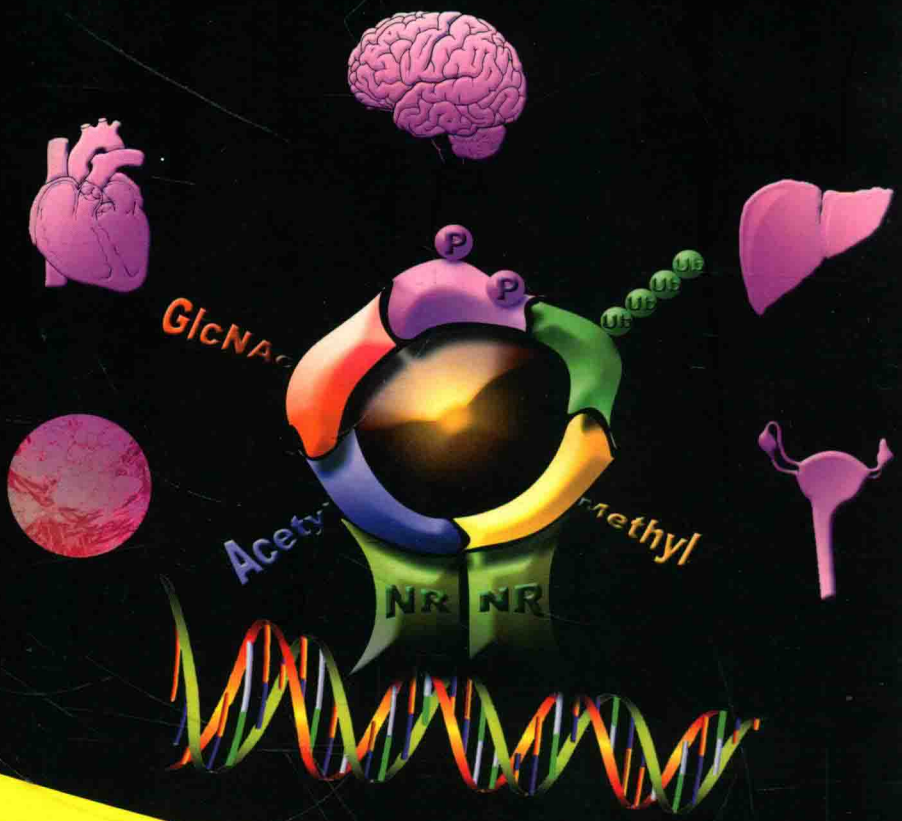
NR Coregulators and Human Diseases

Rakesh Kumar
Bert W. O'Malley



World Scientific

NR Coregulators and Human Diseases

Rakesh Kumar
The University of Texas M. D. Anderson Cancer Center, USA

Bert W. O'Malley
Baylor College of Medicine, USA

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

NUCLEAR RECEPTOR COREGULATORS AND HUMAN DISEASES

ISBN-13 978-981-270-536-5
ISBN-10 981-270-536-8
ISBN-13 978-981-270-537-2 (pbk)
ISBN-10 981-270-537-6 (pbk)

Typeset by Stallion Press
E-mail: enquiries@stallionpress.com

Printed in Singapore by Mainland Press Pte Ltd

NR Coregulators and Human Diseases

About the Editors

Rakesh Kumar

Professor Rakesh Kumar is currently the John G. and Marie Stella Kenedy Memorial Foundation Chair at the M.D. Anderson Cancer Center, where he is Professor of Molecular and Cellular Oncology and of Biochemistry and Molecular Biology. Professor Kumar is also an adjunct Professor at the Baylor College of Medicine. Dr. Kumar's research is directed at defining the mechanisms of estrogen receptor action with a special focus on subcellular localization and master chromatin modifiers. He has discovered the novel targets and functions of the MTA family of nuclear receptor coregulators, and opened new avenues for research. He also was the first to recognize a mechanistic role of PAK1 in cancer cell invasiveness and hormone action, discovering its physiologic substrates, and identifying the nuclear localization and functions of PAK1. He serves on the editorial boards of major cancer journals as well as on peer-review grant panels. Professor Kumar has received several awards and honors for his research excellence.

Bert W. O'Malley

Professor Bert W. O'Malley is currently the Tom Thompson and Distinguished Service Professor at the Baylor College of Medicine. His laboratory discovered that steroid·hormones and nuclear receptors act on genes to regulate the synthesis of messenger RNAs. He then went on to discover the "missing link coregulators" (coactivators/corepressors) that decipher all of the transcriptional instructions in the receptors. Coactivators are "master genes" that have immense regulatory influences on tissue development and physiology because they activate the subfamilies of genes in a manner designed to coordinately regulate cell physiology and metabolism. Of course, the dysfunctions in coactivators (or corepressors) lead to serious disease consequences but can serve as new markers for diagnosis and therapies. Professor O'Malley was a founder of the field of molecular endocrinology and is a member of the National Academy of Sciences and the Institute of Medicine. He has received many honorary doctorate degrees and numerous international awards and honors.

Preface

Nuclear receptor coregulators have experienced an explosive early development over their founding decade. The number of coactivators and corepressors has grown to over 300. The molecular biology of coactivators has informed us of a cadre of diverse and interesting mechanisms of transcriptional action, including chromatin modification and remodeling; initiation of transcription; elongation; alternative RNA splicing and finally, protein degradation. Over the past five years, researchers have demonstrated that coactivators have expanded their pleiotropic actions in multiple cell compartments where they shepherd functions of the numerous gene products required to regulate large physiologic processes such as growth, metabolism, and inflammation. The discovery of coactivators has also resulted in the production of over 90 mouse knockout models for the study of heritable diseases. Of the 300 currently discovered coregulators, about 165 already have been demonstrated to result in human pathologies and heritable dysfunction. They have been demonstrated to be causal in numerous instances of embryonic lethality; growth retardation; maturation; mental retardation; metabolic and endocrine disorders; inflammatory disorders; malignancies; reproductive; and cardiovascular abnormalities. The editors wish to thank the Nuclear Receptor Signaling Atlas Consortium for their continued support in the area of NR coregulator research. The editors also wish to point out that a great deal of primary and unpublished data in this field is summarized on the NURSA web site (www.NURSA.org).

Since these coregulator "master genes" are poised to pay big future dividends to the field of medicine, we felt it timely to compile the first book, written by the top experts in the field and dedicated to the physiologic and pathologic promises of coregulator research. The book will contribute to the foundation of Coregulator Biology as an emerging discipline in medical sciences.

Bert W. O'Malley, M.D. Rakesh Kumar, Ph.D.

Contents

Chapter 1

Nuclear Receptor Coregulators in Human Diseases

Rainer B. Lanz, David M. Lonard and Bert W. O'Malley

One of the major mechanisms through which eukaryotic cells respond to developmental and environmental signals is by altering their patterns of gene expression. By transmitting these signals to transcription factors using complex and tightly regulated processes, the coregulators function as essential signaling integrators for the coordinated control of broad transcriptional programs. Dysregulation of coregulator signaling circuitry has severe consequences for cell homeostasis and often contributes to pathogenesis. This review focuses on coregulators for nuclear receptors and their associations with human diseases. We first reiterate some general aspects of nuclear receptor-mediated transcription, and use selected examples of coregulator function to define the essential class of NR coregulators. We also provide factual evidence for the functional distinctiveness of different coregulators, and substantiate their involvement in numerous human pathologies. Overall, we conclude that the appreciation for coregulator biology is imperative for a fuller understanding of human diseases.

1.1 Nuclear Receptor-Mediated Transcription

The nuclear receptors constitute a large superfamily of structurally related response element-specific transcription factors that are highly versatile in both physiological function and molecular action. They consist primarily of a related group of DNA-binding proteins, some of which are activated by specific, small-molecule ligands (some "orphan"

nuclear receptors may not possess cognate ligands) and play important roles in developmental or endocrine biology. This is true for the steroid receptor subclass of nuclear receptors, which contains the receptors for androgens (AR), estrogens (ER), glucocorticoids (GR), progestins (PR) and mineralocorticoids (MR). They are activated by high-affinity steroids and bind as homodimers to palindromic DNA elements to transactivate target gene expression. They play essential roles in development (GR), reproduction (AR, ER, PR), stress and inflammation (GR), and glucogenesis (GR) and mineral metabolism (MR). The steroid receptors are thought to engage with their target genes primarily when bound to their respective cognate ligands. In contrast, another subgroup of nuclear receptors remain constitutively bound to DNA and actively repress basal transcription in the absence of ligands, but turn into potent transactivators upon binding to cognate high-affinity ligands (e.g. thyroid hormone receptor; TR, retinoic acid receptor; RAR).

Some nuclear receptors do not have a specific, high-affinity cognate ligand but are activated by different weakly binding endogenous cofactors, many of which are metabolic intermediates such as fatty acids, bile acids and/or sterols. These "metabolic sensors" include the peroxisome proliferator activated receptor (PPAR), the liver X receptor (LXR), the farnesol X receptor (FXR) and the hepatocyte nuclear factor 4 (HNF4) that form heteromeric complexes with the retinoid X receptor (RXR).[a] More recent research has shown that some receptors do not require ligands for activity because they already have a high constitutive activity (e.g. liver receptor homolog 1; LRH1). These nuclear receptors are constitutively active but respond to deactivating endogenous ligands (e.g. retinoic acid receptor-related orphan receptor β; RORβ, constitutive androstane receptor; CARβ), or do not appear to have a natural ligand but respond to xenobiotics (Pregnane X receptor: PXR, CARα). Other nuclear receptors may utilize metabolic intermediates such as fatty acids or phospholipids, not as signaling ligands, but rather as constitutive structural cofactors (e.g. HNF4, or Ultraspiracle; usp, the functional homologue of RXR in flies), while again other receptors lack an apparent ligand-binding cavity (e.g. nur-related protein 1; NURR1, chicken ovalbumin upstream promoter transcription factor; COUP-TF, or drosophila hormone receptor-like 38; DHR38).

[a] The 'metabolic sensor' receptors are discussed in more details later in the Metabolic Syndrome section.

While the tedious search for ligands established the importance of nuclear receptors in the eukaryotic cell and revealed the vital roles they play in the physiological state of entire cell systems, ligand binding alone was insufficient to explain how these molecules act to integrate the different signal-dependent cellular programs of transcriptional responses at the molecular level. The quantum leap in nuclear receptor biology came about with the discovery of an entirely different class of "ligands" — the nuclear receptor coregulators (NR coregulators), which were initially thought to be molecules that only influenced the transcriptional potency of nuclear receptors. The cloning and characterization of the first coactivator (SRC-1,[1]) and the first two corepressors (NCoR/RIP13,[2,3] and SMRT[4]) triggered a decade of intensive research in molecules we now collectively call transcriptional coregulators. It soon became clear that the additional information for conveying cell signals to nuclear receptor-mediated transcription must reside in the coregulator molecules and in their modes of function. This new cognition also was fueled by the realization that most transcription factors are not able to bind to cognate DNA response elements on their own unless the nucleosomal structure at promoter regions of target genes has been "prepared" for binding. This preparation is afforded by the enzymatic activities for the modifications of the components of the basic transcription machinery and the chromatin. It is now established that the transcription of nucleosomal DNA by polymerase II requires post-translational modification of histones to induce dynamic changes in the chromatin structure to either inhibit or facilitate transcription factor binding to genomic DNA. The modifying and corresponding counteracting enzymatic activities, histone acetyltransferases (HATs) and histone deacetylases (HDACs), and other nucleosome-modifying enzymatic activities, are usually recruited to promoters as multi-component complexes. The subunits of these complexes work together to collectively modify histones or other proteins.[b] For example, the NR coregulators SRC-1, CBP, p300, SRC-3 and others have been shown to possess HAT activity.[5–9]

[b] While hyperacetylation of histones is generally associated with transcriptionally permissive genes, the effects of acetylation of non-histone proteins varies between substrates, resulting in, for example, alterations in protein-protein interactions, sub-nuclear protein localization, or protein stability.

1.2 What are NR Coregulators?

The initial portrayal of a coactivator as a molecule that either bridges an activated nuclear receptor with one or more general transcription factors of the pre-initiation complex or modified histones soon had to be broadened. Molecules were characterized that did not directly bind to nuclear receptors, nor did they bind to general transcription factors but were still required for nuclear receptor-mediated transcription.[10] These "second generation" coregulators — sometimes referred to as cocoregulators in the literature — are currently still being discovered due to improved technologies and assays. The rate of coregulator discoveries has paralleled novel technologies from their initial discovery. First, the biochemical discovery of NR corepressors added a new functional class of molecules to the NR coregulators,[11,12] then through yeast two-hybrid expression screens which used distinct portions of nuclear receptors as "bait" to find the "classical" NR coregulators that directly interacted with nuclear receptors.[c] Later, RNAi technology allowed screening for "essential" molecules that are functionally necessary for regulated nuclear receptor transactivation. At the present, advanced mass spectrometry is being exploited to characterize the entire multi-protein transcription complexome which is required for integrating signal-dependent programs of NR-mediated transcription.[13]

1.2.1 *Structural determinants*

The task of defining transcriptional coregulators would be facilitated if these proteins had a distinguishing signature such as a sequence or structure motif that would mark their biological function. However, this is not the case. The InterPro database, which provides an integrated view of the commonly used protein signature databases, currently lists the motif termed "nuclear receptor coactivator" (IPR014920) as having 34 entries, all of which represent isoforms or fragments of the structurally related steroid receptor coactivators SRC-1, SRC-2, and SRC-3. Another motif, IPR009110 "(nuclear receptor coactivator, interlocking)" is found in 74 entries, but they all constitute either CREB-binding

[c] To avoid false positive results in the yeast expression system, the activation domain-1-containing N-terminal portion of a nuclear receptor commonly has been omitted for use as bait; most initial screens expressed the ligand binding domain only for finding interacting proteins.

protein (CBP), the related p300,[d] or the SRC coactivators, indicating that no simple protein signature exists for NR coregulators.

The nuclear receptors, on the other hand, share a high level of structural conservation in their central DNA binding domain, and, to a lesser extent, in the carboxyl terminal ligand binding domain (LBD). The intricacy of conserved α-helical structures in the LBD and their positional changes induced upon ligand binding provided the structural determinants for the identification of a conserved motif in proteins that bind to this domain. This so called "LXXLL-motif" (where L is leucine and X is any amino acid) or "NR box" was found to be necessary and sufficient to form a helical structure with which proteins bind to a "charge clamp" formed by the ligand-activated nuclear receptor.[14,15] Because proteins binding to this critical and hormone-sensitive region of the receptor inevitably alter the transactivation potency of nuclear receptors, the LXXLL motif was thought to serve as a signature motif for coregulators. This, however, is only partially valid, as many NR coregulators do not contain an NR box but still associate (through direct binding or tethered to another coregulator) with the nuclear receptors. Moreover, coregulator interactions also occur with other parts of the receptor that are not NR box-dependent. Thus, the LXXLL sequence is a signature motif for only a subset of coactivators or corepressors that bind in a ligand-dependent manner to the charge clamp of activated nuclear receptors.

Many coregulators indeed contain one or more LXXLL motifs. A compilation of nuclear receptor coregulators curated at NURSA (Nuclear Receptor Signaling Atlas; www.nursa.org) at the time of this writing includes 303 coregulator genes of which 149 have NR boxes in their amino acid sequence. While many (84) coregulators on this list have only one LXXLL motif, others have two (36) or several, with PELP1 (proline, glutamic acid and leucine rich protein 1) topping the list with eleven NR boxes.

Soon after the discovery of the SRC-1 coactivator, we realized that the same molecule could also be employed to repress gene function.[16] This unexpected "switch" in activity is realized now to be a frequent trait for both coactivators and coregulators. PELP1, separately cloned

[d] The SRC-family of coregulators as well as the related CBP and p300 are not specific for only the nuclear receptors, as they activate transcription mediated by non-nuclear receptor transcription factors too.

and termed MNAR (modulator of nongenomic activity of estrogen receptor), is a good example of a multi-tasking transcriptional coregulator that participates in both enhancement and repression of gene function. It participates in genomic and nongenomic responses to ER signaling.[17] It was postulated to play an important role in endocrine cancers as it increases E2-mediated cell proliferation and is involved in E2-mediated tumorigenesis and metastasis.[18] PELP1/MNAR coactivates ER- and RXR-mediated transcription but corepresses GR, NUR77, and non-nuclear receptor transcription factors such as AP-1 and NF-κB.[19] This interesting ambivalence towards acting as either positively as coactivator or negatively as corepressor is not uncommon for NR coregulators, nor is their ability to affect many transcription factors. The differential use of the many LXXLL helices may be a mechanism to modulate the efficacy of association with particular nuclear receptors,[20,21] and while a differential use of an activator and a repressor domain within a single coregulator is possible, it is uncommon, and the differential transcriptional activity is more readily achieved through interactions with other molecules carrying enzymatic activities tethered to the coregulator and brought into the transcriptional unit.

In the case of PELP1/MNAR, the N-terminal leucine-rich region was observed to interact with HDAC2 where it exhibits repressive activity when bound to GR. In addition, it was shown that the C-terminal acidic activation domain, which contains homopolymeric glutamic acids stretches, binds to the hypoacetylated histones H3 and H4 and prevents them from becoming substrates of histone acetyltransferases, promoting and maintaining the hypoacetylated state of histones at the target genomic site.[19] Hypoacetylated histones are associated with compact, transcriptionally inert, chromatin. When bound to ER, however, PELP1/MNAR recruits the p300/CBP histone acetyltransferases to the transcription unit,[17] thus reversing its role so as to maintain hyperacetylated histones. Acetylated histones are found in transcriptionally permissive, or "loose", chromatin. This example illustrates one of the hallmarks of NR coregulator function, namely the ability to provide or to recruit diverse enzymatic activities to nuclear receptors for the purpose of modulation of transcription. It is therefore not a surprise that the intense research being done on transactivation is revealing that a network of sequentially exchanged coregulator complexes contribute to transcription, rather than it being the work of individual proteins. These coregulator complexes have a heterogeneous

composition and are not discrete, separable entities but share many molecules and subcomplexes.

1.2.2 *Diversity in enzymatic activities and complexes*

To better understand the unexpectedly diverse roles that the NR coregulators play in human diseases, one must first recognize the value of their extensive involvement in diverse cellular processes. The elucidation of many complexes identified to serve the roles in NR-mediated target gene expression excoriates transcription control from a solely promoter-proximal process and establishes it as a cell system-wide central event that is linked to epigenetic programs. Because coregulators act as part of multi-protein complexes, carrying along enzymatic activities such as acetylating and deacetylating entities, methylases and demethylases, protein kinases and phosphatases, ubiquitin and SUMO ligase activities, pseudouridylases, and ATP-dependent chromatin remodeling activities, they exert stratified actions in cellular processes as diverse as transcription (including transcriptional elongation, RNA splicing and RNA transport), translation, DNA replication, and cell cycle control. The targets for these enzymatic activities are components of the coregulator complexes, nuclear receptors and other transcription factors, components of the basal transcription machinery, and the chromatin adjacent to the genes they regulate.

To complicate matters, the eukaryotic cell makes use of a combination of these enzymatic activities, connecting diverse biological processes with transcription control, which, paradoxically at a first glance, also includes the cessation of transcription via the ubiquitin proteasome pathway.[22,23] Moreover, new evidence is emerging that suggests another level of abstraction in the control and integration of cellular signals by the differential post-translational modification of the NR coregulators themselves. While the histone modification code may have evolved to define the transcriptional state of polymerase II-driven gene expression, a coregulator modification code will define complex composition and its target preferences. All this, and possibly much more, as we are still discovering new NR coregulators and are learning more about their functional diversity, has to be considered in addition to the temporal and spatial aspects of coregulator and nuclear receptor expression, as well as the availability and type of ligands for nuclear receptor activation.

The focus of this essay (and book) is to discuss the involvement of selected NR coregulators in human diseases. We therefore must refer to other reviews for a more detailed account of the biochemical and cellular aspects of these molecules.[24–29] However, we conclude with a more precise definition of "NR coregulator", a prelude of which shall include that this interesting class of molecules lacks a defining signature motif and is unlikely described with available gene ontologies. NR coregulators comprise a heterogeneous and functionally distinct group of molecules that associate with nuclear receptors by direct binding or in a complex with other molecules and integrate cellular signals by tethering varied enzymatic activities to the transcription unit for the purpose of precisely regulating nuclear receptor-mediated target gene expression. The entirety of this regulation is coactivation or corepression.

1.2.3 *Steroid receptor RNA activator*

Although one is often tempted to think of coregulators solely being proteins, such a generalization cannot be made. The reason for this is that one NR coregulator has been described to function as an RNA molecule.[30] Albeit unique for its kind, steroid RNA activator (SRA) displays many properties intrinsic to NR coregulators. This RNA was shown to coactivate some, but not all nuclear receptors *in vitro* and *in vivo* through associations with other NR coregulators in distinct complexes. For example, SRA, which differentially coactivates ERα and ERβ[31,32] and enhances the transactivation of a recombinant AR mutant lacking the amino-terminal activation domain only in the presence of SRC-1,[30] is in a human ER-AF1 complex along with the DEAD-box RNA helicases p68 and p72, SRC-1 and splicing factor SF3A subunit 1.[33,34] This ribonucleoprotein complex interacts via SF3A1 with the ligand-activated ERα that is phosphorylated at Ser118, and via SRA and components of the SRC-1/CBP histone acetyltransferease complex and their LXXLL motifs with the LBD of the receptor. Thus, the ER-AF1 ribonucleoprotein appears to function as a communication link between different pathways such as MAPK-mediated growth factor and estrogen signaling, nuclear receptor transcription via histone acetylation, and ER-phosphorylation-dependent RNA splicing. The physical interaction of RNA helicases p68 and p72 with SRA is required for these proteins to function as ERα-specific coactivators. This suggests that SRA has a central structural role in the RNA-associated proteins in the complex.