(6): p. 3565-8.

59. Otwinowski, Z. and W. Minor, *Processing of X-ray Diffraction Data Collected in Oscillation Mode Methods in Enzymology.* Macromolecular Crystallography, 1997. 276(part A): p. 20.

60. Adams, P.D., et al., *PHENIX: building new software for automated crystallographic structure determination.* Acta Crystallogr D Biol Crystallogr, 2002. 58(Pt 11): p. 1948-54.

61. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallographica Section D-Biological Crystallography, 2004. 60(Part 12 Sp. Iss.): p. 7.

62. Afonine, P.V., R.W. Grosse-Kunstleve, and P.D. Adams, *A robust bulk-solvent correction and anisotropic scaling procedure.* Acta Crystallogr D Biol Crystallogr, 2005. 61(Pt 7): p. 850-5.

63. Afonine, P.V., et al., *On the possibility of the observation of valence electron density for individual bonds in proteins in conventional difference maps.* Acta Crystallogr D Biol Crystallogr, 2004. 60(Pt 2): p. 260-74.

64. Painter, J., Merritt, E.A., *TLSMD web server for the generation of multi-group TLS models.* J. Appl. Cryst., 2006. 39: p. 3.

65. Painter, J. and E.A. Merritt, *Optimal description of a protein structure in terms of multiple groups undergoing TLS motion.* Acta Crystallogr D Biol Crystallogr, 2006. 62(Pt 4): p. 439-50.

66. Abramoff, M.D., Magelhaes, P.J., Ram, S.J., *Image Processing with ImageJ.* Biophotonics International, 2004. 11(7): p. 6.

67. Chen,Y. and M.A. Young, *Structure of a thyroid hormone receptor DNA binding domain homodimer on a inverted palindrome response element.* Molecular Endocrinology, 2010. (to be published)

# Chapter III

# Structural Characteristics of the Thyroid Hormone Receptor Ligand Binding Domain in the Ligand-free State

## 3.1 Abstract

The ligand binding domain (LBD) is a key player in ligand-dependent transactivation function of the thyroid hormone receptor. In the ligand binding domain, the dynamic properties of helix 12 of nuclear receptors are especially critical determinants of transcriptional activity of TR due to their direct involvement in coregulator recruiting. Unfortunately, current available structures of TR LBDs are all in the ligand-bound (*holo*) form with little variations been observed in the helix 12. Only two NR LBD structures in their ligand-free (*apo*) states have been determined so far, which belong to the α isoform of retinoid X receptor (RXR) and the γ isoform of peroxisome proliferator-activated receptor (PPAR). The helix 12 regions display two distinct conformations in the two available *apo* NR LBD structures. In order to evaluate which one of the *apo* NR LBD

structures presents the possible *apo* TR LBD conformation, and thus advance our understanding of the molecular mechanism of the functional switch of TR LBD from the ligand-free state to the ligand-bound state, we have modeled the *apo* TR LBD using the two related *apo* LBD structures of PPAR and RXR. Molecular dynamics (MD) simulations and free energy calculations have been applied on both homology models to determine their structural stabilities as well as their dynamic properties, which may contribute to the TR function. Our data support that the PPAR-based *apo* TR LBD model is the more reliable structure, in which the helix 12 adopts a relatively stable conformation mimicking the active TR LBD structure in the ligand-bound state.

## 3.2 Introduction

Thyroid hormones direct critical events in many physiological processes, such as regulating embryo development, metabolism, heart rhythm and cholesterol level, by binding to thyroid hormone receptors (TR), which in turn regulate gene transcription in the nucleus [1, 2]. The activity of TRs on their target genes results from three major events: recognition and binding of DNA response elements in the promoter region, ligand binding, and exclusive corepressor/coactivator interaction. In the absence of T3, TRs retain corepressor proteins and block target gene transcription. Upon T3 binding, TRs release the corepressors and recruit coactivator proteins to activate the gene transcription. While specific DNA recognition is achieved by the centrally located and highly conserved DNA binding domain, both ligand binding and ligand-dependent coregulator (including the corepressor and coactivator proteins) recruitment is directed by the ligand binding domain (LBD) at the C-terminal region of the TR protein. The LBDs share highly homologous sequences among various TR isoforms but less conservation among

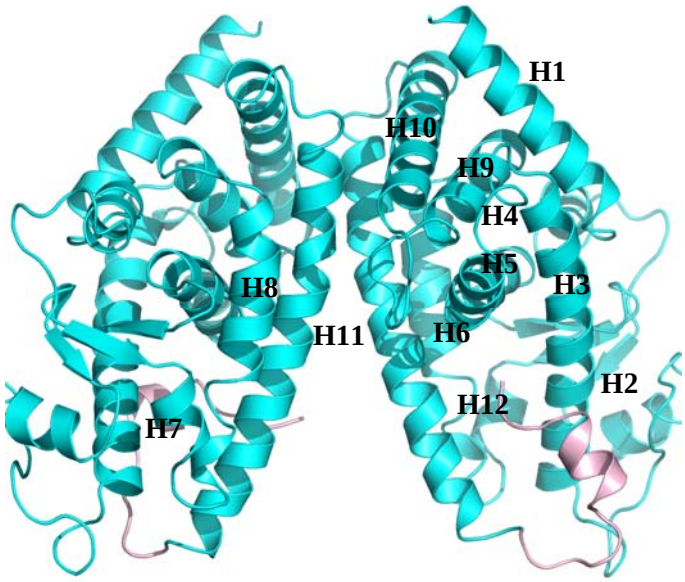
members of the nuclear receptor protein family, as compared to the DNA binding domains [3, 4].

Despite the less conserved amino acid sequences of NR LBDs, the overall architectures of the NR LBDs are well preserved within the NR protein family. Determination of crystal structures of *holo* TR LBD and other NR LBDs have greatly aided our understanding of molecular mechanisms underlying their function. The main body of the NR LBDs consists of 12 helices folding into a three layered, antiparallel α-helical sandwich, in which a central layer formed by helices 5, 9 and 10 is covered by two additional layers of helices and creates a ligand binding cavity in the middle. With ligand bound, NRs initiate the corepressor dissociation and coactivator association. This process is also accompanied by change of oligomerization states of NRs and affinity for a dimerization partner. Without ligand bound, RXRs tend to form tetramers in solution and lead to auto-repression of its activity. Upon ligand binding, dissociation of the homotetramers into homodimers or monomers has been observed in RXRs [9]. Similar ligand-induced effect on alteration of oligomerization states has also been observed for TRs in solution, based on evidence from experiments such as small angle x-ray scattering (SAXS) and dynamic light scattering (DLS) [14, 15]. However, even though the *apo* and *holo* structures of RXR LBDs have been determined, it is not fully understood how ligand binding interferes with oligomerization states of the LBDs.

TR LBDs not only can bind the ligands, but also play critical roles in dimerization, transcriptional activation and basal repression in the ligand-free states. Direct evidence of the TR LBD dimer assemblies has not yet been obtained from the crystallography-based structure studies. However, homodimers and heterodimers have been determined for the

other highly related proteins in the NR family, such as retinoid acid receptor (RAR), retinoid X receptor (RXR), peroxisome proliferator-activated receptor (PPAR) and vitamin D receptor (VDR) [5-8]. Among the NR LBD dimer structures, the homodimers of RXR and PPAR are in the *apo* form, while the RXR-heterodimers are in the *holo* form. Unlikely to be a result of mere coincidence, it has been proposed that the ligand binding can facilitate the heterodimer assemblies with RXR and hinder the homodimer assembly of NRs. As DNA response elements of NRs (NREs) mostly consist of two half-sites for NR binding, NRs often exert their transactivation functions as dimers. Therefore, the dimerization property within the NR LBDs is also associated to dimer assemblies on the DNA response elements and the transcription activation function of NRs [9]. All the non-steroid NR proteins, such as TR, RAR, VDR, and PPAR, can heterodimerize with the RXR and form activation complexes on the DNA response elements--especially the direct repeats type of response elements. In contrast, the steroid NR proteins, such as androgen receptor (AR), glucocorticoid receptor (GR), and estrogen receptor (ER), are more likely to form homodimers on palindromic NREs in a symmetric assembly [10, 11]. Therefore, the ligand-induced alteration of oligomerization states is probably distinct between the two subgroups of the NR protein family and the related conformational switch may also be different. Additionally, transcription activation and basal repression function of the TR LBD is correlated with interactions with coactivators and corepressors through the interface on the LBDs in a ligand-dependent manner [12, 13]. In summary, the LBD is the central piece mediating ligand binding, dimerization and coregulator recruitment function of TRs.

(A)




(B)

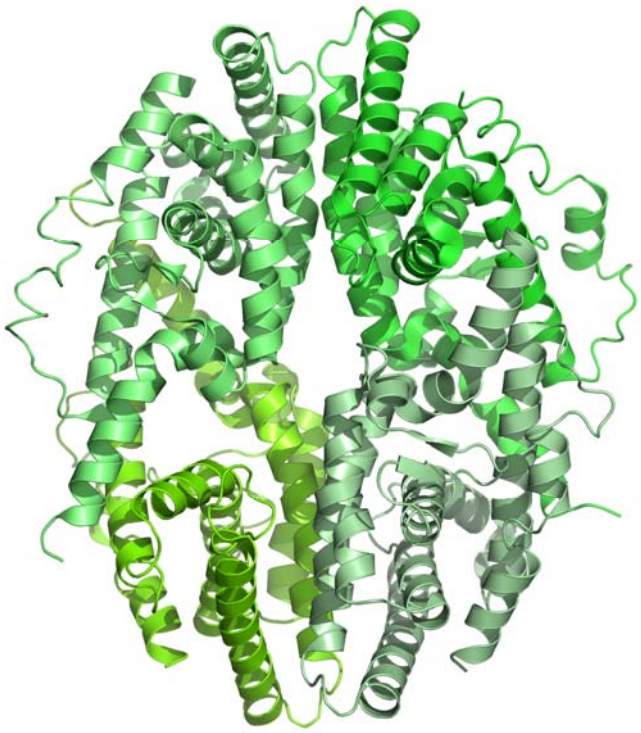


Figure 3.1: (A) The homodimeric structure of the *apo* PPAR LBD, the helices 1-12 are marked as H1-12 (pdb ID: 1PRG). The AF-2 region is colored in pink. (B) The tetrameric structure of the *apo* RXR LBD (pdb ID: 1G1U).

The C-terminal helix region of the NR LBDs is highlighted as the transactivation function 2 (AF-2) motif, with respect to the AF-1 region that has been defined in the N-terminal A/B domain, due to its critical role in the ligand-dependent transactivation function. This hypothesis is primarily based on the structural information from the ligand-free (*apo*) and ligand-bound (*holo*) RXR LBD, an NR protein with about 30% sequence identity in the LBD region as the TR. It has also been assumed that the TR LBD will undergo similar conformation transitions upon ligand binding [5, 9, 16]. In the RXR LBD, the AF-2 is the major region with the most significant structural switch from an extended "open" state to a folded back "closed" state upon binding of its ligand (retinoic acid). Alteration of the AF-2 conformation is directly linked to the coactivator or corepressor recruitment. Since the "closed" state AF-2 is required as a part of the coactivator binding interface; while the "open" state AF-2 can expose some inner hydrophobic residues accessible for the corepressor interactions [12]. However the functional correlated AF-2 structure switch upon ligand binding is not as obvious in another related protein: the PPAR LBDs. In the *apo* PPAR LBD crystal structures (Figure 3.1A), the AF-2 region still adopts a folded conformation resembling the "closed" state rather than the "open" state [6]. Without exposure of the hydrophobic residues buried by the AF-2 motif in the closed conformation the *apo* PPAR LBD can only incorporate the corepressors through an induced fit process by altering the AF-2 conformation upon corepressor binding [17]. Thus the preference of the corepressors over the coactivators in the absence of ligand cannot be directly explained by the *apo* PPAR LBD structure, which may be a reason that this *apo* structure of PPAR LBD has been less appreciated in the related NR LBD studies. Nevertheless, without the actual crystal structure available

for the *apo* TR LBD it is not completely justified to simply assume that TR is behaving the same as the related RXR (Figure 3.1).

Nevertheless, without the actual crystal structure available for the *apo* TR LBD, it is not completely justified to simply assume that TR is behaving the same as RXR. Based on the current knowledge, the AF-2 motif of *apo* TR LBD can adopt either the open state as in the *apo* RXR LBD structure or the closed state as in the *apo* PPAR LBD structure. Different conformations of AF-2 will lead to varied transcriptional competence as a result from their distinct accessibility for the coregulator proteins. If the AF-2 regions adopt similar conformation in a ligand-independent manner, the ligand-dependent adjustments of conformational dynamics within the AF-2 and other functional regions in the TR LBD are the primary factors differentially modulating the coregulator binding properties of the TR LBDs. Therefore, our focus in the homology modeling is to determine which conformation of the *apo* NR LBDs is more likely to be adopted by the *apo* TR LBD and what conformational dynamics may contribute to TR function.

## 3.3 Results

### 3.3.1 Overall structures

As shown in Figure 3.1, the *apo* LBD structures of RXR and PPAR still preserve the α-helix bundle structure as observed in the *holo* TR LBD structures [5, 6]. The loop region between helix 2 and helix 3, which is absent in the *holo* LBD crystal structures, becomes stable enough to be resolved in both *apo* LBD structures. Nevertheless, the most significant structural difference between *apo* LBDs of the RXRα and the PPARγ is located at the AF-2 helix region. The AF-2 helix of *apo* RXR LBD adopts an extended conformation, which is distinct from the *holo* RXR LBD, by dramatically increasing the

angle between helix 12 and helix 11. The same extended conformation or the open state of the AF-2 helix is also observed in the homotetramer structure of *apo* RXR LBDs, in which the AF-2 contributes to dimerization and auto-repression of *apo* RXR LBDs by a domain swapping mechanism [9]. On the other hand, the AF-2 motifs of *apo* PPAR LBDs adopt two slightly varied conformations revealed by the two monomer subunits in the crystal structure. One conformation is almost identical to the closed AF-2 conformation in the *holo* PPAR LBD structure, while the other stands for a partially closed conformation with a folded back but less compacted helix 12 via a slightly loose conformation adopted by the linker between helix 11 and helix 12 [6]. As has been shown in several NR LBD and coregulator peptide complex structures, the open state of the AF-2 helix prepares the *apo* LBD for the corepressor binding, which requires the exposure of several hydrophobic amino acid residues in helix 3 and helix 5, which are otherwise buried by the closed AF-2 helix in *holo* NR LBDs [17-19]. Based on related structural studies, the AF-2 helix of the *apo* TR LBD may exist in either the open or the closed conformation with varied transcriptional competence resulting from their distinct accessibility for the coregulator proteins. Our focus in computational homology modeling is to determine which one of the resolved *apo* NR LBD structures better represents the *apo* TR LBD conformation and what conformation is relatively more stable to be adopted by the TR AF-2 helix.

Two homology models of *apo* TRβLBD have been constructed using either of the crystal structures of the *apo* RXR LBD or the *apo* PPAR LBD with the AF-2 in the completely closed state as the template. The T3 bound *holo* TRβLBD crystal structure

(A)

```
thyroid         ----------EELQKSIGHKPEPTDEEWELIKTVTEAHVATNAQG----SHWKQKRKFL 45
peroxisome      ---------ESADLRALAKHLYDSYIKSFPLTKAKARAILTGKTTDKSPFVIYDMNSLMM 51
retinoid        MGSSHHHHHHSSGLVPRGSHKREAVQEERQRGKDRNENEVESTSSANEDMPVERILEAEL 60
                          *     *  :.  :.    *    .  :  .:                :

thyroid         PED-IGQAPIVNAPEGGK-VDLEAFSHFTKIITPAITRVVDFAKKLPMFCELPCEDQIIL 103
peroxisome      GEDKIKFKHITPLQEQSKEVAIRIFQGCQFRSVEAVQEITEYAKSIPGFVNLDLNDQVTL 111
retinoid        AVEPKTETYVEANMGLNPSSPNDPVTNICQAADKQLFTLVEWAKRIPHFSELPLDDQVIL 120
                  :      :       .        .           :  :.::** :* * :*  :**: *

thyroid         LKGCCMEIMSLRAAVRYDPESETLTLNGEMAVTRGQLK--NGGLGVVSDAIFDLGMSLSS 161
peroxisome      LKYGVHEIIYTMLASLMNKDG-VLISEGQGFMTREFLKSLRKPFGDFMEPKFEFAVKFNA 170
retinoid        LRAGWNELLIASFSHRSIAVKDGILLATGLHVHRNSAHS-AGVGAIFDRVLTELVSKMRD 179
                *:    *::    :          :       : *   :       . .     ::  .:

thyroid         FNLDDTEVALLQAVLLMSSDRPGLACVERIEKYQDSFLLAFEHYINYRKHHVTHFWPKLL 221
peroxisome      LELDDSDLAIFIAVIILSGDRPGLLNVKPIEDIQDNLLQALELQLKLNHPESSQLFAKLL 230
retinoid        MQMDKTELGCLRAIVLFNPDSKGLSNPAEVEALREKVYASLEAYCKHKYPEQPGRFAKLL 239
                ::*.:::. : *::::. *  **      :*  ::..  ::*   : .  . .  :.***

thyroid         MKVTDLRMIGACHASRFLHMKVECPTELFPPLFLEVFED----- 260
peroxisome      QKMTDLRQIVTEHVQLLQVIKKTETDMSLHPLLQEIYKDL---- 270
retinoid        LRLPALRSIGLKCLEHLFFFKLIG-DTPIDTFLMEMLEAPHQMT 282
                 ::. ** *      . :  :*       : .:: *: :
```

(B)

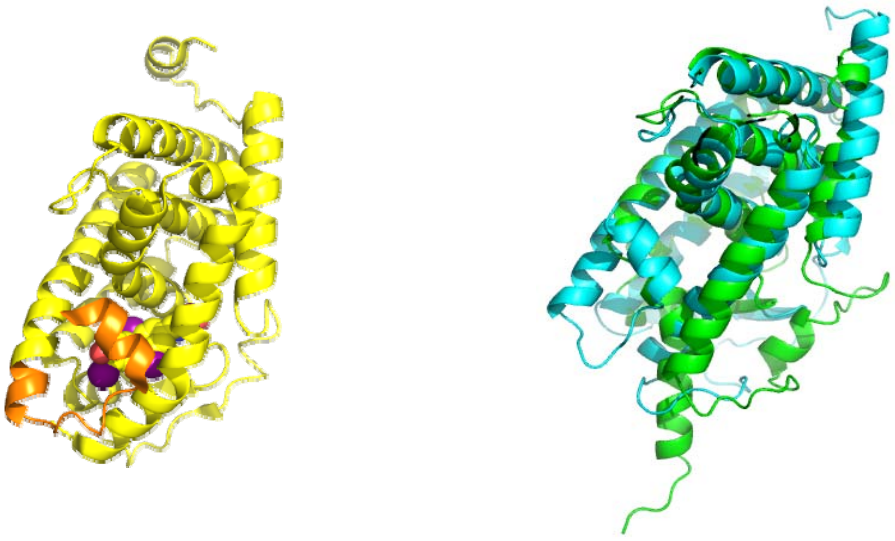


Figure 3.2: (A) Amino acid sequence alignment of the LBDs of TR, RXR and PPAR. The identical amino acids shared by all three proteins are marked by "*"; the homologous amino acids with similar properties such as charge or hydrophobicity are marked by ":"; and the most distinct residues are marked by "." below the sequences. (B) The left panel is the crystal structure of holo TRβLBD with T3 bound (PDB ID: 1XZX), in which the T3 ligand is shown as spheres and the C-terminal region with varied conformations in the *apo* LBDs is colored in orange. The right panel is the overlay of the RXR based- and PPAR based- homology model of the apo TR LBD, colored in green and cyan, respectively.

(PDB code 1XZX) is used as a reference in the structure-based sequence alignment for the modeling input. Because different AF2 conformations have been revealed in the *apo* RXR and *apo* PPAR crystal structures, the outcome homology models of the TR LBD also adopt distinct conformation in the AF2 region. In addition, the two refined model structures of TRβLBD are also distinct in two other regions besides the AF-2: helix 3 orientations at the N-terminal part and the loop region connecting helix 2 and 3 (Figure 3.2 B). The RXR-based model preserves the open form of the AF-2 helix and the loop region with a much longer helix 2 in the middle of the loop region; while the PPAR-based model has the complete closed form of the AF-2 helix and a short helix 2 followed by the loop region interacting with the beta-turn, which is closer to the T3-bound TRβLBD structure. On the other hand in the RXR-based model, the extended form of AF-2 severs between the beta-turn and the helix 3. Consequently, the overall conformation of the loop region around helix 2 is also more significantly altered in the RXR-based model, compared to the *holo* TR LBD structure. also adopt distinct conformations in the AF2 region. In addition, the two model structures of *apo* TRβLBD are also different in two regions besides the AF-2: the N-terminal region of helix 3 and the loop region connecting helix 2 and 3. The RXR-based model incorporates the open form of the AF-2 helix and a loop region with a relatively longer helix 2 in the middle; while the PPAR-based model has the complete closed form of the AF-2 helix and a shorter helix 2 followed by a loop region interacting with the beta-turn. Overall, the PPAR-based model share more similarity compared to the T3-bound TRβLBD structure. On the other hand, in the RXR based model, the extended form of AF2 severs between the beta-turn and the helix 3 and alters the packing in the helix bundle of TR LBD.

Consequently, conformation of the loop region flanking helix 2 is also more significantly altered in the RXR-based model, compared to the T3-bound TR LBD structure.

### 3.3.2 Stability analysis of the *apo* LBD models

Both the RXR-based and the PPAR-based *apo* TR LBD homology models are submitted to MD simulation to evaluate their stabilities. Two-dimensional root mean square deviation (rmsd) matrices, which represent the rmsd between structures at one time-spot and any other time-spot during the simulation, have been calculated for both MD trajectories. The matrices are shown in contour maps as in Figure 3.2. The contour color represents the different levels of rmsd, while the *x, y* axis represent the time points. The diagonal values of the maps are rmsds between the structure snapshots at the same time, which are therefore constant at zero. And if the line along the *x* axis is expanded according to the color-coded rmsd, it will be the same as the conventional one-dimensional rmsd curve with reference to the starting structure along the simulation trajectory, whereas other lines parallel to the *x* or *y* axis represent different rmsds calculated with reference to the structure at the time point of the corresponding *x* or *y* value.

In theory, the relatively stable structure ensembles, corresponding to the local minimum states (if not the global minimum) of the free energy landscape in the conformation space, should be able to maintain a low level of rmsd value through certain time durations. These stable states would be visualized by darker (purple) square regions in the 2-dimensional rmsd contour map, and reappearance of a certain state will present symmetrically as a dark square in the two triangles divided by the central zero-line crossway. Therefore, the much darker regions along the diagonal line of the 2-

dimensional rmsd map of the PPAR-based *apo* TRβLBD structure model (Figure 3.2B) suggests that this model is likely to be in a more stable state with a low degree of structural deviations during the MD. On the other hand, the RXR-based model (Figure 3.2A) shows a relatively high rmsd level all the way through the whole course of simulation, which indicates an evolving or a relatively flexible conformation.

The free energy estimations also support the notion that the PPAR-based *apo* TR LBD is the more favored model. As shown in Figure 3.2C, the RXR model at its starting point shows a very high free energy level, and during the simulation the free energy is gradually minimized while the structure is evolving to an energetically more favored state. On the other hand, the free energy of the PPAR-based model stays relatively on the same level as the starting structure during the MD, and its ability to maintain the free energy during the MD is compelling evidence supporting stability and reliability of this homology model. After about 28ns simulation, the free energy level of the RXR-based model drops to roughly the same level as the PPAR-based model. However, the RXR-based model probably still represents a higher free energy state, considering the fact that the amino acid sequence of the RXR-based model lacks 14 residues in the N-terminus of the PPAR-based model where the structure is not visible in the *apo* RXR LBD. Extra residues within a system will lead to an increase of the total free energy since the bonded terms of internal energies are always positive. Therefore, the shorter amino acid sequence in the RXR-based model directly results in a reduction of the internal energies from bonds, angles and dihedrals within the N-terminus residues, which is about a 300~500 kCal decrease of the free energy compared to the PPAR-based model. Due to the distinct internal energies that will be introduced by different amino acid sequences; the free

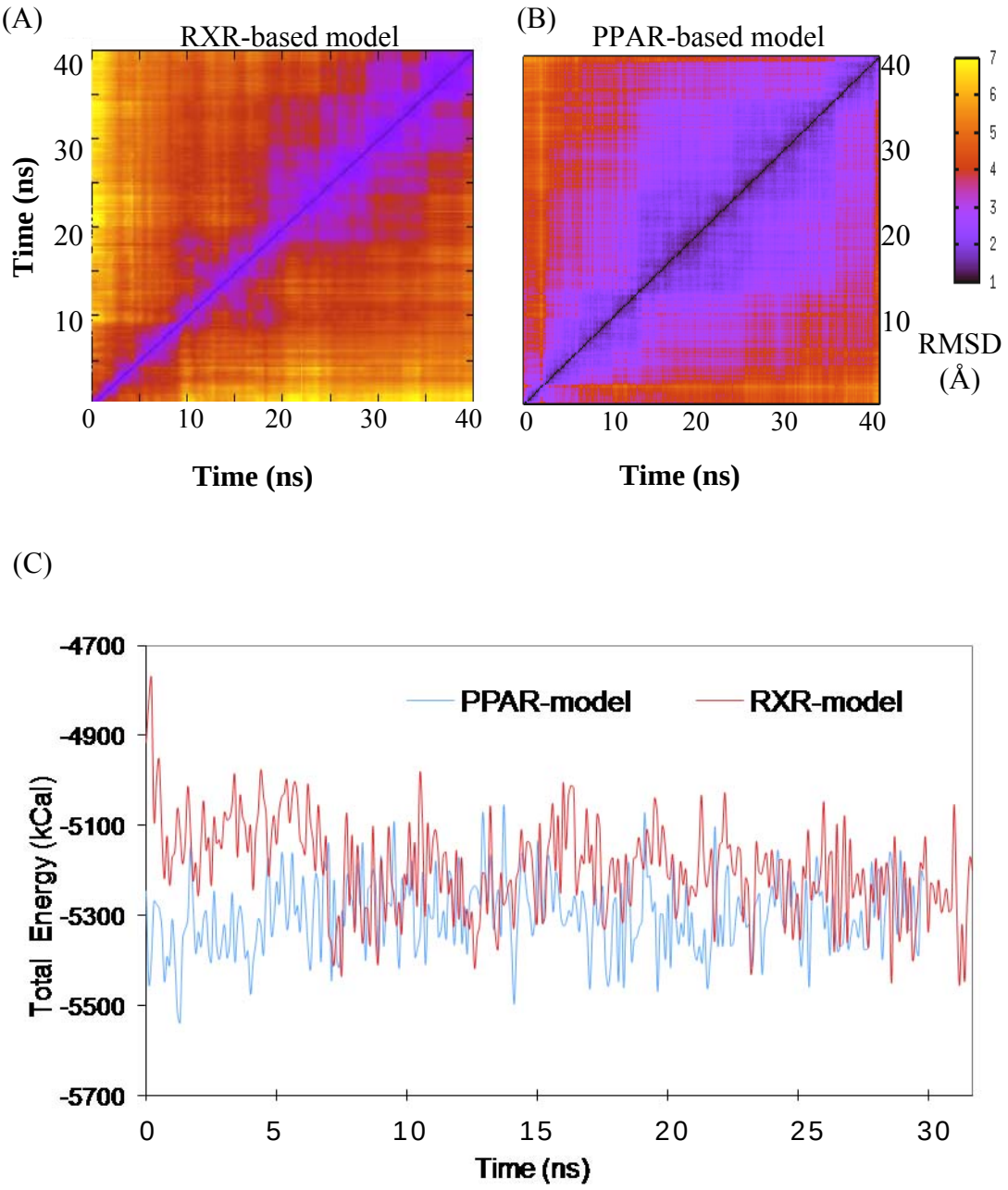



Figure 3.3: (A) and (B) are the two dimensional RMSD map of the RXR-based and PPAR-based *apo* TR LBD models. X, y axis units are in nanoseconds. Color contour level unit is in Angstrom representing the pair-wise RMSD values in each time spots during the MD simulation. (C) The total free energy of the two apo TR LBD models along MD trajectory time. The energy calculation included the gas phase energy components of the system and the solvation free energy using GBSA method.

energy comparison is not credible between two completely different proteins. In this case, however, since the two models share exactly the same sequence from the starting residue in the RXR-base model, the free energy comparison is still informative in evaluating the conformation states as long as the effect from the missing residues is taken into account. As a result, even though the free energies of the two models appear the same, the electrostatic, van der Waals, and solvation parts of the free energies put the PPAR-based model in the more favorable conformation stage. Not to mention in the starting model, the free energy of the PPAR-based model is considerably lower than the RXR-based model which proves that the PPAR-based model is a more stable conformation for *apo* TR LBDs (Figure 3.3C).

### 3.3.3 Clustering of *apo* TR LBD model structures

In order to get a better view of the possible conformation that has been described by the MD simulation, we have clustered all the structures along the MD trajectory into five groups using their rmsd values as a distance measure. Therefore, each cluster group incorporates similar structures with low rmsd values with each other in the group, while between different groups the structure variances are more significant with high rmsd values. From the representative structure from each group, we can directly visualize the major kinds of conformations that have emerged during the MD. Additionally, their relative stability can be quantified by the percentage of occurrence in each cluster group, since the higher percentage of occurrence of the structures in a cluster is likely to be correlated with the more predominant or energetically preferred conformations.

From the overlaid representative structures of the clusters in the RXR-based and PPAR-based *apo* TR LBD models, the variations from the starting models are revealed as

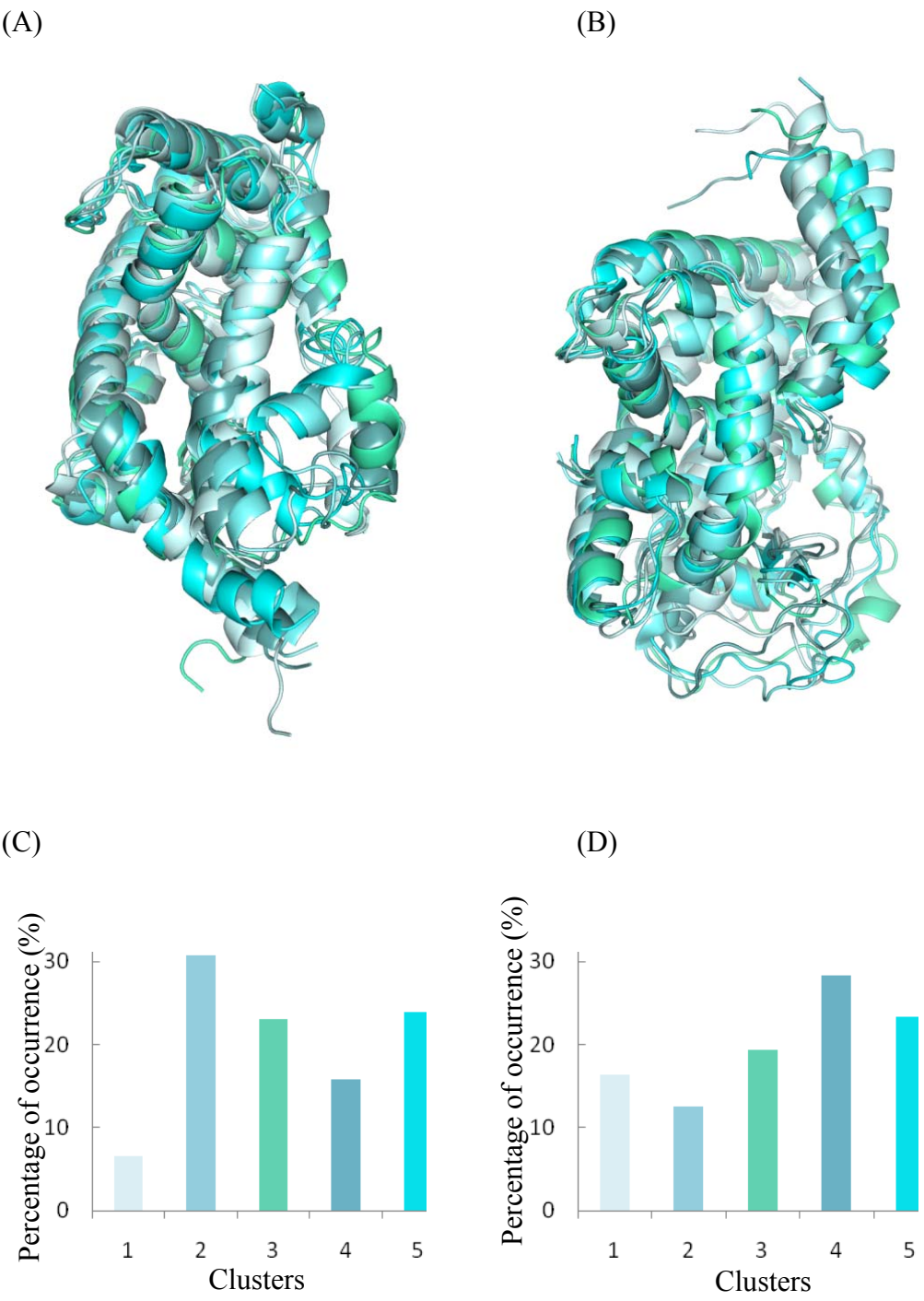



Figure 3.4: (A) and (B) are the representative structures from the five groups generated by clustering of the trajectory structures in the RXR-based and PPAR-based *apo* TR models, respectively. (C) and (D) are diagrams showing the percentage of occurrence of each structure group. The bars are colored according to the coloring scheme used in the structure view above.

in Figure 3.3 A and B. In general, the RXR-based model demonstrates more variation with respect to the starting model. The AF2 helix is significantly bent, compared to the extended conformation in the beginning structure. As shown by the clustering percentage, the straight AF2 conformation represents about only 6.6% of all the trajectory structures, while the bent AF2 is much more prevalent through the later stages of the MD simulations. Meanwhile, the helix 2 and the N-terminal portion of the helix 3 have also displayed highly variable positions. On the contrary, the PPAR-based model displays less dramatic conformation variations between the representative structures in the different clusters (Figure 3.4 B). As shown in Figure 3.4 D, the occurrence percentages of the clusters are closer to average: between 12% and 28%. In the representative structures, both the loop regions flanking helix 2 and helix 3 as well as the AF2 region are preserving their overall orientations and positions in the PPAR-based model structures. These observations are also in agreement with the previous rmsd and free energy calculations, proving the PPAR-based *apo* TR LBD model maintains its structural stability much better than the RXR-based model (Figure 3.4).

### 3.3.4 Regional fluctuations and dynamics

The mobilities of each residue in the *apo* TR LBD models are assessed by their atomic fluctuations in the form of B-factors calculated for each MD trajectory, as shown in Figure 3.5. The loop region between helix 2 and helix 3, the N-terminal part of helix 3, the dimerization interface around the linker between helix 10 and helix 11, and the AF2 region have all displayed much higher mobility in the RXR-based model than those in the PPAR-based model. On the other hand, the β-turn region and the neighboring helix 7 in the PPAR-based model are more flexible than in the RXR-based model. As has been

studied by previous steered MD simulations, the ligand dissociation pathway is most likely via rearrangements in the loop region and the β-turn close to it in the 3-dimensional structure [20]. Thus the varied mobilities observed in the β-turn and the loop between helix 2 and helix 3 may convey different ligand-binding properties, as increased flexibility around the β-turn probably can facilitate entry of the ligand T3. The linker region between helix 11 and helix 12 also displays more mobility in the PPAR-based model than the RXR-based model, even though the helix 12 in the PPAR-based model appears to be more stable.

The regional rmsd plots, as shown in Figure 3.6 A and B, of the *apo* TR LBD model simulations have further demonstrated the diverse pattern of the structural variations in the three functionally related regions. In the PPAR-based model, helix 12 in the *apo* TR LBD model deviates less from the starting point. Helix 2 in the middle of the loop region has the highest deviations from the starting position, which suggests that this region may be either highly flexible or still has not reached the stabilized conformation yet. In the RXR-based model, helix 12 has not yet been stabilized nor has it reached its ideal conformation, compared to the other two regions, with their rmsd curve levels up after about 18ns in the simulation. This result also indicates that the helix 12 conformation, as in the *apo* RXR LBD, is probably not a stable state for the *apo* TR LBD. In contrast to the common assumptions that the AF-2 region containing helix 12 is the most dynamic region within TR LBD in the absence of the ligand T3, our results predict there are other regions, such as the loop connecting helix 2 and helix 3, and the β-turn region, harboring dynamic potentials and may contribute to ligand-induced structural and functional switches in the TR LBDs.

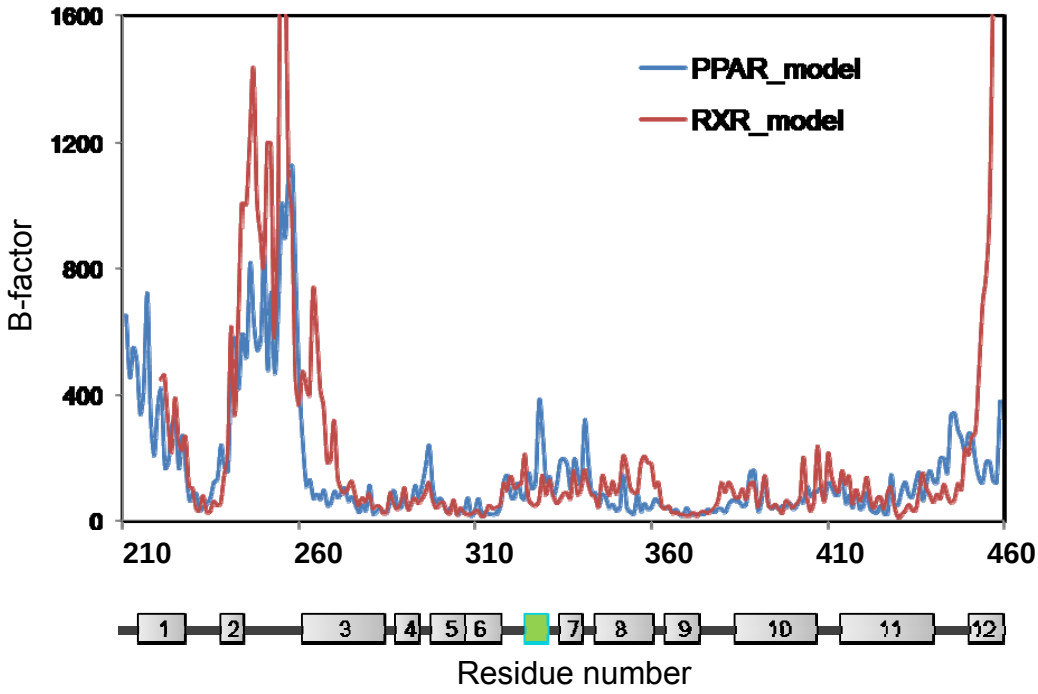
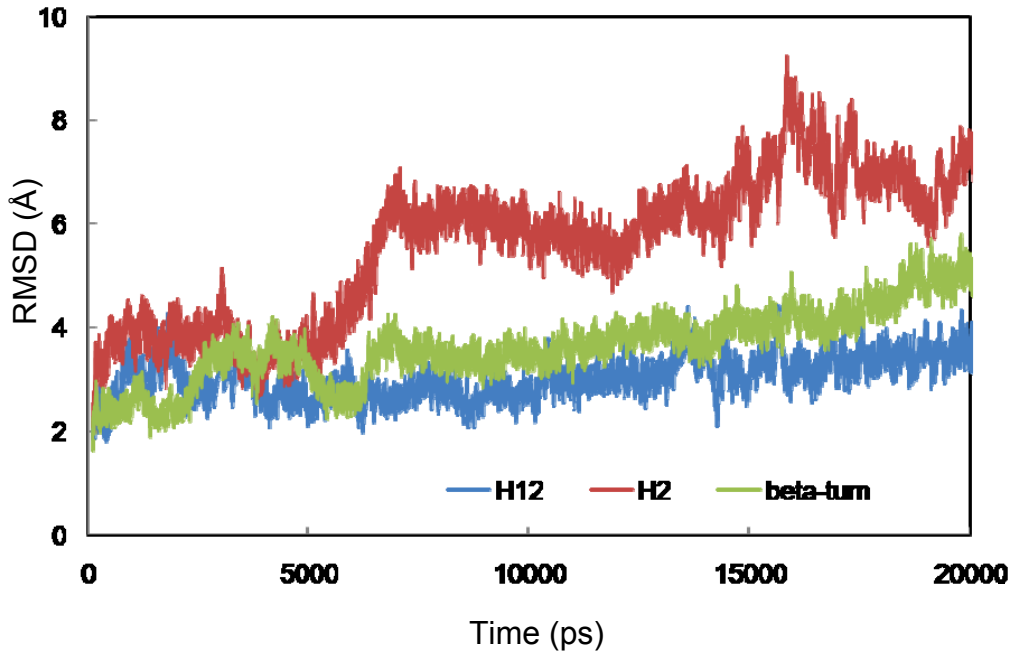



Figure 3.5: Calculated backbone atomic fluctuations in the form of B-factor for each amino acid residue in the RXR-based and PPAR-based model, colored in red and blue, respectively. The diagram below the B-factor plot shows the correlated positions of the twelve α-helices (as gray rectangles) and the β turn (as a green rectangle) in the TR LBD.

(A)




(B)

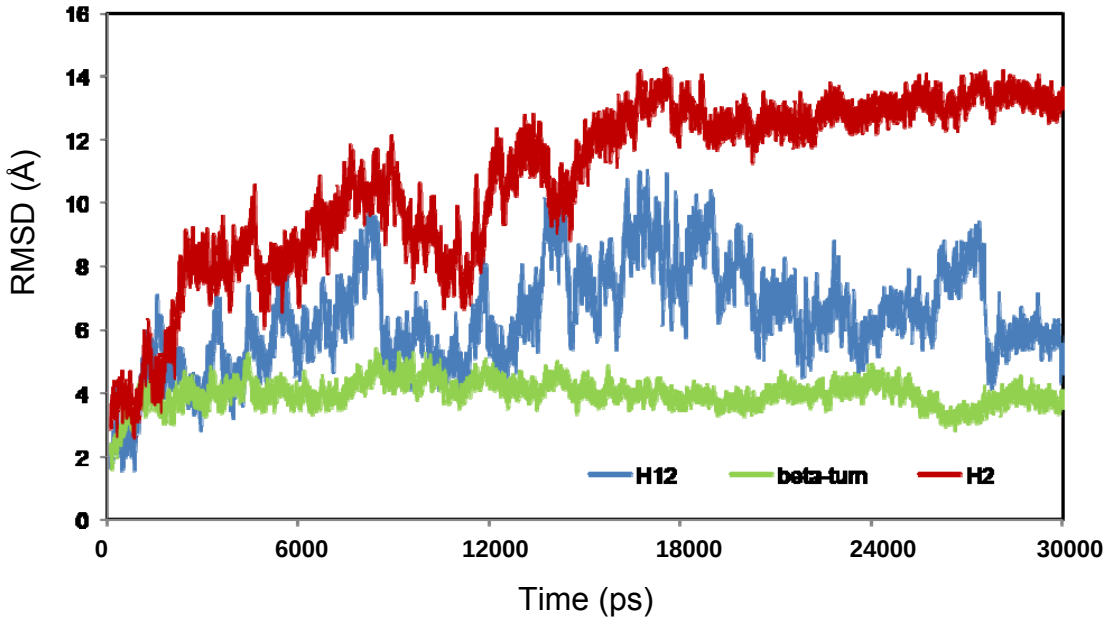



Figure 3.6: The rmsd calculated for three functional-related regions in the PPAR-based (A) and RXR-based (B) TR LBD models, respectively: the helix12 region (colored in blue), helix2 region within the loop connecting helix 1 and helix3 (colored in red), and the β-turn region between helix6 and 7 (colored in green).

## 3.4 Discussion

The homology models of *apo* TR LBD help us to better understand the yet-to-be-resolved structure of the TR LBD in the ligand-free state. Previous hypotheses preferred the structure of the *apo* RXR LBD as a model for the TR LBD in order to explain the functional aspects that have been observed in coregulator recruitment and transcription activation by the TRs. The "open" conformation of the AF-2 helix that has been present in the *apo* RXR LBD structure can directly explain the constitutive preference of corepressor binding since the hydrophobic part of the interface is readily exposed if the *apo* TR LBD adopts such a conformation. The "closed" conformation as observed in the *holo* TR LBD crystal structures is more prepared for the coactivator binding but not the corepressor binding. However, the crystal structure of the *apo* PPAR LBD reveals the AF-2 conformation analogous to the closed state, as in the *holo* TR LBD. If the *apo* TR LBD also adopts similar conformation as the *apo* PPAR, the corepressor cannot be engaged without an alteration of the AF-2 conformation upon binding. Thus, this structural model cannot directly explain the basal repression function that has been reported in the *apo* TRs. Nevertheless, a model fitting the easy explanation of the functional observations is not necessarily the actual right one representing the *apo* TR LBD structure.

In our study, we have modeled the *apo* TR LBD structure using either *apo* RXR LBD or *apo* PPAR LBD as the template and have tested their validities using MD simulation. From the analysis of various dynamics characters presented in the two models, the PPAR-based *apo* TR LBD is probably the more stable and energetically favored conformation, and the closed conformation of the AF-2 region in the *apo* TR LBDs also

appears to be much more stable than the open conformation, even in the absence of ligand. In this case, the binding of corepressors to the *apo* TR LBD will be an induced fit process, which comprises the adjustment of the AF-2 conformation and exposure of the hydrophobic interface only in the presence of the corepressors. Although the closed conformation of AF-2 in the PPAR-based *apo* TR LBD model is close to that being observed in the *holo* TR LBD structure, the *holo* TR preferentially recruits the coactivator proteins instead of the corepressor proteins. The distinct coregulator preferences of the TR LBD in the *apo* form and the *holo* form may be attributed to the change of structural dynamics rather than just the static conformations. In addition, regions other than the AF-2 motif may also contribute to the specified coregulator recruitments through alterations in their structural dynamic properties in ligand-free and ligand-bound states.

The regional dynamics data have also indicated that in addition to the AF-2 region, other parts of the TR LBD are very flexible and may also have the potential of structural or dynamic rearrangements in a ligand-dependent manner, similar to the results from the previous SMD study demonstrating that the ligand dissociation is likely to encompass regions other than AF-2 [20]. The ligand-dependent coregulator recruitment may also rely on the structure dynamics in the other regions besides AF-2 in a direct or allosteric way. This hypothesis is also supported by certain mutant TRs found in some patients with resistance to thyroid hormone (RTH) syndrome. Specifically, the mutant TRs with either or both of the mutations A234T and R243Q have been reported with impaired corepressor release. Rather than directly affecting the AF-2 conformation, these two mutations are located within the helix 2 and the loop connecting helix 2 and helix 3 and decrease the stability within the N-terminal region of the TR LBD [25]. Therefore,

mutation interference on the corepressor release is probably conveyed through an allosteric mechanism, which is still yet to be fully resolved. Furthermore, the crystal structures revealed for the two mutant TR LBDs fail to display any significant conformation change, and part of the crucial loop region is also unresolved in the structures. Therefore, MD simulation may provide a perfect way to study the structure as well as conformational dynamics.

## 3.5 Methods

### 3.5.1 Deriving *apo* TR homology models

Homology models of the *apo* form TRβLBD were developed using the program Modeller v8.2 [21]. The crystal structures of the RXRαLBD and PPARγLBD in their *apo* form (PDB codes 1LBD, 3PRG, respectively) were used as the template to build the two homology models of the *apo* TRβLBD [5, 6]. A sequence alignment of TR/RXR or TR/PPAR was first performed through ClustalW [22], followed by a structure-based sequence alignment by superimposing the *holo* TR LBD crystal structure (PDB code 1XZX) onto the RXR or PPAR template structure performed in PyMol. Both sequence-based and structure-based alignments were examined and tested as the alignment inputs for the Modeller program. With each group input of structural template and alignment files, 5 homology structures were generated and evaluated by the discrete optimized protein energy (DOPE) scoring function in Modeller. The final model with the lowest global DOPE score was subjected to energy minimization and molecular dynamics simulations.

### 3.5.2 Molecular dynamics simulation

The MD simulations of the TR LBD models were conducted using Amber 9. Preparation of the system was done in primarily three steps. First, the TR LBD molecule was solvated by adding explicit water molecules (TIP3P), creating a periodic octahedron box with a minimum distance of 10.0 Å from the TR LBD to the box edge. Then K+ ions were added to neutralize the system. Afterward, K+ and Cl- ions were added to the system to achieve the concentration 0.15M, mimicking the physiological ionic level. The solvated system was energy minimized to remove any side chain clashes and then heated to 303 K in 5ps and equilibrated by 45 ps at a temperature of 298 K. The solvent density was adjusted in the process at 1atm pressure, while harmonic positional restraints on the protein residues were stepwise reduced from 50 kCal/mol to 2.5 kCal/mol. Afterwards, MD simulations were performed at the canonical constant temperature and constant volume. Bonds involving hydrogen atoms were constrained to their equilibrium length by the SHAKE algorithm. The electrostatic energy was calculated by the particle-mesh Ewald (PME) method with the default setting in Amber 9.

### 3.5.3 Free energy calculation

The rmsds of the protein backbone with reference to the starting model and atomic fluctuation in the form of b-factors for each residue were calculated using the ptraj program of AmberTools. The conformation clusters were generated by dividing the structures in the MD trajectory into groups based on pair-wise similarity distance measured by the rmsd values, using a complete linkage algorithm.

Free energy calculation is performed using the MM_PBSA module in Amber 9. The calculated free energy included the gas phase energy components and the solvation free energy contribution depicted as the following equation:

$$Etot = Eelec + Evdw + Eintra + Eelec,desolv + Enp,desolv$$

Eelec and Evdw are the electrostatic and van der Waals energies, respectively. Eintra is the internal energy of the molecule, including the covalent bond, angle and torsion energies. Eelec,desolv and Enp,desolv are the electrostatic and non-polar desolvation energies, respectively. The Eelec,desolv energies are evaluated using a modified version of the generalized Born (GB) method. The non-polar desolvation energy is assumed to be proportional to the solvent-accessible surface area of the molecule with the constant of 0.0072kcal/(mol $Å^2$) [23].

# Bibliography


1. Glass, C.K. and J.M. Holloway, *Regulation of gene expression by the thyroid hormone receptor.* Biochim Biophys Acta, 1990. **1032**(2-3): p. 157-76.
2. Oppenheimer, J.H., Sammuels, H. H., *Molecuar Basis of Thyroid Hormone Action.* Academic Press, New York, 1983.
3. Mangelsdorf, D.J., et al., *The nuclear receptor superfamily: the second decade.* Cell, 1995. **83**(6): p. 835-9.
4. Tsai, M.J. and B.W. O'Malley, *Molecular mechanisms of action of steroid/thyroid receptor superfamily members.* Annu Rev Biochem, 1994. **63**: p. 451-86.
5. Bourguet, W., et al., *Crystal structure of the ligand-binding domain of the human nuclear receptor RXR-alpha.* Nature, 1995. **375**(6530): p. 377-82.
6. Uppenberg, J., et al., *Crystal structure of the ligand binding domain of the human nuclear receptor PPARgamma.* J Biol Chem, 1998. **273**(47): p. 31108-12.
7. Chandra, V., et al., *Structure of the intact PPAR-gamma-RXR-alpha nuclear receptor complex on DNA.* Nature, 2008: p. 350-356.
8. Pogenberg, V., et al., *Characterization of the interaction between retinoic acid receptor/retinoid X receptor (RAR/RXR) heterodimers and transcriptional coactivators through structural and fluorescence anisotropy studies.* J Biol Chem, 2005. **280**(2): p. 1625-33.
9. Gampe, R.T., Jr., et al., *Structural basis for autorepression of retinoid X receptor by tetramer formation and the AF-2 helix.* Genes Dev, 2000. **14**(17): p. 2229-41.

10. Schwabe, J.W., et al., *The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements.* Cell, 1993. **75**(3): p. 567-78.
11. Zilliacus, J., et al., *Structural determinants of DNA-binding specificity by steroid receptors.* Mol Endocrinol, 1995. **9**(4): p. 389-400.
12. Nagy, L., et al., *Mechanism of corepressor binding and release from nuclear hormone receptors.* Genes Dev, 1999. **13**(24): p. 3209-16.
13. Webb, P., et al., *The nuclear receptor corepressor (N-CoR) contains three isoleucine motifs (I/LXXII) that serve as receptor interaction domains (IDs).* Mol Endocrinol, 2000. **14**(12): p. 1976-85.
14. Figueira, A.C., et al., *Human thyroid receptor forms tetramers in solution, which dissociate into dimers upon ligand binding.* Cell Biochem Biophys, 2006. **44**(3): p. 453-62.
15. Figueira, A.C., et al., *Low-resolution structures of thyroid hormone receptor dimers and tetramers in solution.* Biochemistry, 2007. **46**(5): p. 1273-83.
16. Ribeiro, R.C., et al., *Mechanisms of thyroid hormone action: insights from X-ray crystallographic and functional studies.* Recent Prog Horm Res, 1998. **53**: p. 351-92; discussion 392-4.
17. Xu, H.E., et al., *Structural basis for antagonist-mediated recruitment of nuclear co-repressors by PPARalpha.* Nature, 2002. **415**(6873): p. 813-7.
18. Vanhooke, J.L., et al., *Molecular structure of the rat vitamin D receptor ligand binding domain complexed with 2-carbon-substituted vitamin D3 hormone analogues*

*and a LXXLL-containing coactivator peptide.* Biochemistry, 2004. **43**(14): p. 4101-10.

19. Darimont, B.D., et al., *Structure and specificity of nuclear receptor-coactivator interactions.* Genes Dev, 1998. **12**(21): p. 3343-56.
20. Martinez, L., et al., *Molecular dynamics simulations of ligand dissociation from thyroid hormone receptors: evidence of the likeliest escape pathway and its implications for the design of novel ligands.* J Med Chem, 2006. **49**(1): p. 23-6.
21. Fiser, A. and A. Sali, *Modeller: generation and refinement of homology-based protein structure models.* Methods Enzymol, 2003. **374**: p. 461-91.
22. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0.* Bioinformatics, 2007. **23**(21): p. 2947-8.
23. Gohlke, H., C. Kiel, and D.A. Case, *Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes.* J Mol Biol, 2003. **330**(4): p. 891-913.
24. Diallo, E.M., D.L. Thompson, and R.J. Koenig, *A method for efficient production of recombinant thyroid hormone receptors reveals that receptor homodimer-DNA binding is enhanced by the coactivator TIF2.* Protein Expr Purif, 2005. **40**(2): p. 292-8.
25. Huber, B.R., et al., *Thyroid Hormone Receptor-β Mutations Conferring Hormone Resistance and Reduced Corepressor Release Exhibit Decreased Stability in the NTerminal Ligand-Binding Domain.* Mol Endocrinol, 2003. **17**(1): p. 107-116.

# Chapter IV

# Corepressor Recruitment by the Thyroid Hormone Receptor Ligand Binding Domain

## 4.1 Abstract

Thyroid hormone receptors (TRs) recruit nuclear corepressor proteins in the absence of ligands to assemble the transcription repression complex on target genes. The two major nuclear corepressors, the nuclear receptor corepressor (NCoR) and the silencing mediator of retinoic acid and thyroid hormone receptors (SMRT), are related in their domain organizations as well as in their amino acid sequences in their functional domains. Despite the similarity in the two corepressors, TR is found to preferentially recruit NCoR as its functional partner. Previous studies have established the regions that directly bind NRs, namely the nuclear receptor interaction domains (IDs), within the corepressor proteins. The underlying mechanisms contributing to the nuclear receptor specificities are still not completely clear. We have applied computation homology modeling and molecular dynamics (MD) simulation to study the corepressor interaction with the TR ligand binding domain (LBD). All the homology models of TR LBD with

different corepressor ID peptides are stable enough to stay bound during the MD simulation. From the data in MD, we predict that the amino acids flanking the conserved "IxxII" (CoRNR) box are as important as the ones inside the box for directing the nuclear receptor specificity and the dynamics in binding. Specifically, data from the MD studies have also demonstrated that the unique ID domain present solely in NCoR but not in SMRT is crucial in determining TR preference, which is consistent with some of the previous experimental results. In addition, during the simulation, it has also been observed that incorporation of the T3 ligand into the TR-corepressor complex will alter the mobilities at several regions in TR LBDs and cause the corepressor peptide to drift away from the TR LBD as well.

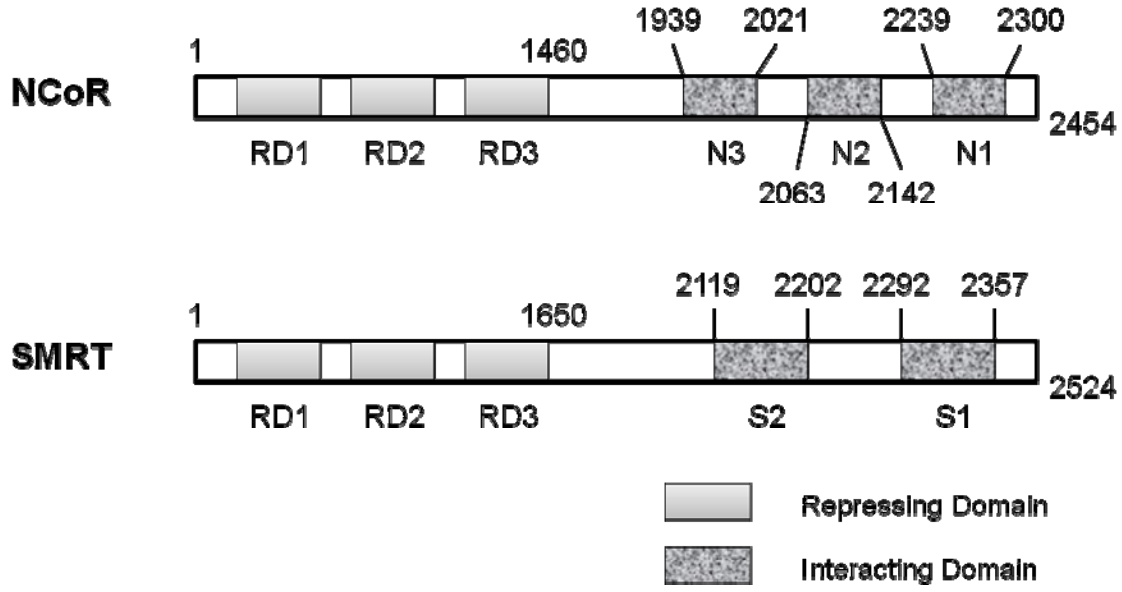
## 4.2 Introduction

Thyroid hormone receptor (TR) and retinoid acid receptor (RAR), two members in the nuclear hormone receptor protein family, are found to repress gene transcription in the absence of ligands through recruitment of corepressor proteins [1-3]. It has been established that both RARs and TRs function mostly via formation of heterodimeric complexes with the 9-*cis* retinoid X receptors (RXRs) [4, 5]. In vivo, the TR/RXR heterodimer has been shown to bind DNA in the context of chromatin/nucleosome assembly. The association of the thyroid hormone with the chromatin-bound TR/RXR heterodimer leads to the disruption of local chromatin structures in a transcription-independent process. Thus, chromatin structure has substantial links to the TR regulation of gene expression in vivo: the TR/RXR heterodimer recognizes the response element within the chromatin, TR/RXR makes use of the chromatin assembly to repress transcription more efficiently, and TR/RXR directs the disruption of the local chromatin

structure in response to thyroid hormones [6]. On the other hand, similar corepressor binding to homodimers of the nuclear receptors was also found mostly through in vitro assays and the presence of the corepressors helps to stabilize the homodimer assembly on the direct repeat type of DNA response elements in addition to their chromatin modification abilities [7, 8]. Importantly, an oncogenic mutant of TR, v-erbA acts as a constitutive repressor and, when coexpressed with the receptor, blocks transcription activation by thyroid hormones. The result of this inhibition is a loss of the hormone responsiveness and the constraint on the hormone-induced cell differentiation and proliferation. Mutations within v-erbA have been found to further stabilize a homodimeric assembly and prohibit the ligand-induced corepressor release [35]. Therefore, the oncogenic activity of v-erbA is directly linked to the transcriptional repression function of TRs [9, 10]. In addition, deletion or a transcription frame-shift mutation of the AF-2 domain at the C-terminus of RAR or TR generates a dominant negative mutant as a potent transcriptional repressor insensitive to ligand binding, and this type of mutation has been shown to cause serious defects in differentiation and development [11-14]. In conclusion, transcriptional repression by unliganded nuclear receptors appears to play an important role in regulating cell growth and differentiation conducted through modification of the chromatin structure.

Two major nuclear repressor proteins associated with the TR and RAR function that have been identified are NCoR (nuclear corepressor protein) and SMRT (silent mediator for retinoid and thyroid hormone receptors) [2, 3]. NCoR and SMRT are two related proteins of molecular weight of about 270kDa, with similar domain organizations

A)




B)

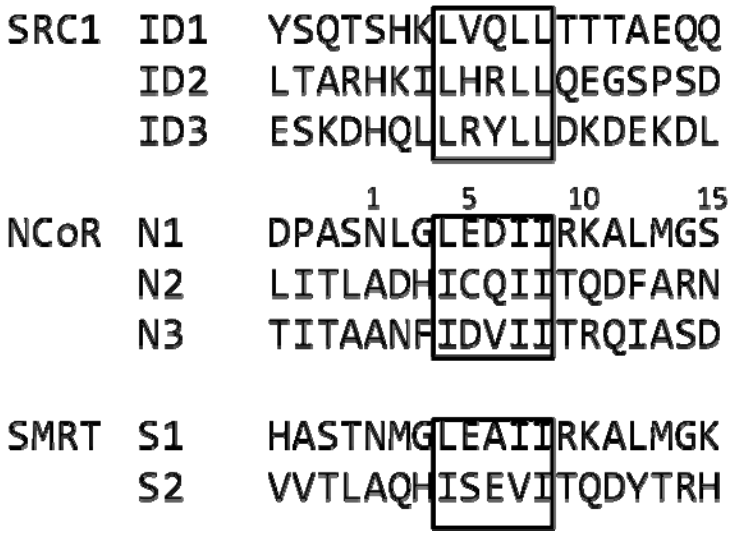



Figure 4.1: A) Schematic diagram of the domain organization in the nuclear corepressor protein NCoR and SMRT. B) The amino acid sequence around the NR boxes of the coactivator SRC1 and the CoRNR boxes in the corepressors. The box regions are marked by squares. The numbers above the NCoR sequences starts from the n-terminal residue that has been included in the homology modeling.

and abilities to repress transcription by directly or indirectly modifying the chromatin structure. Using a yeast genetic system, the cDNAs encoding the nuclear proteins that specifically interact with the ligand-binding domain of human retinoid X receptor-alpha (RXRα) were isolated for the first time. With activities dependent on nuclear receptor binding and as yet unidentified specific small-molecule ligands or on activation by other processes simultaneously, nuclear corepressors were identified as binding partners for the nuclear receptors while the repression complex is bound to DNA response elements [1]. Subsequently, the two major proteins and their natural splicing variants were both identified within the human as important mediators of hormone actions [2, 16, 17]. Experimental results from functional assays, such as that a fusion protein containing the GAL4 DNA-binding domain and SMRT functions as a repressor of the GAL4-dependent reporter, have established their role in the ligand-independent transcription repression [3].

Sequence and function studies of the wild-type and natural splicing variants have demonstrated the presence of three or four amino terminal domains in corepressors, which repress transcription in a cooperative as well as mechanistically distinct fashion [16]. Functional analysis reveals that SMRT contain at least three distinct repression domains in the N-terminal region, and the corresponding regions in NCoR are also responsible for repressing the basal transcription in the ligand-free state with abilities to directly or indirectly modify the chromatin structures. With evidence from in vivo and in vitro experiments such as the mammalian two-hybrid assay and far-western analysis, the repression domains in the SMRT and the NCoR demonstrate abilities to interact with Sin3A/B, which in turn associates with histone deacetylases (HDACs), to form the transcription repression complex [18, 19].

Two domains and three domains in the C-terminal region of SMRT and NCoR, respectively, are found to interact with the nuclear receptors. The two earlier discovered nuclear receptor interaction domains (IDs) in NCoR, named N1 and N2, are highly related to the two IDs found in SMRT, named S1 and S2. On the other hand, the later identified third ID (N3) in NCoR is not present in SMRT and specifically binds TRs but not other NRs [20, 21]. Inside each of the corepressors, there is a CoRNR box with a conserved "I/LxxI/VI" sequence, which is quite similar to the NR box "LxxLL" found in the coactivator proteins. Studies have shown that the CoRNR box is essential for NR interactions and that short CoRNR peptides are effective in blocking the function of corepressors in vitro and in vivo [22]. While the overall domain organizations as well as sequences of IDs from the SMRT and the NCoR show a great deal of similarity, they do target specific nuclear receptors during the transcription repression process (Figure 4.1 and Figure 4.7). It has been proposed that regulation of the transcription repression process of NRs relies on the selective combination of the CoRNR box and a certain NR LBD. In addition, for the function of TR, NCoR is the preferred corepressor partner that is normally recruits to the repression complex [21, 23]. The orphan receptor RevErb can recruit NCoR too, but RevErb only required N2 instead of N3 in the process [23]. In contrast, SMRT prefers to bind RAR preferably through the S2 domain [21, 23]. The RXR, as the common heterodimerization partner for the non-steroid NRs, can bind the NCoR or SMRT but preferably thorough the N1 or S1 domain [24, 25]. The separate NR preferences may also help to stabilize the NR homodimer or heterodimer assembly on various DNA response elements.

Both NCoR and SMRT are important targets for cell signaling pathways, which influence their expression levels, subcellular localization, and association with other proteins. Recently, the biological importance of these proteins has been revealed by studies of genetically engineered mice and patients with diseases such as acute promyelocytic leukemia (APL) and resistance to thyroid hormone (RTH) [26]. APL is often caused by fusion of the promyelocytic leukemia protein to RARα. Treatment of APL patients with an RAR ligand results in complete remission [27]. However, patients often generate RA-resistance due to mutations in the RAR moiety that prevent corepressor dissociation [28]. Thus the NR repression locking effect of CoRNR peptides can be an alternative method of APL treatment.

Our focus in the study of corepressor and TR interactions is to understand the structural features determining the binding specificity and the coregulator switching mechanisms directed by ligand binding. Using an optimized co-expression protocol, we are able to demonstrate the binding between the NCoR and the TR LBD. In order to further study the conformational changes and dynamic characteristics in the corepressor and TR complexes, we use homology modeling and MD techniques to study the 3-dimentional structures of the CoRNR peptides and the TRβLBD complexes based on the only resolved crystal structure of PPAR-SMRT S1 complex. In addition, the ligand binding effects have also been analyzed through the modeling and simulation of the T3-bound TR LBD-corepressor complex and comparing the dynamics properties to the T3-free TR LBD-corepressor complex.

## 4.3 Results

### 4.3.1 Corepressor binding to TR LBD

In order verify the binding of the corepressor with TR, we have co-expressed and co-purified the NCoR IDs with the TR LBD in *E. coli*. The amino acid sequence of the corepressor protein NCoR or of SMRT predicts they are all without defined secondary structure except for the short helical CoRNR box. Because both corepressors possess such a long amino acid sequence, ~2500aa, which is about 270kD in molecular weight, the lack of a stable secondary structure is likely to result in a disordered protein. Due to the sequence characteristics, it is not surprising that the expression and production of soluble full-length corepressors are quite difficult. As a result, related crystallography or NMR structural studies are all very limited and can only include a short region within the full-length corepressors.

The short NCoR (NCoRsh) construct being used in our study contains all three nuclear interaction domains. Despite the relative low expression temperature (16˚C) and slow growth of the culture, the protein is inclined to form inclusion bodies while being expressed in *E. coli* by itself. The result of limited soluble protein is likely due to the lack of a stable tertiary structure of the protein and the plastic architecture. Furthermore, in the case of the NCoRsh construct, inclusion of N-terminal tags such as GST or SUMO, which are very soluble proteins, still cannot help improve the expression and solubility of NCoRsh. In order to overcome the solubility problem, we have included the TR LBD protein to be coexpressed together with the NCoRsh protein in *E. coli* (Rossetta 2 AI). The TRβLBD and NCoRsh protein constructs are inserted into the ppSUMO and the

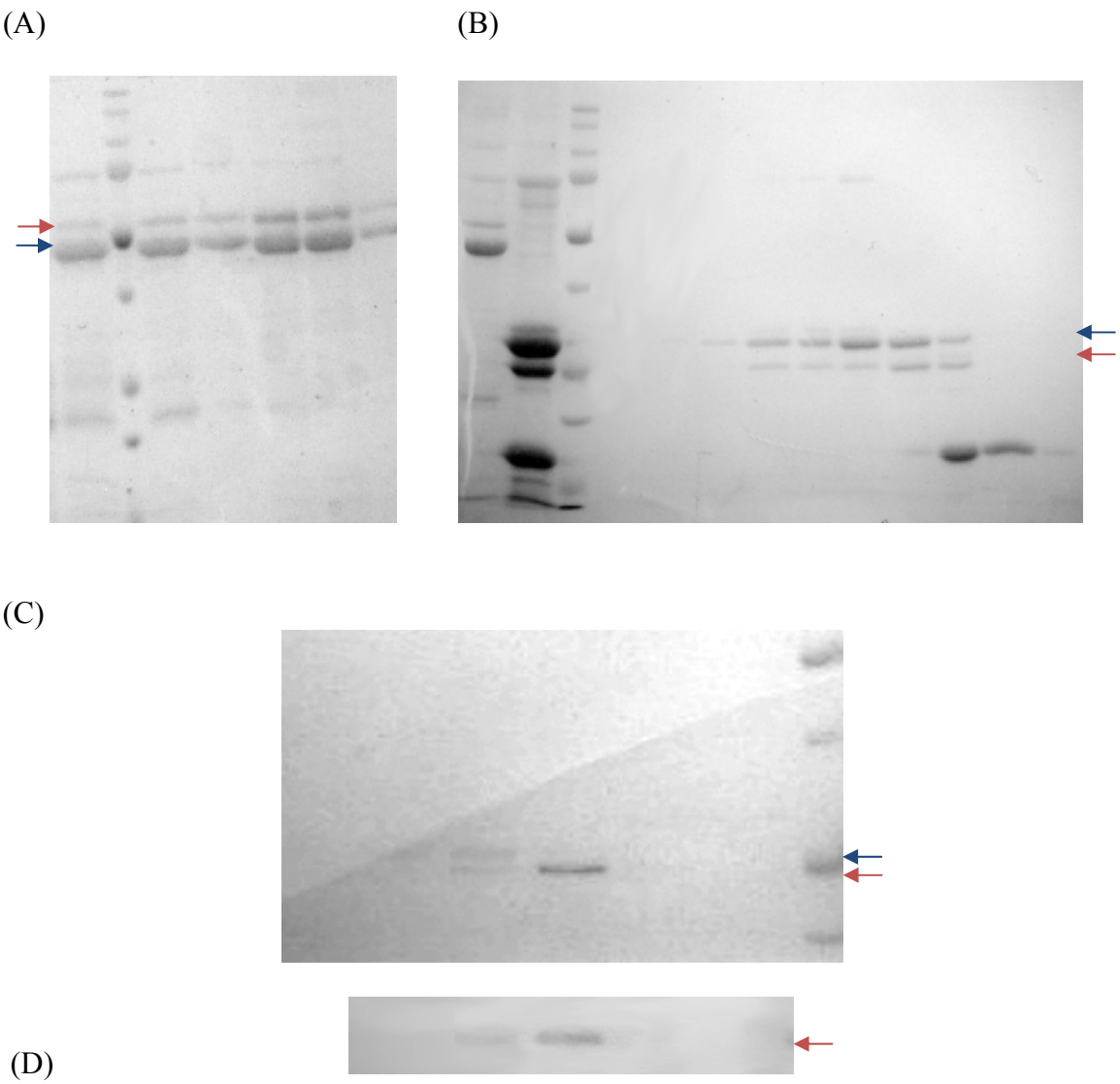



Figure 4.2: Co-expression and purification of the TR LBD – NCoRsh complex. The red arrows represent the positions of NCoRsh bands, and the blue arrows represent the positions of TRβLBD bands. (A) The fractions from the first NTA affinity column. The lanes left to the marker are the fractions from the flow-through portion. The right panels are fraction from the peak in the wash process, and they have been collected for the next purification. (B) The proteins with expression tags and the tag-free proteins (GST or SUMO is cut off) are in the first and second lane, respectively. The lanes to the right of the marker are the flow-through fractions from the second NTA affinity column. The last three fractions are the SUMO and GST tags that have been cut off the proteins. (C) The peak fractions from the Superdex200 gel filtration column. (D) Western blot of the fractions from the gel filtration peak, using the antibody against a peptide from the NCoR N2.

pGex6p plasmid, respectively. Antibiotics including kanamycin, ampicillin and chloramphenicol are used to maintain the expression plasmids and the supplemented tRNAs in the Rosetta 2 AI *E. coli* strain. The auto inhibition properties of the AI cell strain repress the expression of target proteins while the cells grow into the log phase. After the IPTG and L-arabinose are added to the culture, the repression is relieved and both TR LBD and NCoRsh begin to be expressed. As a result of the combinatory factors, the co-expression protocol is able to provide much better yields of soluble NCoRsh proteins in complex with TR LBD, which ends up being around 0.5mg product per liter of culture after purification.

As shown in Figure 4.2, western blot using antibodies against part of the N2 region in NCoR confirms that the protein being purified is the NCoRsh. From the PAGE gels at different stages of purification, it is obvious that NCoRsh has been associated with the coexpressed TR LBD all the time. However, a significant amount of TR LBD protein has been separated from the fraction containing NCoR during the purification process, probably due to the higher expression level of the TR LBD proteins, which usually generate ~5mg per liter of culture while being expressed alone. Nevertheless, the final gel filtration chromatography shows there is still coexisting TR LBD protein in the NCoR fractions. The results suggest that the binding of TR to NCoRsh is strong enough to persist through all steps of purification and probably also helps prevent NCoRsh from precipitation or aggregation. This outcome may also be true *in vivo*, where the binding of NRs is very likely to stabilize the mostly disordered structure of corepressor proteins. On the other hand, the corepressor, with its multiple IDs in a single molecule, may also help stabilize the oligomeric states of the NR proteins.

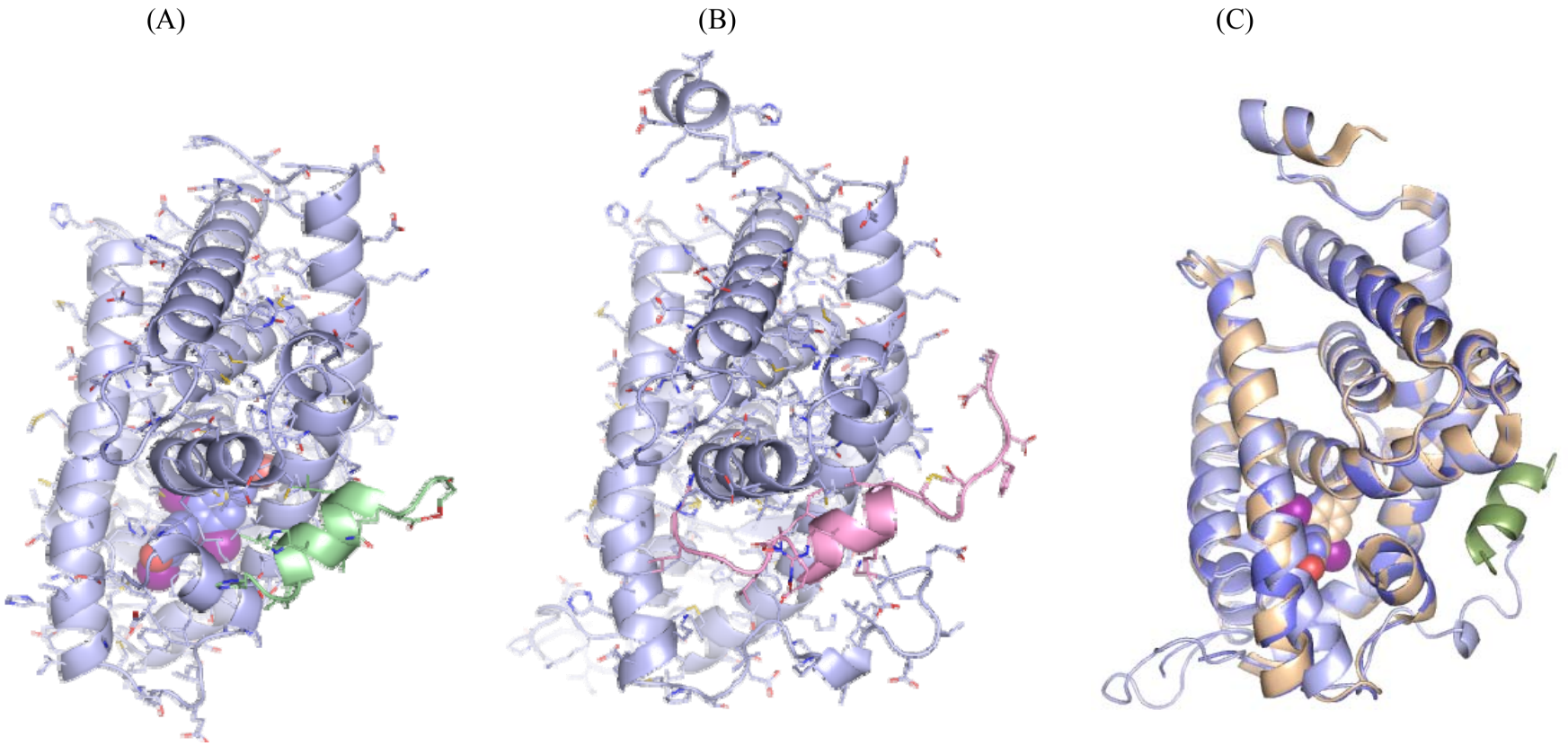


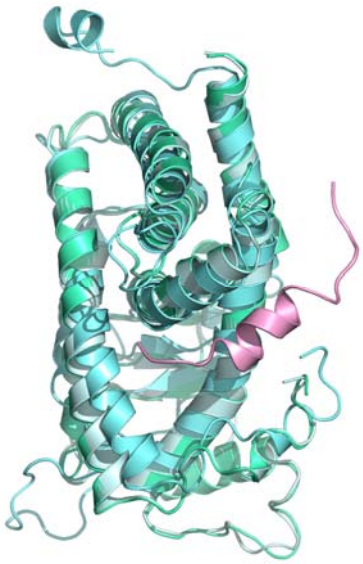

Figure 4.3: A) The crystal structure of the TR LBD and coactivator peptide. B) The homology model of TR LBD and SMRT S2 peptide complex. C) The overlay of the coactivator complex (in blue) and the corepressor complex (in light blue) with the *holo* TR crystal structure (in wheat) in a slightly rotated angle to better display the AF2 region.

### 4.3.2 Computational homology model of corepressor binding

The only corepressor and nuclear receptor complex structure that has been revealed so far is the complex of the PPAR LBD and the CoRNR peptide from the first nuclear interaction domain (S2) of SMRT. The relative lack of structural information suggests that the corepressor bound form of TR LBD, like the *apo* form, may have a significant degree of flexibility in certain regions - and the thermal motions in these flexible regions will interfere with the proper formation of an organized protein crystal lattice. Therefore, we can apply an in-silicon study using the available PPAR-SMRT crystal structure, which will circumvent the difficulties and limitations in the experimental research. In order to study the corepressor and TR interaction mechanism, we have modeled the 3-dimensional structures of the CoRNR box and TRβLBD complex based on the crystal structure of the PPAR/SMRT complex. The 19 residues of the SMRT peptide that are resolved in the PPAR/SMRT crystal structure start from the third residue to the N-terminus of the CoRNR box. Therefore, for simplification, we re-numbered the corepressor peptide residue with this position as a reference (as shown in Figure 4.1 B). Hopefully, by using the homology models of the complex structures, we can better understand the mechanisms controlling specific corepressor interactions.

The stability and reliability of homology models are primarily dependent on the sequence alignment between the target protein and the template protein. The alignment file is a crucial input for the modeling software Modeller. Sequence alignment of either protein or DNA has always been a central part of various bioinformatics studies. The related homology structural study is also an extended approach to utilize and extract the information that has been encoded by the genome. In order to determine the optimum

(A)



(B)

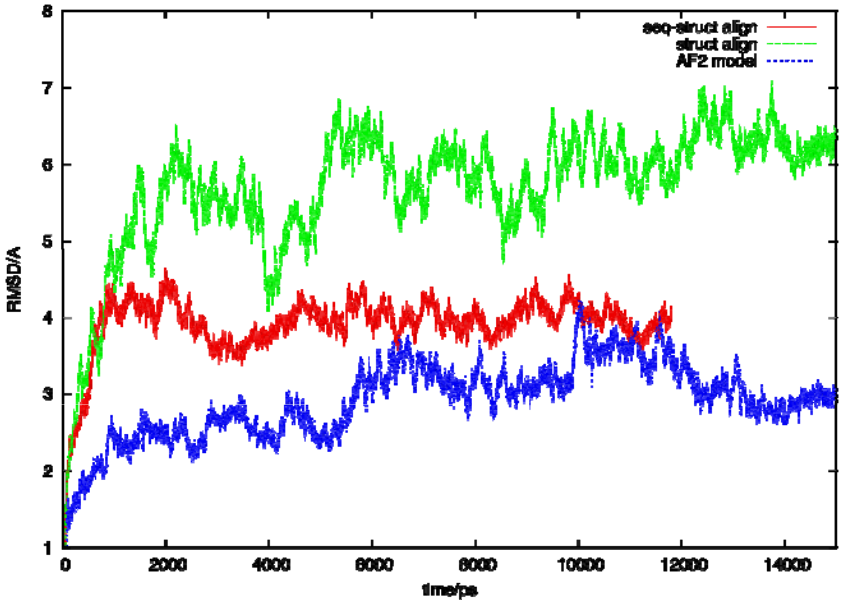



Figure 4.4: The homology models of TR in complex with SMRT S2 peptide, using different sequence alignment protocols: sequence and structure combinatory alignment is colored in green, structure-based alignment is colored in pale-cyan, and the AF-2 partial modeling is colored in cyan. The three structures are aligned together. The SMRT S2 peptides share the same conformation in the three models, which is colored in pink. (B) The rmsd plots of the homology models using different alignment protocols with reference to the starting model.

sequence alignment process suitable for the homology modeling in this study, we have tested the stabilities and energetics of various models using molecular dynamic simulations. We have selected the TRβLBD and the SMRT S2 peptide as the main target system. The alignment protocols being tested include three kinds: a combined sequence-structure alignment, a structural based alignment, and a localized partial alignment at the AF-2 region. The sequence alignment is generated for TRβLBD and PPARγLBD by the online ClustalW2 program (www.ebi.ac.uk/clustalw/) using the default Blosum matrix for the similarity estimation. The structural-based sequence alignment is generated by the alignment command in the structure visualization program PyMol. The localized partial alignment stands for a homology model only at the AF2 region of the TRβLBD, where the most obvious structure deviation occurs while switching between the corepressor bound and coactivator bound conformation (Figure 4.4). The auto sequence alignment function incorporated in the Modeller program is also tested, but only the discrete optimized protein energy (DOPE) scoring function has been applied to evaluate the outcome models. From the DOPE scores (data not shown), the sequence alignment function in Modeller, which is essentially also a structural-based alignment, proves to generate slightly better models than the structural-based sequence alignment generated in PyMol in this case. This result suggests that the gap parameterization in the alignment function of Modeller is more appropriate for the homology modeling.

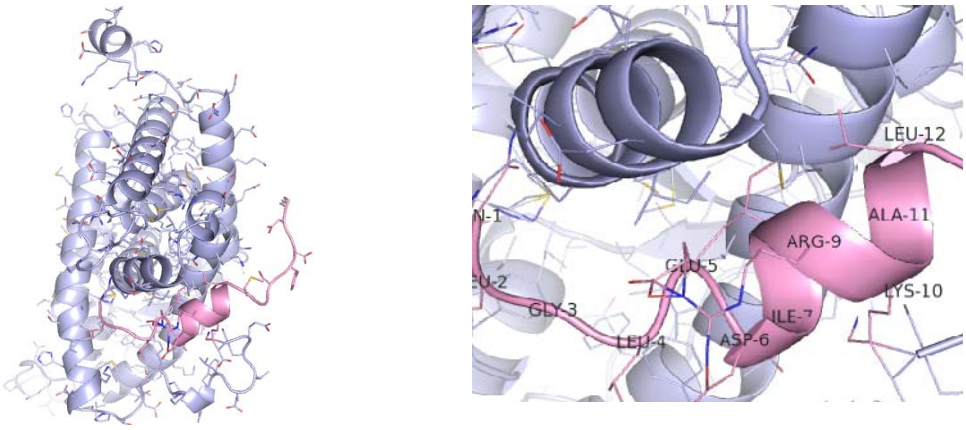
A better way to evaluate the homology model than the evaluation function provided by Modeller is to use molecular dynamic simulations. During the MD simulation, the side chains and local geometry will be energetically optimized while unstable regions or models will display more instability or overall flexibilities. As the rmsd plot shows in

Figure 4.4 B, by comparing the rmsd values of the MD trajectories with reference to the starting model, the AF-2 partial modeling protocol exhibits the minimum deviation from the starting model and appears to be the most stable structure over the simulation time. At the other end of the spectrum, the structure-based modeling appears to be the least stable with more variation of the rmsds. Considering the fact that the only obvious conformation alteration observed in the corepressor complex as compared to the *holo* TR LBD is located around the AF-2 region (Figure 4.3 C), it may also be only necessary to model this region, and the perturbance to the original TR LBD crystal structure can be minimized. As a result, we have decided to use the AF-2 partial modeling protocol in the homology modeling of the other similar complexes formed by TR LBD and ID peptides in NCoR. Accordingly, the homology models are generated and optimized for the complexes formed by TR LBD and the N1, N2 and N3 peptides as described in the Methods section. The resulting model structures are shown in Figure 4.5 for each corepressor peptide complex and for the structural details at the binding interface.

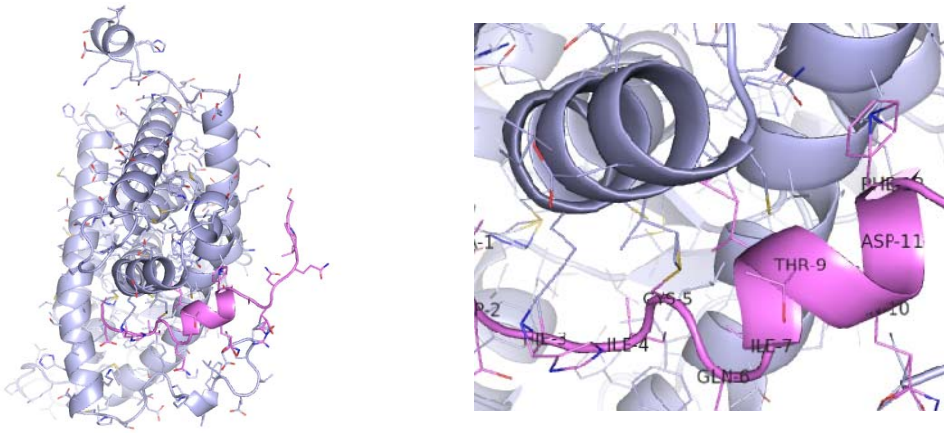
### 4.3.3 Corepressor binding interface

From the homology models of TR with the NCoR IDs (Figure 4.5), we observe that the interfaces between the corepressor peptides and the TRβLBD mostly consist of hydrophobic residues. In particular, Ile280, Thr281, Ile286, Val282, Val283, Ile302, and Leu305 are all within the TRβLBD region that forms the corepressor interfaces. From the average structure calculated from the model MD simulation, we have visualized which of those residues are actually contributing to the hydrophobic interactions with the

A)




B)




C)

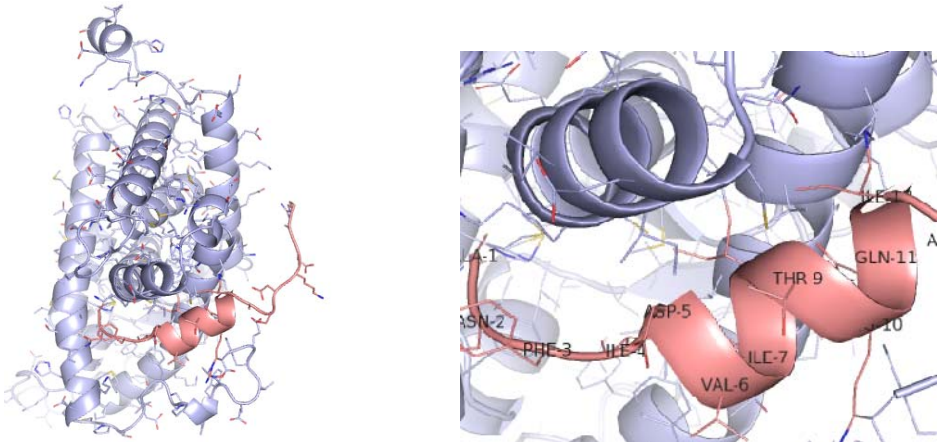


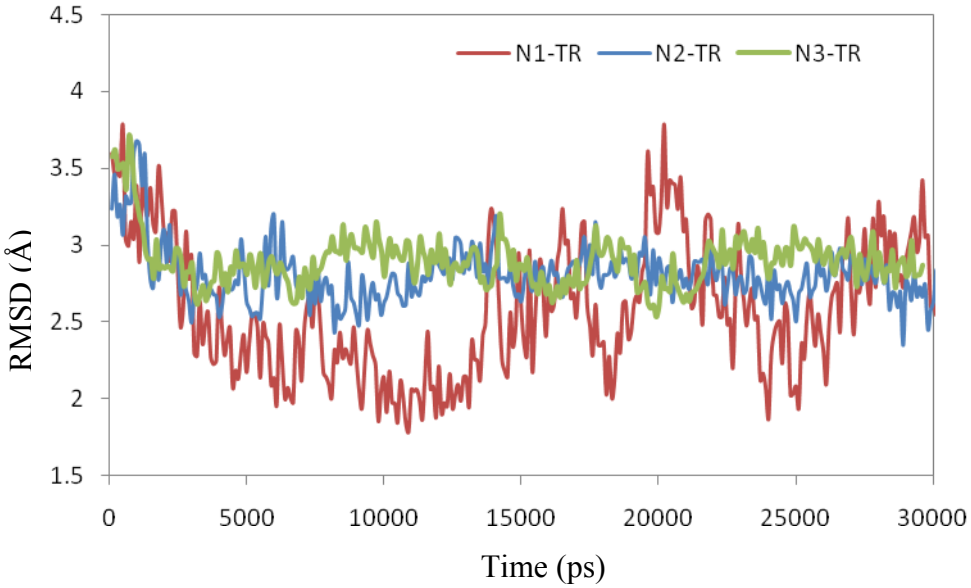

Figure 4.5: The homology models of TR in complex with A) N1 peptide, B) N2 peptide, and C) N3 peptide. On the left is the full view of the complex and on the right is the zoomed-in view at the corepressor binding pocket. a structure-based alignment with an optimized gap-scoring component, generally works slightly better than the alignment generated by PyMol.

corepressor peptide. Especially, Val283 from helix 3, Ile302, and Leu305 from helix 5 are directly facing the short helical region of the corepressor peptide, which contains the most conserved hydrophobic residues in the CoRNR box, namely the isoleucine or valine in the two C-terminal positions of the "I/LXXI/VI" box in all the corepressor peptide. In addition, the N-terminus isoleucine or leucine of the CoRNR box, which is located at a more buried position within the binding pocket, is stabilized by the presence of Ile286 in the helix 3 of the TRβLBD. These preserved hydrophobic effects probably contribute to the residue conservation within the CoRNR boxes of NCoR or SMRT.

The homology models of TR LBD-corepressor complexes also help to elucidate previous mutagenesis mapping studies, which discovered the distinct levels of effect of each individual residue [23, 34]. For instance, the mutation of I280 to either a charged residue or a bulky residue such as methionine will significantly reduce the binding efficacy of all three corepressor peptides of NCoR. This effect is unlikely to result from the direct interaction with the corepressors; rather, it is mostly due to the impact of the mutation on the hydrophobic interactions inside TR LBD and the overall protein folding and stability. On the other hand, the T281 can be mutated to either hydrophobic residues such as leucine or isoleucine with even increased corepressor binding efficiency. This is also not a result from direct interference with corepressor binding, but is more likely due to the stabilizing of some hydrophobic residue side chains in helix 12 in the altered conformation when corepressors are bound.

On the other hand, the electrostatic or polar contacts are rather limited in the corepressor interface. The only contact that is observed in the starting models is at position 10 of the corepressor peptide, which is demonstrated as a positively charged

(A)




(B)

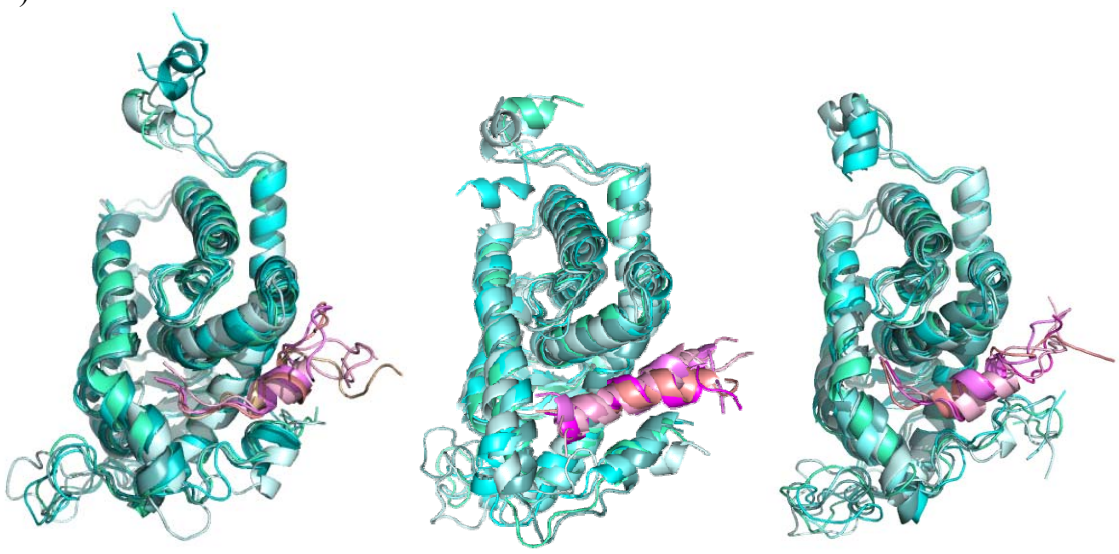


Figure 4.6: (A) Rmsd plot of the TR LBD-NCoR peptide complex models with reference to the average structure in each trajectory. Blue is the TR LBD with NCoR N2, green represents TR LBD with NCoR N3, and red represents TR LBD with NCoR N1. This plot shows the structural variations with respect to the average structure of the simulation and represents the relative flexibilities of the different complexes during the simulation. (B) The representative structures from the MD trajectory clusters of the complex of the TR and NCoR peptides. The trajectory structures are clustered using rmsd as the distance measuring criteria and the complete linkage clustering algorithm. The occurrence of each cluster in the trajectory varies from 10 to 40 percent. The TR LBD and the corepressor peptide are colored separately in cyan and magenta, respectively.

residue or a polar residue with a long side chain, plus the NH2 group present at the terminus of their side chains can reach the backbone oxygen atom of the residues at the very C-terminal of helix 12 in TR LBD. During MD simulation, side chain clashes in the starting model were gradually eliminated. In addition, coordinates of corepressor peptides are better equilibrated and additional polar contacts between the TR LBD and corepressor peptides began to emerge. From the average structure of the MD trajectory, we can get extra information on the polar contacts. For instance, Lys10 in N1 no longer reaches to the backbone oxygen, but has established contact with the acidic side chain of Glu457 instead. The N-terminal Asn1 is also able to contact the backbone atoms in helix5 and helix 10 at both sides. The N2 complex displays more polar contacts, including those between the residues Arg14 and Asp461, Asp11 and Lys288, Gln10 and Glu460, and Asp2 and Lys306, from the N2 peptide and the TR LBD, respectively. As a possible result of the geometry restraints posed by the abundant polar contacts, the conformation of the N2 peptide evolves into a relatively long and stable α-helical structure, compared to the starting model or to the other TR LBD-corepressor peptide complexes. In the N3/TR LBD complex, the polar contacts appear between Gln11 and Thr281, Asp15 and Lys288, and Arg19 and Gln301, in the N3 peptide and the TR LBD, respectively. In addition, there are also transient contacts between the Ala1 in the N3 peptides and the backbone atoms in helix 11 of TR LBD, which keep the N-terminus of the N3 peptide in a position closer to the center of TR LBD, as compared to N1 and N2 (Figure 4.6).

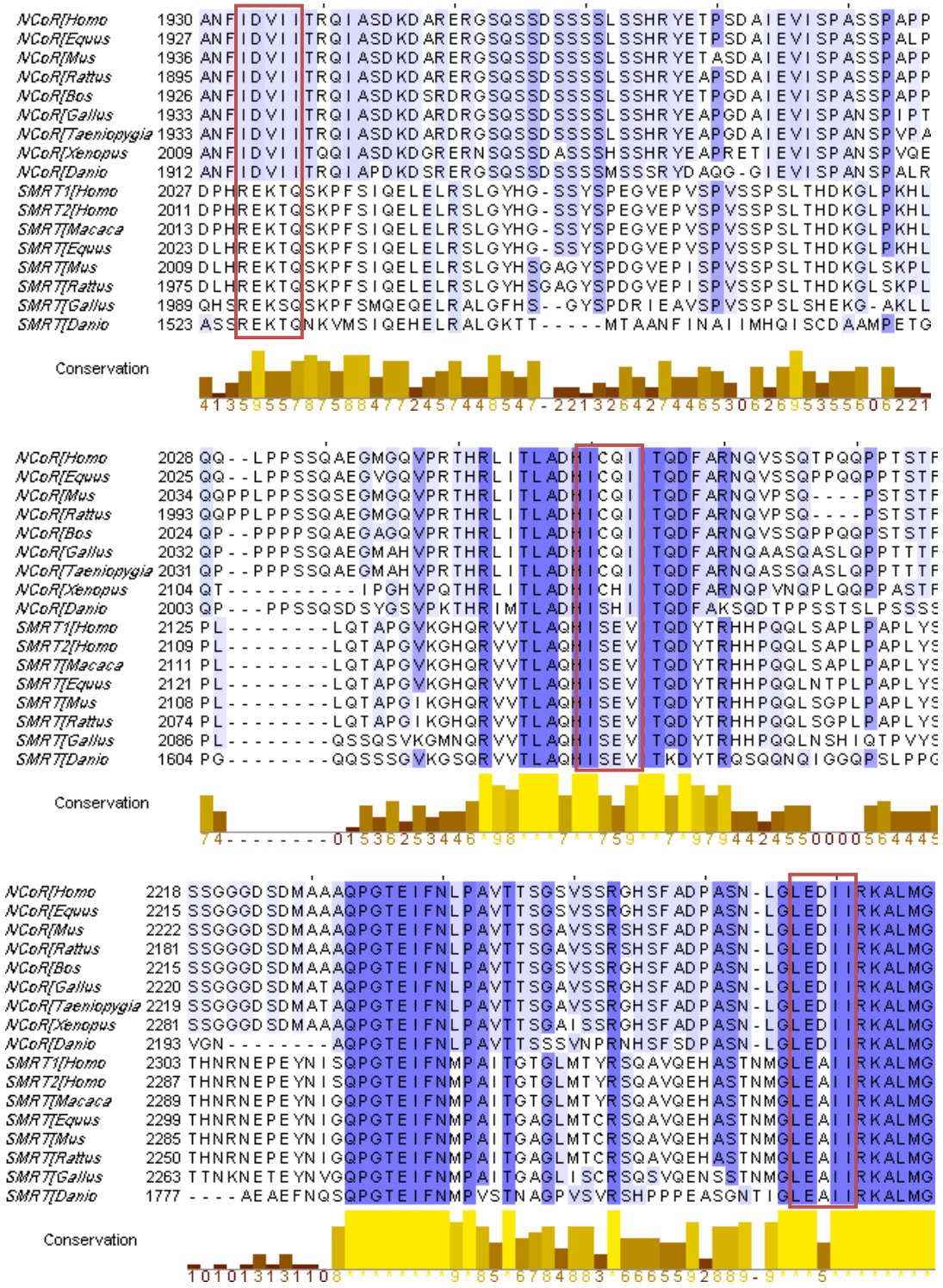


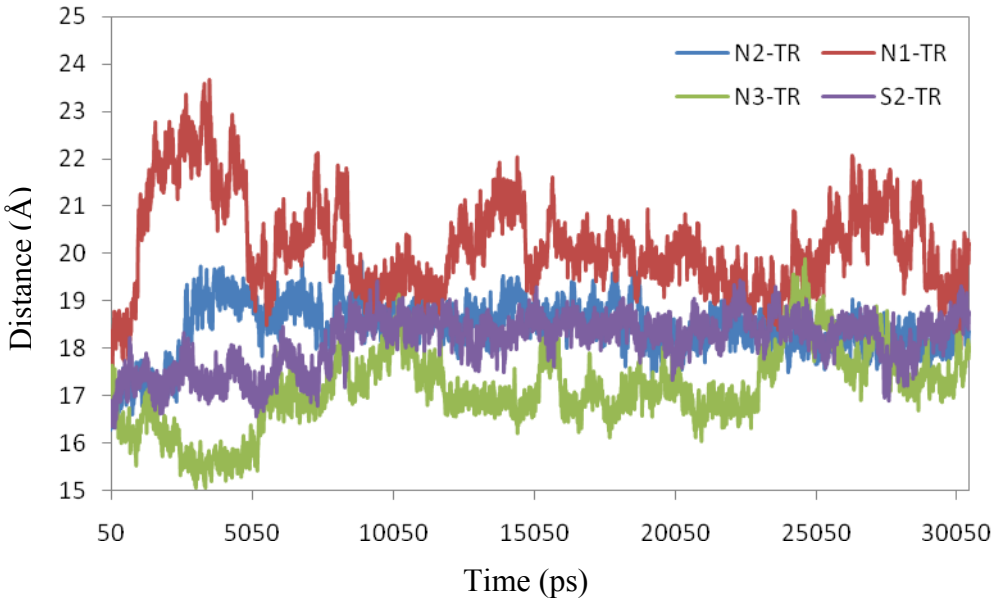

Figure 4.7: Amino acid sequence conservation around the regions of the CoRNR boxes in the NCoR and the SMRT homologs of different organism. The CoRNR box regions are marked by the red squares. The sequence coloring represents the conservation level of each residue. The panels at bottom display the relative conservation score of each residue.

### 4.3.4 Characterization of the distinct nuclear receptor interaction domains

Although the two corepressor proteins NCoR and SMRT are highly related and contain similar domain organization as well (Figure 4.1 and Figure 4.7), TR still preferentially recruits NCoR as its partner in the transcription repression process. On the other hand, TR and RXR among all the other highly related NR proteins are also especially preferred by NCoR, while RAR and PPAR are more preferred by SMRT. The mechanism underlying the specificity of the corepressor and NR interaction is not fully resolved yet. It has been proposed that the presence of a third nuclear receptor interaction domain (ID) at the N-terminal side of the two earlier identified IDs in NCoR is attributed as the key for the favoring of TR; and the absence of the third ID in SMRT is considered the reason for being a less suitable partner for TR. Previous studies have also suggested that the third ID in the NCoR, referred to as N3, is specialized in interacting with TR within the NR members [7]. On the other hand, another important function of the corepressors is to stabilize the formation of the TR homodimer on the DNA response elements. Due to the stoichiometry, only two IDs are necessary to keep contact with the two TRs in the dimer assembly. However, SMRT, which contains only two IDs, is unable to help stabilizing dimerization of TRs [8]. Therefore, it remains unclear how the third domain helps in TR recognition and dimerization and which two of the three NCoR IDs are involved in the TR transcription repression complex.

The interaction of corepressors with the TR is believed to be mediated by the helical motif centered at the CoRNR box in each of the IDs. Although all the CoRNR box sequences in each corepressor ID have the "I/LXXI/VI" form, as shown in Figure 4.1B and Figure 4.7, the exact amino acid composition inside and near the box may contribute

A)




B)

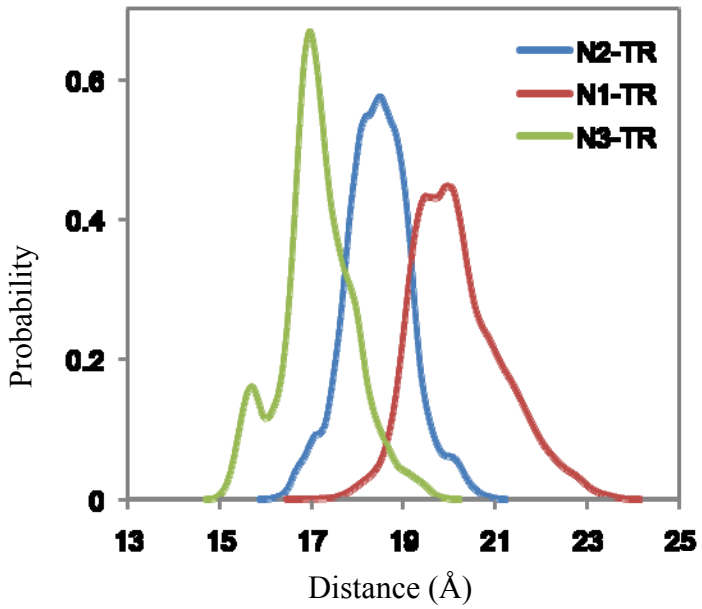


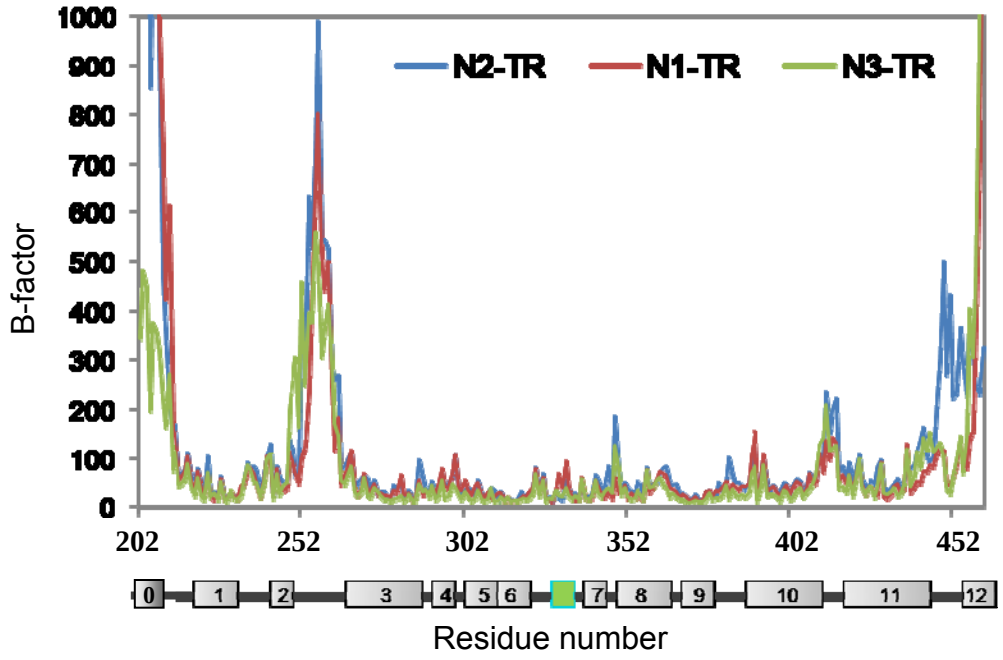

Figure 4.8: Corepressor distances to the TR LBD center. A) The distance values plotted against the simulation time for each TR-corepressor complex simulation. B) The probability distribution of different NCoR corepressor peptides. The x-axis is the distance in angstrom, and the y-axis is the probability values using Gaussian approximation calculated in R.

to the fine characterization of the NR binding specificity. More importantly, this may also be essential for the corepressor and coactivator discrimination. The functional importance probably helps to maintain the significantly high degree of sequence conservation around the CoRNR box through evolution, while most other parts of the corepressors have poor sequence conservation. Mutagenesis studies have proved that a helix-breaking mutation with a proline next to the amino-terminal of the CoRNR box completely blocks the binding of the ID to the TR; whereas an alanine residue introduced at the same position only marginally decreases the binding [8]. Considering the corepressor bound complex model structures, besides the interference on helical formation, the blocking effect of the proline mutation may also come from the rigidity introduced by proline residue, which can create steric clashes with the TR LBD inside the corepressor binding pocket.

With MD simulations of TR-NCoR peptides complexes, we have obtained the dynamic information in addition to the static pictures of the complex structures. In order to understand the dynamics of the interaction between the corepressors and TR, MD simulations have been performed on the homology-modeled complexes formed by the TRβLBD and the corepressor peptides containing the CoRNR box from the N1, N2, N3 in NCoR, or the S2 in SMRT. To measure the closeness as well as the relative movements of the corepressor peptide, we have calculated the distance between the geometric centers of the backbone atoms of corepressor peptide and TRβLBD with no mass weighting. Therefore, the different amino acid compositions of the corepressor peptide can not affect the center position. Only changes in the relative locations of the corepressor peptide, with reference to the TR LBD, are the main factor determining the distance values. The results, as shown in Figure 4.8, show the distinct movements of the

three NCoR peptides. The N1 peptide forms the most distant contact with the TRβLBD and appears to be the most dynamic one as well, as the distances remain the largest and most highly variable during the simulation time. On the other hand, the N3 peptide is closest to the TRβLBD as shown by the distance curve (Figure 4.8 A) and the distribution chart (Figure 4.8B). However, the N3 peptide still displays more variations than the N2 peptide, as judged by the evolving distance value. From the appearance of the peaks and "shoulder" in the probability distribution curve of the distances, the N3 position may have three sub-states with a small, medium and large distance with reference to TRβLBD. In contrast, the N2 peptide is the most fixed one within the TRβLBD binding site, displaying the least variation of the distance values. In addition, the S2 peptide from SMRT seems to share similar characteristics of the binding as the N2 peptide, which include distance values at about the same level and a relative static position as well. Considering the amino acid compositions flanking the CoRNR box of N2 and S2 (Figure 4.7 middle panel), the sequence conservation in the neighboring region probably helps to establish the similarities in their TR binding characteristics, even though two residues within the CoRNR box are distinctive. Accordingly, the even higher resemblance between the amino acid composition of N1 and S1 may lead to similar TR binding scenarios. As a result, although we have not modeled the TR-S1 complex, it is reasonable to assume that our analysis of the TR-N1 complex may also apply to the TR-S1 complex.

(A)




(B)

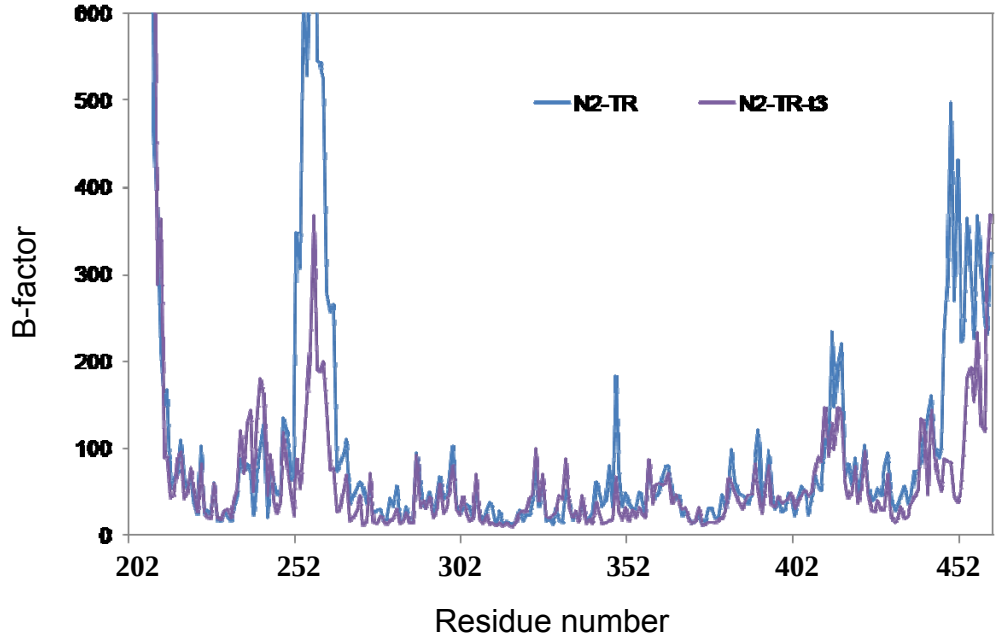



Figure 4.9: B-factor charts of (A) the TR LBD in complex with the three ID peptides in the NCoR; (B) the N2-TR LBD complex in the presence or absence of the T3 ligand. The y-axis of the second chart is adjusted to reveal more details in the regions with smaller range of the b-factor values, while the first chart provides a more complete view at the larger scale. The diagram between two b-factor charts shows the relative position of the helices in the TR LBDs by the rectangles in grey, and the β-turn as the green rectangle.

### 4.3.5 Corepressor induced conformational alterations

As shown by the representative structure from clusters of the MD trajectory (Figure 4.6), the binding of different corepressor peptides generates conformational alterations in several parts of the TR LBD. As part of the corepressor binding interface, helix 12 and helix 11 of TR LBD exhibit the most significant variation in their conformations. In the N1- and the N2-bound complex structures, helix 11 exhibits a more obvious bending, compared to the straight conformation in the starting model. The linker region between helix11 and helix 12 displays a relaxed and flexible conformation in the N1-TR LBD complex at a relative distant position to the main body of TR LBD, while in the N3-bound complex, the linker is more compact and relatively stable, even though the helix 12 almost lost its original α helical conformation. Since the helix11 and helix12 contribute to the corepressor binding interface, these conformational changes can be attributed to direct results of corepressor binding to the TR LBD. On the other hand, as the corepressor is rescued from a disordered state by binding to TR LBD, the peptides also revealed different conformation features such as the increased length of α helix within the N2 peptide, or a decreased length and minor relocation of the α helix within the N1 peptide.

MD also predicts that some regions distant from the corepressor binding site are also affected by corepressor binding. The representative structures from the cluster groups have revealed the subtle conformational variations in the N-terminal hinge helix (helix 0) and the loop region connecting helix 2 and helix 3. In addition, the computed b-factor (atomic fluctuations calculated for the backbone atoms), which represents the mobility of each residue during the MD simulation, is plotted for the corepressor peptide bound and

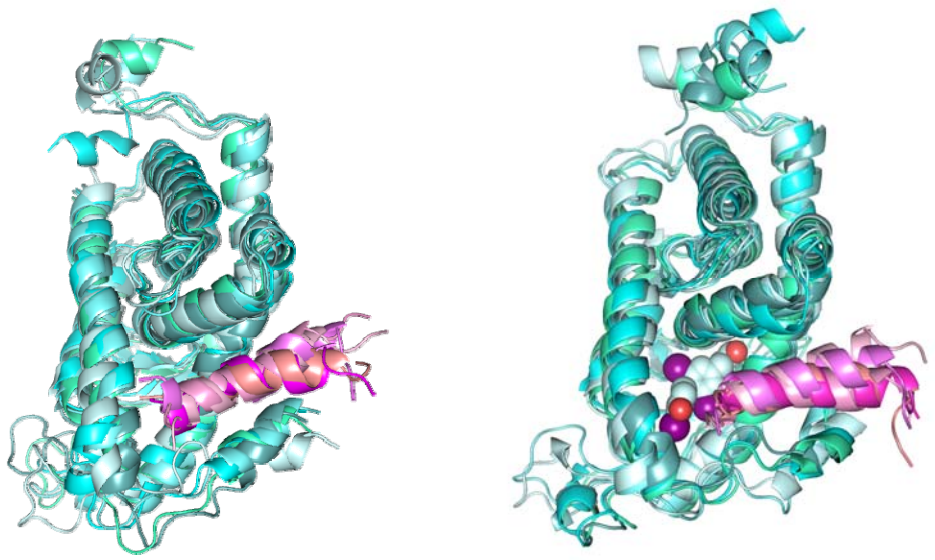
the *apo* state TR LBD (Figure 4.9 A). The relative changes of residual mobilities in different regions of the TR LBDs cannot be simplified by a single rule. Some alterations are common for all corepressor bound complexes as compared to the *apo* state, such as the increased mobility in helix 2, helix 12, and the bending point of helix 11. Other alterations are specified signatures of the individual corepressor peptide. For the N3-bound TR LBD, the flexibility of helix 0 and the loop region between helices 2 and 3 are significantly decreased, compared to the N1, N2 bound, or the *apo* state TR. The modified dynamics of helix 0, as part of the hinge region connecting the TR LBD and the DBD, may be able to transfer the signal of corepressor binding on the LBD, to the DNA binding domain of the TR. The increased mobilities around the connecting region between helix7 and helix 8, as well as between helix 10 and helix 11, are specific in the N2- and N3-bound TR LBDs, as opposed to the N1-bound or *apo* state. The linker between helix 7 and helix 8 is involved in the direct contact with the ligand T3 in the *holo* TR LBD, while the loop between helix 10 and helix 11 is part of the dimerization interface that has been revealed in the symmetric homodimer and heterodimer of related NRs [29-31]. These results suggest that binding of the N2 or N3 peptide to the TR LBD may be more effectively coupled with the functionally related events such as ligand binding or LBD dimerization than binding of the N1 peptide.

### 4.3.6 Ligand induced effect on corepressor binding

One major function of TR LBD is that, in response to the T3 binding, corepressors will be discharged and coactivators will be recruited. Therefore, we have simulated the complex of N2 peptide and the T3-bound TR LBD in order to discover the structural

(A)

(B)



(C)

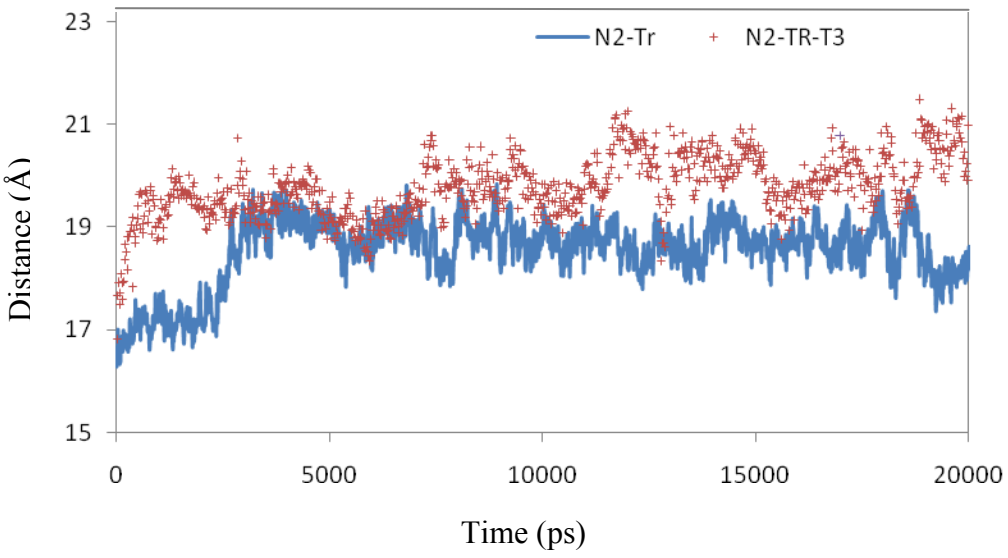



Figure 4.10: (A) and (B) are representative structures from MD trajectory clustering results of the complex of TR and N2 peptides without or with T3 bound, respectively. The trajectory structures are clustered using rmsd as the distance measuring criteria and the complete linkage clustering algorithm. And the occupation of each cluster in the trajectory varies from 10 to 40 percent. The TR LBD and corepressor peptide are colored separately in cyan and magenta, respectively. (C) The distance between the center of N2 peptide and TR LBD, calculated for the TR-N2 trajectories with and without T3 bound.

mechanism behind this function. Because of the limited time scale that can be performed in MD, which is about 20ns in this study, the corepressor release event has not occurred yet. However, the changes of the protein dynamics, which may be responsible for the alteration of corepressor affinity, have begun to manifest within the simulation time. In order to directly measure the propensity of the peptide dissociation, the distances between the center of the peptide and the center of the TR LBD are calculated and plotted against the simulation time in Figure 4.10 C. The steady increase of the distance values in the T3 bound TR-corepressor complex indicates that the T3 binding is indeed driving the corepressor peptide dissociation from the TR LBD during the simulation. Notably, the relative steep increasing step of the distance curve revealed in both simulations is likely due to the obvious establishment of the extended α-helical conformation of the corepressor peptides in both complexes (Figure 4.10 A and B). Additionally, probably due to the augmented dynamics of the N2 peptide as a result of the T3 binding, the N2 peptide within the T3-bound TR LBD is more flexible, and thus is able to adapt the extended α-helical conformation more rapidly than that in the N2-TR LBD complex without T3.

In order to understand the mechanism facilitating the release of the N2 peptide, we have also looked into conformational and dynamic alterations induced by the T3 binding. These alterations are displayed in several regions that are directly or indirectly interfering with corepressor binding. From the representative structures of five rmsd-based clusters of the *holo* TR and N2 complex (Figure 4.10 B), the most obvious changes in the conformation compared to the *apo* TR LBD and N2 complex (Figure 4.10 A) are located at the helix 0, the loop between helix 2 and helix 3, and the linker region between helix

11 and helix 12. In the *holo* TR LBD and N2 complex simulation, a nearly 90˚ twist in the middle of the loop connecting helix 0 and helix 1 causes the reorientation of helix 0. A short helical turn starts to emerge in the middle of the originally disordered loop between helix 2 and helix 3. The helical turn formation in the loop region between helix 2 and helix 3 is accompanied by a significant reduction of the residual mobility in this region, even though the helix 2 flexibility is slightly increased in the T3 bound state. The loop between helix 11 and helix 12 becomes more elevated and adjacent to the corepressor binding pocket, which is correlated with a slight receding of the N-terminus of the corepressor peptide from the binding site. In addition to the conformational alteration, the residue dynamics have also shifted accordingly with the T3 binding. As shown in Figure 4.9B, the b-factor value is also notably reduced in the AF-2 region and the dimerization interface at the connecting loop between helix 10 and helix 11, compared to the apo TR-corepressor complexes. The fluctuation changes may represent possible conformational entropy change within these regions, which can alter the energetics of the corepressor binding to the TR LBD. We are still in the process of analyzing the energetic properties of the corepressor and the TR LBD interactions. Hopefully, the free energy estimations will provide us better quantitative measurements for the binding affinities and specificities as well as the ligand-binding induced effect.

## 4.4 Discussion

From the co-expression and co-purification process, we have proved the relative stable binding between NCoR and TR LBD, and also found that TR LBD binding does help to regularize the NCoRsh conformations and increase the solubility. The

improvements in the experimental procedures can help provide better quantities of the proteins for future experimental studies.

In order to set up a reliable homology modeling protocol, we have compared the MD simulations of three models generated with different sequence alignment methods. As expected, the partial modeling that only alters the conformation in the AF2 region in order to incorporate the corepressor peptide is the most stable and probably most credible model of the TR-corepressor complex. With the minimized alteration from the original holo TR LBD crystal structure, the major parts of the molecule are in their comfortable position and help to maintain the overall stability of the molecule. The incorporation of the corepressor peptide is enough to induce conformational dynamic alterations that can be discovered by MD simulations. The combinatory sequence and structure protocol is similar to that has been employed in the Chapter III homology modeling studies, which also proved to be relatively stable. However, due to the more dramatic conformational alteration in the core region of the LBD of the *apo* RXR, compared to the *holo* TR LBD crystal structures, a partial homology model is not applicable in the case of the *apo* TR LBD homology modeling using the *apo* RXR LBD as the template.

From our homology models of TR and corepressor peptides, a refined picture of TR-corepressor interactions is revealed for the first time. The interface between corepressor peptides and the TRβLBD in the starting models mostly consists of hydrophobic residues that contribute to the binding. MD simulations of the models have restabilized the side chain positions and also revealed alteration in the backbone coordinates of both TR LBD and the corepressor peptides. The gradual appearance of extra polar interactions between the TR LBD and the corepressors through the MD simulation are the indispensable force

driving the reshaping of the corepressor peptides. During the simulation, the three NCoR peptides have revealed distinct conformations and dynamic characteristics. From their conformational and positional dynamics, TR LBDs have shown preference for binding of the N2 and N3 peptides in the NCoR, which is consistent with previous experimental results. Moreover, the divergent N3 domain present solely in the NCoR but not in SMRT is the primary cause of the differentiated fondness of the two corepressor proteins for TR. Furthermore, incorporation of T3 ligand into the TR-corepressor complex proves to facilitate the corepressor peptide dissociation from TR LBD. T3 binding also alters the mobilities at several regions within TR LBD, which may contribute to the dimerization of the TR LBDs or communications with the TR DBD.

## 4.5 Methods

### 4.5.1 Protein expression and purification

The ligand binding domain fragment of the rTRβ and the NCoRsh were expressed in BL21 (Rossetta 2 AI) E.*coli* cells as a GST-fusion and a SUMO-fusion protein, respectively. The Rossetta2 E.*coli* strain proved to effectively eliminate truncated protein expression due to its additional tRNA genes for rare codons located within the construct [36]. The auto-inhibition control is provided by inserting a chromosomal copy of the T7 RNA polymerase gene under the tight control of the arabinose-inducible araBAD promoter. The extra arabinose-inducible control helps to synchronize and manage the expression of the two proteins only after the induction in an altered low temperature condition. The construct of the NCoRsh containing all three IDs covers the amino acids from 1925 to 2300. Glucose was included in the medium to repress expression during the growth for 5-6 hours in 37˚C until OD600 reached about 1. Then, the cells were induced

with 0.2mM IPTG and 0.2%w/v arabinose for 24 hours at 16°C. After harvesting, the cells were resuspended and lysed in a buffer containing 20mM potassium phosphate, pH 7.5, 150 NaCl, 10% glycerol, 5mM dithiothreitol and freshly added 1mM PMSF. The bacterial lysates were centrifuged at 30,000rpm for 45 minutes and the supernatants were collected. The supernatants were subjected to purification first on a Ni-NTA column (GE Healthcare), and dialyzed to remove the extra imidazole in the buffer while removing the SUMO and GST tags by proteolysis at 4˚C overnight. Subsequently, the proteins were loaded onto a second Ni-NTA column (GE Healthcare) to remove the tags and other proteins with Ni affinities. Afterwards, the rTRβ LBD and NCoRsh protein were concentrated and purified to homogeneity on the Superdex200 (GE Healthcare) gel filtration column. The final protein concentration was estimated based on calculated extinction coefficient values.

### 4.5.2 Homology Modeling

Homology models of the complex of TR LBD bound to peptides including the interacting box of SMRT or NCoR were developed using the program Modeller [32]. The structure template came from the crystal structure of the PPAR-SMRT complex with antagonist GW6471 bound (PDB code 1KKQ) [25]. A sequence alignment of TR/PPAR was first performed through ClustalW [33], followed by a structure-based sequence alignment by superimposing the *holo* TR LBD crystal structure (PDB code 1XZX) onto the PPAR part of the template structure performed in Pymol or with the structure-based alignment program included in the Modeller [32]. Both sequence-based and structure-based alignments were examined and used as alignment inputs for the Modeller program. The partial AF2 modeling was generated by using the sequence alignment only for the

AF2 region as input, and the other part of the TR LBD came from the holo TR LBD crystal structure with a filled-in loop between helix 2 and helix 3 from the partial antagonist (KB131084)-bound TR LBD crystal structure (PDB code 2J4A). With each group input of structural template and alignment files, 5 homology structures were generated and evaluated by the discrete optimized protein energy (DOPE) scoring function in Modeller. The final model with the lowest global DOPE score was subjected to energy minimization and molecular dynamics simulations.

### 4.5.3 Molecular Dynamics Simulations

The MD simulations of the TR LBD and corepressor peptide complex model were conducted using Amber9. Preparation of the system was done in three steps. First, the TR LBD molecule was solvated by adding water molecules (TIP3P), creating a periodic octahedron box with a minimum distance of 10.0 Å from the TR LBD to the box edge. Then K+ ions were added to neutralize the system. Afterward, K+ and Cl- ions were added to the system to achieve the concentration 0.15M mimicking the physiological ionic level. The final system went through an initial step of energy minimization and then was heated to 303 K in 5ps and equilibrated by 45 ps at a temperature of 298 K in several steps with decreasing levels of harmonic positional restraints from the highest 50 kCal/mol to as low as 2.5 kCal/mol, first on all the protein atoms and last only on the main chain atoms of the protein residues. After the setup, the coordinates and velocities in the system were equilibrated and prepared for MD. Then the positional restraints were removed and MD simulations were performed at constant pressure of 1atm (101,325 Pa). Bonds involving hydrogen atoms were constrained to their equilibrium length by the

SHAKE algorithm. The electrostatic energy was calculated by the particle-mesh Ewald method with the default setting in Amber 9.

# Bibliography


1. Seol, W., H.S. Choi, and D.D. Moore, *Isolation of proteins that interact specifically with the retinoid X receptor: two novel orphan receptors.* Mol Endocrinol, 1995. **9**(1): p. 72-85.
2. Horlein, A.J., et al., *Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor.* Nature, 1995. **377**(6548): p. 397-404.
3. Chen, J.D. and R.M. Evans, *A transcriptional co-repressor that interacts with nuclear hormone receptors.* Nature, 1995. **377**(6548): p. 454-7.
4. Yu, V.C., et al., *RXR beta: a coregulator that enhances binding of retinoic acid, thyroid hormone, and vitamin D receptors to their cognate response elements.* Cell, 1991. **67**(6): p. 1251-66.
5. Beato, M., P. Herrlich, and G. Schutz, *Steroid hormone receptors: many actors in search of a plot.* Cell, 1995. **83**(6): p. 851-7.
6. Wong, J., Y.B. Shi, and A.P. Wolffe, *A role for nucleosome assembly in both silencing and activation of the Xenopus TR beta A gene by the thyroid hormone receptor.* Genes Dev, 1995. **9**(21): p. 2696-711.
7. Cohen, R.N., et al., *The nuclear corepressors recognize distinct nuclear receptor complexes.* Mol Endocrinol, 2000. **14**(6): p. 900-14.
8. Makowski, A., et al., *Determination of nuclear receptor corepressor interactions with the thyroid hormone receptor.* Mol Endocrinol, 2003. **17**(2): p. 273-86.
9. Damm, K., C.C. Thompson, and R.M. Evans, *Protein encoded by v-erbA functions as a thyroid-hormone receptor antagonist.* Nature, 1989. **339**(6226): p. 593-7.

10. Sap, J., et al., *The c-erb-A protein is a high-affinity receptor for thyroid hormone.* Nature, 1986. **324**(6098): p. 635-40.
11. Damm, K., et al., *Functional inhibition of retinoic acid response by dominant negative retinoic acid receptor mutants.* Proc Natl Acad Sci U S A, 1993. **90**(7): p. 2989-93.
12. Tsai, S. and S.J. Collins, *A dominant negative retinoic acid receptor blocks neutrophil differentiation at the promyelocyte stage.* Proc Natl Acad Sci U S A, 1993. **90**(15): p. 7153-7.
13. Renaud, J.P., et al., *Crystal structure of the RAR-gamma ligand-binding domain bound to all-trans retinoic acid.* Nature, 1995. **378**(6558): p. 681-9.
14. Yen, P.M., *Molecular basis of resistance to thyroid hormone.* Trends Endocrinol Metab, 2003. **14**(7): p. 327-33.
15. Johnson, A.D., *The price of repression.* Cell, 1995. **81**(5): p. 655-8.
16. Ordentlich, P., et al., *Unique forms of human and mouse nuclear receptor corepressor SMRT.* Proc Natl Acad Sci U S A, 1999. **96**(6): p. 2639-44.
17. Park, E.J., et al., *SMRTe, a silencing mediator for retinoid and thyroid hormone receptors-extended isoform that is more related to the nuclear receptor corepressor.* Proc Natl Acad Sci U S A, 1999. **96**(7): p. 3519-24.
18. Li, H., et al., *Characterization of receptor interaction and transcriptional repression by the corepressor SMRT.* Mol Endocrinol, 1997. **11**(13): p. 2025-37.
19. Jones, P.L. and Y.B. Shi, *N-CoR-HDAC corepressor complexes: roles in transcriptional regulation by nuclear hormone receptors.* Curr Top Microbiol Immunol, 2003. **274**: p. 237-68.

20. Webb, P., et al., *The nuclear receptor corepressor (N-CoR) contains three isoleucine motifs (I/LXXII) that serve as receptor interaction domains (IDs).* Mol Endocrinol, 2000. **14**(12): p. 1976-85.

21. Cohen, R.N., et al., *The specificity of interactions between nuclear hormone receptors and corepressors is mediated by distinct amino acid sequences within the interacting domains.* Mol Endocrinol, 2001. **15**(7): p. 1049-61.

22. Hu, X. and M.A. Lazar, *The CoRNR motif controls the recruitment of corepressors by nuclear hormone receptors.* Nature, 1999. **402**(6757): p. 93-6.

23. Hu, X., Y. Li, and M.A. Lazar, *Determinants of CoRNR-dependent repression complex assembly on nuclear hormone receptors.* Mol Cell Biol, 2001. **21**(5): p. 1747-58.

24. Ghosh, J.C., et al., *Interactions that determine the assembly of a retinoid X receptor/corepressor complex.* Proc Natl Acad Sci U S A, 2002. **99**(9): p. 5842-7.

25. Xu, H.E., et al., *Structural basis for antagonist-mediated recruitment of nuclear co-repressors by PPARalpha.* Nature, 2002. **415**(6873): p. 813-7.

26. Jepsen, K. and M.G. Rosenfeld, *Biological roles and mechanistic actions of co-repressor complexes.* J Cell Sci, 2002. **115**(Pt 4): p. 689-98.

27. Huang, M.E., et al., *Use of all-trans retinoic acid in the treatment of acute promyelocytic leukemia.* Blood, 1988. **72**(2): p. 567-72.

28. Melnick, A. and J.D. Licht, *Deconstructing a disease: RARalpha, its fusion partners, and their roles in the pathogenesis of acute promyelocytic leukemia.* Blood, 1999. **93**(10): p. 3167-215.

29. Bourguet, W., et al., *Crystal structure of the ligand-binding domain of the human nuclear receptor RXR-alpha.* Nature, 1995. **375**(6530): p. 377-82.

30. Chandra, V., et al., *Structure of the intact PPAR-gamma-RXR-alpha nuclear receptor complex on DNA*. Nature, 2008: p. 350-356.

31. Uppenberg, J., et al., *Crystal structure of the ligand binding domain of the human nuclear receptor PPARgamma.* J Biol Chem, 1998. **273**(47): p. 31108-12.

32. Fiser, A. and A. Sali, *Modeller: generation and refinement of homology-based protein structure models.* Methods Enzymol, 2003. **374**: p. 461-91.

33. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0.* Bioinformatics, 2007. **23**(21): p. 2947-8.

34. Marimuthu, A., et al., *TR surfaces and conformations required to bind nuclear corepressor*. Mol Endocrinol, 2002, 16: p.271-86.

35. Lee, S. and Privalsky, M.L., *Multiple mutations contribute to repression by the v-Erb A oncoprotein*. Oncogene, 2005, 24(45): p.6737-52.

# Chapter V

# Conclusions and Perspectives

From recognizing the linkages between TR function and physiological conditions to the discovery of the functional pathways of nuclear receptor mediated gene transcription regulation, significant advances have already been made in the study of thyroid hormone actions in the past few decades. Based on previous studies, connections between thyroid hormone actions and the regulation of gene expression by thyroid hormone receptors have been revealed. As a member of the nuclear receptor protein family, TRs share a common functional theme with the other NRs, which is the transmission of signals manifested as fluctuations of cellular hormone levels to the specific up- or down-regulation of the expression levels of hundreds of target protein genes. To achieve very specific regulation of such a large number of genes with a limited number of NR proteins, the actions of the NRs are fine-tuned by many different factors. Some of the factors that have been identified include: 1) binding of ligand; 2) sequence and organization of the DNA response element in the promoter region of target genes; 3) cross-talk with other signal pathways via competitive heterodimerization and binding to DNA response elements; 4) cellular levels of coregulator proteins and their binding effect; 5)

developmental stage dependent and tissue specific expression of different TR isoforms. The work in this thesis is primarily focused on the first four factors: how they influence the TR activity and the inter-relationship between these factors.

Our knowledge of TH and TR actions has benefitted and co-advanced with many key developments in biomedical science and technologies. For example, recent human genome project and technologies such as microarrays, fluorescent spectroscopy, and genetic engineered mouse models have provided us with many powerful new tools to study the complex aspects in the TR action. As has been discussed in the previous chapters of this dissertation, more advanced biophysical and computational techniques have started to be applied in the study of TR functions. I have been fortunate to take advantage of the well-developed fluorescent anisotropy (FA) technique to quantify affinity of the TR DBD and the DNA binding in Chapter II. The combined application of computational modeling and MD simulations in Chapter III and Chapter IV also extends our understanding in the area.

The cooperative DNA binding ability that was investigated in Chapter II is an important component of the transcription regulation function of TR. Previous transcription assays have shown that, in the absence of RXR, TR can activate transcription on the F2 TRE to a much higher level than it can on the DR4 and Pal TREs. The positive cooperativity that we have observed for TR homodimer binding to the F2 TRE but not to the DR4 may represent an important component of the differential activation because it can strongly influence both the TR homodimer affinity to F2 as well as the subsequent recruitment of coregulator proteins. Our crystal structure of the TRβDBD homodimer on the F2 TRE have revealed possible inter-monomer contacts in

the C-terminus region of the DBDs. While both the native gel shift assays and FP experiments clearly support cooperative binding of two TRβDBD subunits to the F2 TRE, the mechanism that underlines this cooperativity is not completely revealed by our crystal structure. So far, few nuclear receptors have high-resolution structures known for the monomer-DNA and DNA-free states in addition to the dimer-DNA complex, which limits our understanding of the molecular mechanism responsible for the observed cooperativity. My mutagenesis studies suggest that the CTE helix of the TRβDBD contribute some, but not all, of the binding cooperativity via its ability to form inter-subunit contacts at the C-terminal tail region. Varying the inter-half-site spacing can alter the cooperativity and shift the preference of monomer or dimer formation on the F2 TRE. At the same time, elimination of any possible direct inter-subunit contacts by the deletion of a considerable portion of the CTE helix failed to completely abolish cooperative subunit binding. This leaves two possible explanations for the observed cooperativity: 1) non-bonded, i.e. electrostatic, interactions between the DBD subunits, or 2) DNA-mediated inter-subunit effects.

Although a DNA-mediated effect was proposed to be a major factor contributing to the cooperativity of NR binding in the glucocorticoid receptor by Meijsing *et al.* in 2009, there is currently no direct evidence supporting this hypothesis. One model for the DNA-mediated effects is that the binding of the first subunit reduces the overall conformational entropy of the F2 TRE so that the binding of the second subunit pays less of an entropic cost and thus is more favorable. As a result of the positive cooperativity, the dimer-DNA complex will always be the predominant species within any mixture of the TR DBD and the F2 TRE. Therefore, it is not possible to isolate monomer-DNA complexes in solution.

Even though there are experimental methods such as isothermal titration calorimetry (ITC), which can measure the enthalpy change, and by data processing can also estimate the entropy change during a binding process, it is still not possible to separately study the monomer binding and the dimer binding energetics within the context of the same DBD and DNA molecules. If we constrain the binding in the monomer-binding mode by using a scrambled F2 DNA with a single binding site, the binding energetics will also change due to the altered DNA sequence, which is also difficult to be quantified and excluded from the comparison between monomer binding and dimer binding. As a result, the DNA entropy effect is difficult to test experimentally but will be the subject of future computational simulation studies. Fortunately, one of the advantages of MD simulation is that we can easily setup and study a system of the TR DBD monomer binding to the F2 TRE, and analyze conformational dynamics as well as energetics for this system. In this case, the MM_PBSA module in AMBER will provide a powerful tool to assess the entropic contribution to the cooperative binding. Moreover, detailed analysis on the dynamic features of NRE DNA or NR conformations during simulations of different NR bound states may reveal factors contributing to alterations in the DNA binding affinities and the observed allosteric and cooperative binding effects. Interestingly, a remarkable advance in the study of protein-DNA complexes by the MD simulation is started with a system containing a homodimer of glucocorticoid receptor DBD with a DNA fragment published by Eriksson *et al.* in 1995. However, this early work is only focused on analysis of conformational dynamics.

We also need keep in mind that both experimental and computational techniques have

quite different from the actual TR-DNA binding within the cell. The TR binding sites are located within much longer DNA helices in chromatin, and the DNA length as well as chromatin structure will have significant effect on the TR-DNA binding and conformational dynamics.

The cooperative binding not only improves the DNA binding affinity, but may also improve dimerization of full-length TRs, which will increase their affinity for coregulator proteins that incorporates multiple NR binding site within a single molecule and may lead to a further increase of transcriptional activity. A straight forward way to test the effect of cooperative DNA binding on the transcription activity of TR is to perform the transcription assay with mutant TRs that abolishes the cooperativity conveyed by the DBDs. For instance, the Leu99Glu mutant has shown decrease cooperativity as quantified by our FA data. A double mutation including Leu99Glu and Lys101Glu may be even more effective abolishing the cooperativity, as the Lys101Glu mutation will eliminate the only positive charge within the C-terminus of the TR DBD and present a C-terminus region full of negatively charge residues to increase electrostatic repulsions.

The flexible orientation of the CTE of the TR DBD, enabled by the T-box hinge, is another novel feature revealed by our crystal structure in Chapter II. The disordered CTEs of some other NR DBDs, such as the RXR DBD and the PPAR DBD, allow the positioning of the hinge region and the LBDs in a relatively wide range. Comparatively, the more rigid α-helical structure of the CTE of TR DBD restricts the positioning of the hinge region, which connects to the TR LBD. Therefore the extra freedom provided by the T-box hinge is especially important for TR. The flexibility can facilitate the formation of a symmetric LBD dimer associating with a DBD dimer in either a symmetric

arrangement on F2 or an asymmetric arrangement on DR4. Moreover, when LBD dimerization is destabilized by ligand binding, the flexibility of the CTE of TR DBD will allow TR to stay bound to the TRE as dimer even when the LBD dimer assembly is disrupted. However, with a disordered CTE in RXR DBDs, the different orientation in the LBDs and the DBDs of the TR-RXR heterodimer will be more easily accommodated. This may contribute to the preferred TR-RXR heterodimer assembly on DR4 compared to the TR homodimer. Furthermore, the vitamin D receptor, which also displays an α-helical conformation in the CTE of the DBD, may share similar features as TR DBDs in order to coordinate dimers of the DBD with dimers of the LBD. To further analyze the influence of a more elastic or a more rigid CTE on TR function, a mutant TR could be designed. For instance, a double mutation change the two residues in the CTE of TR DBDs, which contact DNA backbone phosphates in the canonical conformation, to negatively charged residues could free the tethering of the CTE to DNA and increase its flexibility. On the other hand, mutations in the C-terminal portion of the T-box region to positively charged residues could introduce more contacts with the DNA backbone and probably decrease the freedom of the CTE.

The combination of crystallography and molecular dynamics simulation has provided us additional perspective on the molecular structures in their static and dynamic states corresponding to different functional situations. The past crystallographic studies have revealed the structures of TR LBDs in complex with the natural ligand T3, some synthetic agonists, and a partial antagonist. However, previous TR LBD structures all resemble the *holo* TR LBD conformation and no conformation changes in the AF-2 domain have ever been directly observed in crystal structures. Accordingly, Chapter III in

this thesis reported homology modeling of the *apo* TR LBD based on two currently resolved *apo* NR LBD structures with varied AF-2 conformations. Recent advances in computational power and theoretical techniques enable the more reliable homology modeling, analysis, and evaluation as well. From my analysis of conformational dynamics presented by homology models of the TR LBD, it is quite surprising that the PPAR-based model of *apo* TR LBD is the more stable and energetically favored conformation, compared to the RXR-based model. A "closed" conformation of the AF2 region in the *apo* TR LBDs, similar to that in the *holo* TR LBD structure, also appears to be more stable than an "open" conformation as found in the *apo* RXR LBD, even in the absence of ligand. These results indicate that there may be no significant structural differences between ligand-free and ligand-bound TRs. Also, my results disagree with the previous theory that ligand-induced functional transition of TR relies on a ligand-dependent structural switch in the AF-2 region, mimicking the structural switch of RXR LBDs in the ligand-free state or ligand-bound state. Therefore, molecular aspects of the ligand binding effect on TR may be more complicated than previously assumed. Further studies should consider the possibility that change of conformational dynamics in addition to alteration of static structures may be an important factor determining ligand-dependent TR activation.

If we could visualize a progression of the RXR-based *apo* LBD model, which in the end converges to the PPAR-based model, it will be more compelling evidence proving that the PPAR-base model is what the *apo* TR LBD looks like. However, this structural transition will not only involve conformational change in the AF-2 region, but also requires a significant bending of the helix 3, which will present a significant energy

barrier between the two conformational states. Current computer power may limit our observation of possible structural ensembles that are distant to the starting point in the conformational space. Several MD methods have been developed to address this problem. For instance, targeted MD (TMD) and steered MD (STD) can accelerate a conformation transition by adding artificial constraints or forces given defined start and finish structures. A negative friction mode that has been described by a talent former colleague in University of Michigan, Jeff Wereszcynski, can also help to propel the exploration in the conformational space (not necessarily in the correct direction though). The local enhanced sampling (LES) may help to more effectively study several possible local conformations within a region simultaneously, i.e. the AF-2 in this study. However, when the structural variation is not just limited within the region, as in the RXR-based and the PPAR-based model, LES can not readily be applied to the system. But if we consider two models based on the "partial-closed" and "closed" AF-2 conformation in the two monomers in the *apo* PPAR LBD structure, LES can be applied to study the dynamics of the AF-2 in both conformations at the same time.

It is known that TR recruits corepressors in the absence of ligands, and is able to switch to coactivator binding in response to ligand binding. More importantly, the level of transcription activity of TR is dependent on this dynamic exchange between corepressors and coactivators. However, my results have shown that the *apo* TR LBD is more similar to the coactivator-bound conformation rather than the corepressor-bound conformation, especially in the AF-2 region. Based on such considerations, binding of corepressors to the *apo* TR LBD will be an induced fit process, which comprises adjustment of the AF-2 conformation only in the presence of corepressors. Thus,

although the structures are similar, the *apo* and *holo* TRs may differ in their abilities to stabilize various complexes with corepressors or coactivators. Therefore, it is very likely that TRs will present different degrees of dynamic behavior in accordance with their constitutive repression and ligand-dependent transcription activation, which is not just a simple on-off switch. How the dynamics of the AF-2 contribute to different stages through which it finally modulates vital processes *in vivo* deserves further investigation.

My studies also suggest that there are regions other than the AF-2 within the TR LBD, which also contribute to TR actions. The regional rmsd data have shown that some other parts of the TR LBD may also contain the potential to undergo structural and dynamic rearrangements in a ligand-dependent manner. Certain TR mutants have been identified in some patients with the resistance to thyroid hormone (RTH) syndrome, accompanied by impaired corepressor release. Interestingly, the mutated residues are not within AF-2 or close to the corepressor binding surface, as described in Chapter IV, but are located in the loop region connecting helix 2 and helix 3 in the TR LBD. Therefore, defective corepressor release caused by these mutations will be an allosteric effect mediated by the TR LBD. In fact, the crystal structures of these point mutations fail to show significant alteration compared to the wild-type TR LBD, while part of the loop region is not resolved. Therefore, the conformational dynamics is probably of primary importance directing the behavior of mutant TRs. In this case, MD simulation would be the best tool to supplement the crystal structure studies with information on the altered conformational dynamics introduced by the mutations. As supported by the analysis shown in the rmsd and b-factor calculations of the *apo* TR LBD, the loop region connecting helix 2 and helix 3 in the TR LBD is highly mobile, and may participate in

ligand binding induced dynamics switch and coregulator recruitment. In the crystal structures of the holo NR LBDs, this region is missing probably also due to its high degree of flexibility. Because of this disordered conformation and flexibility, the homology modeling and MD simulation can be especially helpful in elucidating the functional correlated conformational dynamic properties of the TR LBDs with mutations in this region.

In order to learn more about TRs in their repression state, we have extended our studies to the interaction between TR and corepressors. From our homology models of TR and corepressor peptides, a refined picture of the complex interactions are revealed for the first time in Chapter IV. The interface between the corepressor peptide and the TRβLBD in the starting models mostly consists of hydrophobic residues that contribute to the binding. MD simulations of the complex models have stabilized the side chain positions and also revealed alteration in the backbone coordinates of the TR LBD and corepressor peptides. The gradual appearance of extra polar interactions between the TR LBD and corepressors through the MD simulation are indispensable force driving the reshaping of corepressor peptides. During the simulation, three NCoR peptides have revealed distinct conformations and dynamic characteristics. From their conformational and positional dynamics, TR LBDs have shown preference for the binding of the N2 and the N3 peptide in the NCoR, which is consistent with previous experimental results. Moreover, the distinct N3 domain present solely in NCoR but not in SMRT is probably the primary cause of different affinities of the two corepressors when binding to TR. MD simulations have also confirmed that the T3 binding is indeed driving the corepressor peptide dissociation from the TR LBD, even during the limited time within simulation. In

addition, assessment of binding energetics is still in progress, which will better quantify the different binding affinities with the different corepressor peptide, and also distinguish the corepressor affinity of the *holo* TR LBD from that of the *apo* TR LBD.

The multiple NR interaction domains presented in coregulator proteins are also associated with dimerization of NRs in various functional states. It has been proved by *in vitro* assays that presence of corepressor or coactivators can stabilize the homodimer assembly of TRs on TREs. The different binding affinities between a NR LBD and a certain NR ID in coregulators will also pose additional selectivity on the homo- or heterodimers that can be assembles on TREs. Three IDs presented in NCoR or the coactivator SRC1 may be able to help with conversion between homodimer and heterodimer of TRs by recruiting a RXR monomer to the C-terminal ID while the two N-terminal IDs maintain their interactions with two TR monomers. The ability to reconstruct different NR dimers will be distinctive depending on the various levels of binding affinities between NRs and corepressor IDs.

Furthermore, it is also possible to construct homology model of a corepressor peptide in complex with the mutant TR LBDs with impaired corepressor release. Such study will advance our understanding on altered ability of ligand-induced corepressor release conducted by mutations found in TRs of some patients with RTH as mentioned above. Similar structural dynamic studies can also be extended to study coactivator interactions with TR and to illustrate ligand binding effect on conformational dynamics in the context of coactivator peptides. With dynamics information from both the corepressor-bound and the coactivator-bound TRs, a more complete picture of the structure-correlated functional transitions of the TR will be revealed.

The studies in Chapter III and Chapter IV have broadened our perspectives on molecular mechanisms of TR in the constitutive states associated with transcription repression. The dynamic behavior of the TR LBD and the AF-2 motif seems particularly pertinent to the coregulator recruitment and transcriptional activity of TR, which is modulated by very fine-tuned and interrelated factors. Finally, understanding the mechanisms responsible for TR constitutive repression, in addition to the ligand-dependent activation, provides valuable information that could be applied to drug design. Especially, conformational characteristics observed in the TR LBD – corepressor complexes will better guide antagonist design to achieve increased or decreased affinity for certain corepressor ID. On the other hand, antagonists or agonists with desired potential and specificity in stabilizing TR-coregulator complexes and activating transcription of certain genes could be very useful to overcome various side effects associated with the current therapeutic agents. In addition to treating thyroid-related diseases, TR agonist are also potentially promising therapeutic molecules in preventing atherosclerosis, promoting metabolism as well as fat loss, and treating diabetes. However, a major problem in developing TR-targeting drugs is minimizing undesired effects on heart, bone and muscle, which are also influenced by thyroid hormones. The TRE selectivity of various TR-regulated genes, the differential tissue-selective uptake, as well as diverse levels of coregulators within the cells will lead to very distinctive physiological responses of synthetic TR ligands. This problem is further complicated by selective actions of TR ligands on two TR subtypes (TR$\alpha$ and TR$\beta$). Although the two subtypes have remarkably similarity in the ligand binding pocket, with only a single amino acid difference in the area, they do vary in the loop region flanking helix 2 of the

LBDs. Therefore, one may need to consider ligand binding effect on the conformational dynamics in this loop region, in addition to the AF-2 region, in order to achieve a desired therapeutic profile. At the molecular level, specificity of TR action also result from the selective recruitment of coregulators and thus a specific subset of transcription cofactors, for which an understanding of the molecular basis of the ligand-dependent conformational and dynamic change is of prime importance.