Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

William H. Hudson	Date

Two for one: The glucocorticoid receptor as a DNA- and RNA-binding protein

by

William H. Hudson
Doctor of Philosophy

Molecular and Systems Pharmacology
Graduate Division of Biological and Biomedical Sciences

Eric A. Ortlund, Ph.D.
Advisor

Graeme L. Conn, Ph.D.
Committee Member

Christine M. Dunham, Ph.D.
Committee Member

Yue Feng, Ph.D.
Committee Member

Haian Fu, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Two for one: The glucocorticoid receptor as a DNA- and RNA-binding protein

by

William H. Hudson
B.S., Georgia Institute of Technology, 2009

Advisor: Eric A. Ortlund, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Molecular and Systems Pharmacology in 2015.

Abstract

Two for one: The glucocorticoid receptor as a DNA- and RNA-binding protein
by William H. Hudson

Glucocorticoids are a class of small molecule hormones that control inflammation, metabolism, and responses to stress in vertebrates. Due to their potent anti-inflammatory effects, glucocorticoids are prescribed for a multitude of conditions, including asthma, arthritis, organ transplant rejection, cancer, and endocrine disorders. Glucocorticoids act by binding to the intracellular glucocorticoid receptor, which is expressed in nearly all human tissue and binds to genomic DNA to control the expression of hundreds of genes. Recently, the receptor has also been shown to bind RNA, including an emerging class of cellular molecules, long intergenic non-coding RNA. This work describes the molecular mechanisms by which the glucocorticoid receptor interacts with both DNA and RNA and subsequently controls transcription of target genes. This understanding is critical not only for designing improved glucocorticoid receptor agonists but also understanding the increasing number of proteins that bind both DNA and RNA to control cellular processes such as protein expression, cell division, and genomic repair. I show that proteins that bind both RNA and DNA are numerous - approximately 2% of the human proteome - and represent an efficient mechanism of controlling important cellular processes. I then demonstrate the specific mechanisms by which such a protein - the glucocorticoid receptor - recognizes both a long non-coding RNA as well as genomic DNA elements that mediate the repression of pro-inflammatory genes. In each case, the receptor shows a capability to bind a diverse set of nucleic acid sequences in a specific manner that is evolutionarily unique from the other, related steroid receptors. Collectively, this work expands the number of human proteins known to bind both DNA and RNA and provides an in-depth analysis of how one such protein can recognize multiple nucleic substrates.

Two for one: The glucocorticoid receptor as a DNA- and RNA-binding protein

by

William H. Hudson
B.S., Georgia Institute of Technology, 2009

Advisor: Eric A. Ortlund, Ph.D.

A dissertation submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Molecular and Systems Pharmacology in 2015.

Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Definition
ACTB	Actin-β
ADAR1	Adenosine deaminase 1
AR	Androgen receptor
AML1	Acute myeloid leukemia 1
AncCR	Ancestral corticosteroid receptor
AncGR	Ancestral glucocorticoid receptor
AncGR2	Ancestral glucocorticoid receptor 2
AncMR	Ancestral mineralocorticoid receptor
AncSR2	Ancestral steroid receptor 2
AncSR2	Ancestral steroid receptor 3
ANOVA	Analysis of variance
ATP	Adenosine triphosphate
bp	Base pairs
CCL2	Chemokine (C-C motif) ligand 2
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
CpdA	Compound A
CSD	Cold shock domain
CSDE1	CSD-containing protein E1
Csp	Cold shock protein
CT	Threshold cycle
DAVID	Database for Annotation, Visualization and Integrated Discovery
DBD	DNA binding domain
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DRBP	DNA/RNA binding protein
dsDNA	Double-stranded DNA
dsRBD	Double-stranded RNA binding domain
dsRNA	Double-stranded RNA
DTT	Dithiothreitol
EMSA	Electrophoretic mobility shift assay
ENO1	α-enolase
ER	Estrogen receptor
ERE	Estrogen response element

EPHA2	EPH receptor A2
FAM	Fluorescein amidite
FGFR3	Fibroblast growth factor receptor 3
FKBP5	FK506 binding protein 5
GABRB3	Gamma-aminobutyric acid A receptor, β-3
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GAR	Glycine/arginine rich (domain)
Gas5	Growth arrest specific 5
GPU	Graphics processor unit
GR	Glucocorticoid receptor
GRE	Glucocorticoid response element
GREM	Glucocorticoid response element mimic
GST	Glutathione-S-transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HOTAIR	HOX transcript antisense RNA
HSQC	Heteronuclear single quantum coherence
INS	Insulin
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IVT	*In vitro* transcription
K_d	Dissociation constant
K_i	Inhibitory constant
KH	K homology (domain)
LBD	Ligand binding domain
lincRNA	Long intergenic non-coding RNA
MDN1	Midasin AAA ATPase 1
MECP2	Methyl-CpG-binding protein 2
miRNA	MicroRNA
MME	Monomethyl ether
MR	Mineralocorticoid receptor
mRNA	Messenger RNA
Mya	Million years ago
NAPEPLD1	N-acyl phosphatidylethanolamine-specific phospholipase D
NCoR	Nuclear receptor co-repressor 1
NF90	Nuclear factor of activated T cells 90 kDa
NF-Y	Nuclear factor Y
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
nGRE	Negative glucocorticoid response element
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser effect spectroscopy
nt	nucleotide
NTP	Nucleoside triphsophate
ORC	origin recognition complex
PANDA	P21-associated ncRNA DNA-damage activated
PDB	Protein data bank
PEG	Polyethylene glycol

Pit-1	POU domain, class 1, transcription factor 1
PPAR	Peroxisome proliferator-activated receptor
PR	Progesterone receptor
pre-miRNA	Precursor microRNA
pri-miRNA	Primary microRNA
RANBP2	RAN-binding protein 2
RAR	Retinoic acid receptor
RBD	RNA binding domain
rDNA	Ribosomal DNA
RBM3	RNA-binding motif 3
RMS	Root mean square
RMST	rhabdomyosarcoma 2-associated transcript
RNA	Ribonucleic acid
RNA-IP	RNA-immunoprecipitation
RNP	Ribonucleoprotein (domain)
RNP-1	Ribonucleoprotein 1
RRM	RNA Recognition Motif
RsmE	Ribosomal RNA small subunit methyltransferase E
RT-PCR	Reverse transcriptase polymerase chain reaction
RXR	Retinoid X receptor
SELEX	Systematic evolution of ligands by exponential enrichment
SF-1	Steroidogenic-factor 1
snoRNA	Small nucleolar RNA
SMRT	Silencing mediator for retinoid or thyroid-hormone receptors (or NCoR2)
SNP	Single nucleotide polymorphism
SR	Steroid receptor
SRA	Steroid receptor RNA activator 1
SRC1	Steroid receptor coactivator 1
ssDNA	Single-stranded DNA
ssRNA	Single-stranded RNA
SV2C	Synaptic vesicle glycoprotein 2c
TAF15	TBP-associated factor 15
TBE	Tris/borate/EDTA (buffer)
TDP43	TAR DNA-binding protein 43
TERRA	Telomere repeat-encoding RNA
TFIIIA	Transcription factor IIIA
T_m	Melting temperature
TOP	Terminal oligopyrimidine tract
TR	Thyroid receptor
TRF2	Telomeric Repeat binding Factor-2
tRNA	Transfer RNA
TSLP	Thymic stromal lymphopoetin
UTR	Untranslated region
UV	Ultraviolet
WCSP1	Wheat cold shock protein 1

WT	Wild-type
Xist	X-inactive-specific transcript
YB-1	Y box 1

Chapter 1

Introduction: Proteins that bind DNA and RNA

1.1 Abstract

Proteins that bind both DNA and RNA typify the ability of a single gene product to perform multiple functions. Such DNA- and RNA-binding proteins (DRBPs) have unique functional characteristics that stem from their specific structural features; these developed early in evolution and are widely conserved. Proteins that bind RNA have typically been considered as functionally distinct from proteins that bind DNA and studied independently. This practice is becoming outdated, in part owing to the discovery of long non-coding RNAs (lincRNAs) that target DNA-binding proteins. The majority of this dissertation will describe how a particular DRBP - the glucocorticoid receptor - interacts with both DNA and RNA. This opening chapter focuses on enumerating and describing the function of human DRBPs, finding that DRBPs regulate many cellular processes, including transcription, translation, gene silencing, microRNA biogenesis and telomere maintenance.

This chapter is adapted from the manuscript:
Hudson WH and Ortlund EA. The structure, function, and evolution of proteins that bind DNA and RNA. *Nat Rev Mol Cell Biol.* 2014 Nov;15(11):749-60.

1.2 Introduction

Proteins that bind DNA or RNA are often considered and studied independently of one another. For example, transcription factors are usually modeled relatively simply: they bind to genomic promoters and control target gene expression by activating or repressing RNA polymerases. Following transcription, RNA-binding proteins modulate protein expression by regulating the stability and translation of mRNAs. However, the consideration of DNA- and RNA-binding functions within proteins as separate entities is becoming outdated. The unappreciated dual DNA- and RNA-binding capacity of a growing body of proteins plays a key part in modulating gene expression, cell survival and homeostasis. Recent studies have demonstrated that many transcription factors are capable of binding diverse types of RNA, which enables them to bind to the mRNA products of transcription to regulate their turnover and to integrate other signals, such as responses to stress.[1–7] Additionally, the prevalence and emerging functions of long intergenic non-coding RNAs (lincRNAs) have revealed that non-coding RNAs target many types of proteins through direct interactions.[2,8–11]

In this analysis, we enumerate these DNA- and RNA-binding proteins (DRBPs) and describe their functions, structures and evolution. We first broadly discuss the prevalence of DRBPs within the human genome. We highlight known functions of DRBPs with specific examples of how simultaneous and serial RNA and DNA interaction allows better gene targeting, finer control of gene expression and integration of metabolic state or stress to modulate protein activity. We discuss the structural features of DRBPs that enable dual nucleic acid specificity, focusing on the limited number of solved structures that allow direct comparison of a DRBP complexed with either DNA or RNA. Finally, we discuss the evolution of dual DNA- and RNA-binding domains within DRBPs, including ancient domains, for which dual DNA and RNA binding conferred a selective advantage, and more modern domains, which have recently been targeted by rapidly evolving lincRNAs.

1.3 Defining DRBPs

Defining the subset of human proteins that bind both DNA and RNA is a difficult task. Using gene ontology searches, only 64 human protein-coding genes in the QuickGO gene ontology database[12] (European Bioinformatics Institute) are identified as having direct and specific experimental evidence for both RNA binding (GO:0003723) and DNA binding (GO:0003677) (Figure 1.1a). The PROTEOME database (BioBase) returns 122 such proteins, although direct evidence is lacking for many.

An alternative approach involves combining evidence from studies that have separately attempted to catalog all human proteins that bind DNA or RNA (Figure 1.1b). A study using protein microarrays and bioinformatics approaches identified over 4,000 human proteins that directly interact with double-stranded DNA (dsDNA) *in vitro*.[13] Gene ontology analysis of these proteins reveals that the term RNA binding is highly enriched for ($p < 1 \times 10^{-40}$), indicating that RNA binding may be a common feature of DNA binding proteins (Figure 1.1c). Among these dsDNA-binding proteins, the ontology term 'dsRNA binding' is much more represented than ssRNA binding.

Another study used a crosslinking- and mass spectrometry-based approach to identify 860 mRNA-binding proteins from HeLa cells termed the mRNA interactome.[14] Functional analysis of these proteins again indicates that dual nucleic acid binding is a widespread phenomenon (Figure 1.1d), as they are significantly enriched for both ssDNA binding and dsDNA binding ($P = 8.9 \times 10^{-21}$ and $P = 5.5 \times 10^{-8}$, respectively). Notably, of the 860 proteins identified as mRNA binding, 407 (47.3%) were independently characterized as dsDNA binding in reference 13. Together, the two studies indicate that DRBPs are widespread, perhaps constituting 2% of the human proteome (407 of 20,300 proteins, Figure 1.1b). This number would probably increase if the studies included proteins that are expressed in other cell types, proteins that require ligand binding-dependent signals for nucleic acid binding, or proteins that bind other types of DNA or RNA.

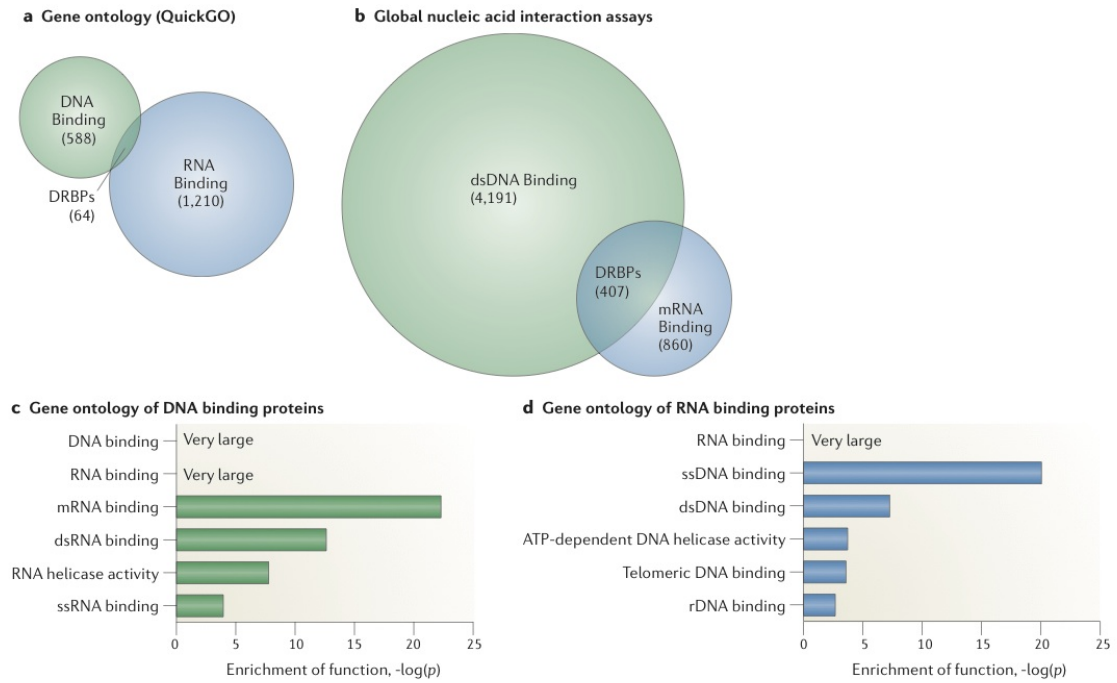


Figure 1.1: Defining human DRBPs. (a) Venn diagram shows DNA-binding proteins and RNA-binding proteins in the QuickGO database supported by low-throughput experimental evidence (as of July 2014).[12] The overlap of these two sets represents human DNA- and RNA-binding proteins (DRBPs), which consists of 64 proteins. (b) Venn diagram shows DNA-binding proteins and RNA-binding proteins identified in high-throughput studies defining the human mRNA and double-stranded DNA (dsDNA) interactomes.[13,14] There are 407 proteins found in both studies, indicating that they may bind both mRNA and dsDNA. In parts (a) and (b), circles are drawn to scale. (c) Molecular function gene ontology analysis reveals that RNA binding is a potentially major function of the dsDNA-binding proteins identified in reference 13. (d) Gene ontology analysis reveals that DNA binding is potentially a major function of the mRNA-binding proteins identified in reference 14. In parts (c) and (d), only selected molecular function attributes are shown for brevity. P values in parts (c) and (d) indicate the probability that the over-representation of the stated ontology term in the selected 407 genes compared with all human genes is due to chance. These were calculated in the TRANSFAC + PROTEOME database (BioBase) using the hypergeometric distribution; very large indicates a P value of $<1 \times 10^{-40}$ ($-\log(P) > 40$). rDNA, ribosomal DNA; ssDNA, single-stranded DNA.

We note that many of the proteins identified in references 13 and 14 as DNA and/or RNA binding lack corroborating evidence from other studies, and these findings should thus be interpreted with caution. For example, many identified proteins, such as polymerase subunits, may bind nucleic acid-bound proteins without binding DNA and/or RNA directly. Additionally, many proteins that bind DNA or RNA *in vitro* may not bind them *in vivo*. However, the two studies provide a reasonable estimate of potential human DRBPs owing to their wide coverage of the human proteome, and we discuss below examples in which the demonstration of protein-nucleic acid binding *in vitro* has preceded the discovery of such binding *in vivo*, sometimes by decades.

In Table 1.1, we provide a detailed list of 149 human DRBPs, with comments on their nucleic acid-binding properties, structures and functions. These proteins were selected based on experimental evidence demonstrating their ability to bind directly to both DNA and RNA, generally obtained from studies using more traditional experimental approaches than the high-throughput studies[13,14] discussed above. Although many of the proteins in Table 1.1 have only been shown to bind DNA and/or RNA *in vitro*, the remainder of this analysis focuses on selected human DRBPs with known cellular roles.

1.4 Functions of DRBPs

We carried out gene ontology and domain enrichment analyses (Figure 1.2) to illuminate the main biological functions of our list of human DRBPs (see Table 1.1). The gene ontology analysis revealed expected biological processes, such as transcriptional regulation, mRNA processing and DNA replication. However, several surprising functions are also implicated, including the DNA-damage response, apoptosis and responses to extreme temperatures (Figure 1.2a).

Ultimately, DRBP functions are governed by their inherent structural and biochemical properties. One can envision DRBPs capitalizing on both RNA and DNA binding in numerous ways; for example, a transcription factor that binds DNA and RNA may interact orthogonally with RNAs that compete with DNA binding to repress transcription, or simultaneously with a promoter and an RNA co-activator to upregulate transcription. The following section focuses on DRBPs that bind DNA and RNA competitively (Figure 1.3a).

Table 1.1: Human proteins demonstrated to bind DNA and RNA. DRBPs are alphabetically ordered by their official gene symbol, with other common names given in parentheses. Some of these proteins have only been shown to bind DNA and/or RNA *in vitro*; list of nucleic acid and general protein functions may not be comprehensive. Structural domains were retrieved from Uniprot[15] unless otherwise noted.

Protein/ gene	DNA bound by protein	RNA bound by protein	Structural domains	General protein function(s)	Notes
ADAR	Z-DNA[16]	dsRNA (secondary structure-specific[17,18]) Z-RNA[19] miRNA[20]	Z-DNA and Z-RNA: Zα domain dsRNA: Deaminase domain	pro- and anti-viral RNA editing[21] mRNA editing[21]	Crystal structures exist of ADAR-DNA and RNA complexes[19,22] ADAR has some sequence specificity for RNA editing substrates[23]
AICDA (AID)	ssDNA[24]	mRNA[25]	C-terminus required for RNA binding[25]	Antibody diversification and class switching[26,27]	Binds DNA bubbles Contains cytosine deaminase activity[24]
ANXA2 (Annexin A2)	dsDNA[28]	mRNA (ex: c-myc)[29]		Endo/exocytosis[30] Phagocytosis[30] Anticoagulation[30]	Prefers G-rich RNA sequences[29] RNA binding is Ca^{2+} dependent[29] Other annexins bind B- and Z-DNA in Ca^{2+} dependent manner[31]
APTX (Aprataxin)	dsDNA[32]	A-form RNA at DNA/RNA junctions[33] Hairpin loops[32]	Deadenylase core: C2H2 Zn finger and histidine triad[33]	DNA Repair[34] Abortive DNA ligation reversal[34]	Crystal structure exists of APTX bound to an RNA-DNA complex[33]
BAZ2A (TIP5)	dsDNA[35]	ssRNA[36] Noncoding pRNA[37]	AT hooks, TAM domain, and PHD/bromodomain[35,38]	Chromatin organization[35]	
BCLAF1 (Btf)	dsDNA[39] (ex: TP53 promoter[40])	mRNA[14,41]	bZIP[39]	mRNA trafficking[42]	
BRCA1	Branched DNA and dsDNA[43]	miRNA[44]	Not fully characterized; AAs 230-534[45]	DNA Repair[46] Apoptosis[46]	Same domain used for DNA and RNA binding[44]
C1D	dsDNA[47]	G-rich RNAs[48] tRNAs[48]		Transcriptional repression[49] DNA Repair[50]	
CARHSP1 (CRHSP-24)	ssDNA[51]	3' UTRs[52]	Cold shock domain	mRNA stability[52] Transcriptional regulation[53]	Binds T-rich DNA[51]
CBX5 (HP1)	dsDNA[54]	Telomere-repeat-encoding RNA[55] heterochromatic mRNAs[56]	Chromodomain and hinge[57]	Gene silencing[58] Transcriptional regulation[58]	DNA/RNA binding conserved in amoeba and yeast[56,59] RNA binding is isoform-specific[55,57] DNA binding is sequence-independent[54]

Protein/ gene	DNA bound by protein	RNA bound by protein	Structural domains	General protein function(s)	Notes
CBX7	dsDNA (ex: CCNE1 promoter)[60]	lincRNA ANRIL[61]	Chromodomain	Transcriptional silencing and chromatin modification[62]	Member of polycomb repressive complex Reports on DNA binding ability conflict[60, 61] Chromodomain used for both DNA and RNA binding[60, 61]
CENPC	dsDNA at centromeres[63]	ssRNA51 Centromeric RNA[64]		Kinetochore assembly[65]	ssRNA binding stabilizes maize CENP-C binding to DNA[64] CENP-A and CENP-W may share similar properties[66, 67] Crystal structure exists of the CENPC-nucleosome complex[68]
CEBPZ (NOC1)	dsDNA[69]	mRNA[14]		Transcriptional regulation[69]	Binds CCAAT DNA elements[69]
CNBP	ssDNA (sterol regulatory elements)[70]	5' UTRs[71]	Seven CCHC Zn fingers and RGG box	Development[72] Transcriptional regulation[72] Cell death and proliferation[72]	Prefers G-rich sequence as DNA or RNA[73] RNA/DNA preference depends on dimerization status[73]
CSDE1 (UNR)	ssDNA and dsDNA[74]	ssRNA[74] mRNA[14]	Five cold shock domains[74]	RNA turnover[75] Apoptosis[76]	Higher affinity for ssDNA than dsDNA[74]
CSTF2	dsDNA[13, 77]	3 UTRs (ex: papillomavirus)[78]	RRM	Cell proliferation[79] RNA processing[80]	DNA binding shown via ChIP[77] and protein microarray;[13] direct binding not definitively demonstrated
DDX3X (DDX3)	dsDNA[81, 82]	5 UTRs (ex: HIV-1)[83]	DEAD box	mRNA metabolism[84] Translational regulation[83] Transcriptional repression[82]	DDX1 and DDX5 also bind DNA[85, 86] and RNA[14, 87, 88] Contains DNA and RNA helicase activity[81]
DEK	dsDNA[89] Supercoiled and four-way junction DNA[90]	mRNA[91]	Winged helix[89]	RNA splicing[91] DNA damage response[92]	DNA binding is not sequence specific[90]
DHX36	DNA quadruplex structures[93, 94]	RNA quadruplex structures[94] mRNA[14]	DEAH box, ATP-dependent helicase domain	mRNA metabolism[95] Transcriptional regulation[96]	RNA binds via the N-terminus[97] DHX9 also binds DNA[98] and RNA[99] A crystal structure of DHX9 in complex with DNA has been solved[100] Resolves G4 DNA structures[93]
DLX2	dsDNA (ex: Wnt enhancer)[101, 102]	Evf-2 ncRNA[101] (ssRNA)	Homeodomain	Embryonic development[103] Lymphocyte maturation[102] Transcriptional regulation[104]	
DNMT1	Unmethylated CpG DNA[105]	ecCEBPA RNA[106]	CXXC zinc finger[105, 107] Methyltransferase catalytic domain	DNA methylation[108] Embryonic development	Higher affinity for RNA than DNA[106] RNA binding is secondary-structure dependent[106] TRDMT1 (DNMT2), similar in sequence to DNA methyltransferases, methylates tRNA[109] DNMT3A also binds DNA and RNA, potentially in a competitive manner[110] Crystal structure exists of DNMT1 in complex with DNA9[111]

Protein/ gene	DNA bound by protein	RNA bound by protein	Structural domains	General protein function(s)	Notes
DNT-TIP2	ssDNA and dsDNA[112]	mRNA[14,41]	Helix-turn-helix and AT hook motifs101	Transcriptional regulation[113] V(D)J recombination[112]	DNA binding is sequence specific[113] Evidence for RNA binding comes from high-throughput crosslinking studies
DPPA3 (Stella)	dsDNA[114]	rRNA[114]		DNA methylation[114] Development[115]	DNA and RNA binding may be nonspecific[114]
DUSP11	ssDNA[116]	ssRNA[116]	PTPase domain	RNA 5 triphosphate and protein phosphatase[117]	Higher affinity for RNA than DNA[116] May bind the lncRNA mrhl[87]
ENO1 (α-enolase)	dsDNA (ex: c-myc promoter)[118]	CUG triplets[119] mRNA[14]		Glycolysis[120] Transcriptional repression[121]	MBP-1, which binds the c-myc promoter, arises from a later start codon of the ENO1 transcript[121]
ERCC6 (CSB)	dsDNA[122,123]	Forked RNA duplexes[124]	DEGH box ATP-dependent helicase domain	DNA Repair[125] Chromatin remodeling[126]	Contains a DNA- (but not RNA-) dependent ATPase activity[124]
ESR1 (Estrogen receptor)	dsDNA (estrogen response elements)[127]	G-rich sequences[128]	Two C4 Zinc fingers	Female reproduction[129] Hormone responses[130] Transcriptional regulation[130]	RNA and DNA binding are competitive[128] ESR1-DNA crystal structure has been solved[127]
FUBP1 (FBP)	ssDNA[131]	mRNA (5 and 3 UTRs; ex: p27Kip1)[132,133]	Novel DNA binding domain[131] 4 KH domains	Transcriptional regulation[134] Translational regulation[134] Apoptosis and cell proliferation[134]	FUBP2 (KHSRP) and FUBP3 also bind DNA[135] and RNA[134,136] FUBP1-ssDNA crystal structure has been solved[137] FUBP2 (KHSRP) RNA solution structure has been solved[138]
FUS (TLS)	ssDNA and dsDNA[139]	ncRNAs[140] mRNA[14,141]	RRM Zinc finger (RanBP2-type)	Transcriptional regulation[142] Neurodegeneration[143]	Promotes D-loop formation and DNA annealing[139] Similar to TARDBP (TDP-43)[143] Binds G-rich RNA[144]
G3BP1	dsDNA[145]	mRNA[14] dsRNA[145]	RRM NTF2 domain	Cell motility[146] RNA turnover[147]	Contains ATP-dependent DNA and RNA helicase activity[145]
GAPDH	ssDNA and dsDNA[148] Telomeric ssDNA[148]	tRNA[149] mRNA[150]	NAD+ binding domain (Rossmann fold)[148,150]	Glycolysis[151] RNA metabolism[152] Transcriptional regulation[152,153]	Higher affinity for ssDNA than dsDNA[148,154] Both DNA[148] and RNA[150] compete for NAD+ (cofactor) binding Binds AU-rich RNA[150]
GTF3A (TFIIIA)	dsDNA[155]	5S rRNA[156]	Nine C2H2 zinc fingers	Transcriptional control[157]	Distinct zinc fingers contribute to RNA and DNA specificity in *Xenopus* GTF3A[158] Structures of *Xenopus* protein in complex with RNA and DNA have been solved[159,160]
HBP1	dsDNA[161]	A- and U- rich sequences[162]	RNA: AXH domain[162] DNA: HMG box[161]	Transcriptional repression[163] Senescence[164]	Structure of an HMG box DNA complex has been solved (see HMGB1) DNA binding and subsequent transcriptional regulation is sequence specific[161,165]
HDGF	dsDNA[166]	mRNA[14]	PWWP domain[166]	Angiogenesis[167] Transcriptional regulation[168]	RNA binding may be indirect[168,169]
HMGB1 (HMG-1)	dsDNA[170] damaged DNA[170]	Branched RNA[171]	Two HMG boxes	Neuronal development[172] Innate immunity[173]	HMGB2 may share similar properties[14] HMGB1-DNA crystal structure has been solved[174]

Protein/ gene	DNA bound by protein	RNA bound by protein	Structural domains	General protein function(s)	Notes
HNRNPA1 (hnRNP A1; UP1)	telomeric DNA[175] dsDNA[176]	Telomerase RNA[175] mRNA[177] miRNA[178]	Two RRMs[175]	RNA processing[179] Transcriptional regulation[179]	May bind RNA and DNA simultaneously[175] hnRNP A2/B1 shows similar properties[180–182] Crystal structure in complex with DNA has been solved[183] RNA binding has been proposed to be both sequence- specific[177] and non-specific;[184] DNA binding can be sequence specific[183]
HNRNPC (hnRNP C)	ssDNA[185]	5 UTRs and 3 UTRs (ex: uPAR)[186,187]	One RRM[188]	mRNA metabolism[186,187] DNA damage and stress responses[189,190]	Binds U-rich sequences[191,192]
HNRNPD (hnRNP D; AUF1)	dsDNA[193] G-quadruplexes[194]	ssRNA[195] mRNA[196]	Two RRMs[188]	mRNA metabolism[197]	Binds AU-rich elements[197] and G-rich elements[198] HNRNPDL (hnRNP D-like) shows similar properties[199] DNA-hnRNP D NMR structure has been solved[200] Interacts with nucleolin (see below)[201]
HNRNPK (hnRNP K)	ssDNA and dsDNA (ex: Myc promoter[202])	mRNA[203] lincRNA-p21[204]	Three KH domains[188]	Transcriptional regulation[205] RNA processing[205]	DNA/RNA binding preference differs among the three KH domains[206] Prefers C-rich RNA and DNA[191,207] Structure with DNA has been solved[208]
HNRNPL (hnRNP L)	ssDNA[209]	3 UTR (ex: VEGF)[209]	Four RRMs[188]	mRNA processing[210] RNA splicing[211]	Prefers CA-rich rich elements and CA clusters[211] HNRNPLL (HNRNPL-like) also binds DNA[212] and RNA[213]
HNRNPU (hnRNP U)	ssDNA and dsDNA[214]	3 UTRs (ex:TNF-α)[215] lincRNAs (ex: Xist,[216] Firre[217]) snRNAs[218]	RGG box[219]	RNA splicing[218] X inactivation[220] Transcriptional regulation[221]	Likely binds RNA, DNA via separate domains[222] HNRNPUL2 (hnRNP U like-2) may show similar properties[14,223]
IGHMBP2 (SMUPB-2)	dsDNA[224]	dsRNA[224] tRNA[225]	RNA/DNA helicase domain[225] R3H domain[225]	Translational control[225] Transcriptional control[226]	Has 5'→3 DNA and RNA helicase activity[224] May contain an AN1-like zinc finger[227] Structure with RNA has been solved[228] DNA/RNA binding is likely not sequence-specific[224]
ILF3 (NF90)	ssDNA[229] and dsDNA (ex: IL-2 promoter[4])	dsRNA[230,231] mRNA27	DZF domain Two DRBM domains	miRNA biogenesis221 mRNA metabolism[232]	ILF2 shows similar properties[229] Represses AU-rich motif-containing mRNAs[233]
KHDRBS1 (Sam68)	ssDNA and dsDNA[234]	mRNA[234,235]	STAR domain (contains a KH domain)[236]	RNA splicing[236] Signal transduction[236]	Prefers U-rich RNA[237] KHDRBS3 also binds DNA and RNA[13,238] RNA binding ability may depend on post-translational modifications[239]
KIN (KIN17)	dsDNA[240]	G-rich RNA[241]	RNA: C-terminal SH3- like domains[242] DNA: C2H2 Zn finger	DNA replication[243] DNA damage response[244]	

Protein/ gene	DNA bound by protein	RNA bound by protein	Structural domains	General protein function(s)	Notes
Ku (*Saccharomyces*)	dsDNA[245] Damaged DNA[246]	TLC1 telomerase RNA[247] mRNA[14]	Ku70/Ku80 ring around DNA[248]	Non-homologous end-joining and DNA repair[249] Telomere maintenance[249]	DNA, RNA binding mutually exclusive[250] Ku is a dimer of the Ku70/Ku80 subunits Human homologs XRCC5 and XRCC6 also bind DNA and RNA[14] Ku-DNA structure has been solved[248]
LDHA (Lactate dehydrogenase)	ssDNA[251,252]	AU-rich elements[253]	Rossmann fold (NAD+ binding region)[253]	Cancer cell growth[254] Anaerobic glycolysis[255] Transcriptional regulation[153]	Higher affinity for ssDNA than dsDNA[154] RNA binding is competitive with cofactor binding[253]
LIN28A	ssDNA[256]	miRNA[257] mRNA (ex: Oct4)[258]	Cold shock domain Two CCHC zinc fingers	miRNA biogenesis[259] Translational control[259] Nutrient response[260]	Both CSD and Zinc fingers are important for miRNA binding[261] LIN28A-RNA structure has been solved[262]
LONP1	ssDNA[263] G-quartet DNA[264]	ssRNA[263]	LON domain	Proteolysis[265] mtDNA damage response[266]	Binds GT- and GU-rich sequences[263] The *E. coli* homolog binds dsDNA[267] Protein substrate binding increases its affinity for DNA[268] DNA and RNA binding is inhibited by ATP , but ATPase domain is not required for nucleic acid binding[268]
LRPPRC (LRP130)	ssDNA[269]	polyadenylated RNAs[270] mtRNA[270]	Contains pentatricopeptide repeats[271]	mRNA metabolism[272,273] Activation of mitochondrial transcription[274]	DNA and RNA bind via the same domain[271] Prefers G-, U-, C-rich RNA[271]
LRRFIP1	dsDNA[275]	dsRNA[276] ncRNA[277]	Residues 429-528 (DNA binding)[275]	Viral response[278] Transcriptional regulation[279]	Prefers GC-rich RNA[276] Higher affinity for dsRNA than dsDNA[276]
MECP2	Four-way junction DNA[280] dsDNA[281]	mRNA[14,282]	AT hooks[283]	Synapse function[284] Transcriptional regulation[285]	Binds methyl-cytosine containing DNA[281] Structure with DNA has been solved[286]
NABP1	ssDNA[287]	ssRNA[287]	OB fold[287]	DNA Repair[288]	Prefers C- and T-rich DNA sequences[287]
NACA (α-NAC)	dsDNA[289,290] ssDNA282	rRNA[290] tRNA[290]	NAC domain[291]	Transcriptional regulation[292] Chaperone of nascent polypeptides[293]	Binds RNA with higher affinity than DNA[291] Dimerization partners affect DNA binding ability[291] Binds RNA with higher affinity than DNA[291] Activates transcription in a sequence-specific manner[294]
NAT10 (ALP)	dsDNA (ex: TERT promoter)[295] rDNA[296]	snoRNAs[296]	N-acetyltransferase domain	rRNA processing[296] Cell division[297]	Contains histone acetylase activity[295]
NCL (Nucleolin)	dsDNA (ex: CD34 promoter)[298]	mRNA (ex: APP 3 UTR)[299]	Four RRMs and RGG domain[300]	RNA metabolism[301] DNA replication[301] Ribosome assembly[301]	Contains DNA and RNA helicase activity[302] Binds G-rich RNA sequences[300] Solution structure of Nucleolin-RNA complex has been solved[303]
NFKB1 (and RELA)	dsDNA (ex:IFN-β)[304]	RNA aptamers[10,305] Lethe RNA[9]	Rel homology domain[306]	Immune signaling[306]	NF-κB is a hetero- or homodimer of Rel-family proteins, such as NFKB1 and RELA[306] Structure of p50 subunit of NF-κB with DNA and RNA have been solved[307,308]

Protein/ gene	DNA bound by protein	RNA bound by protein	Structural domains	General protein function(s)	Notes
NFX1 (NF-X1)	dsDNA[309]	mRNA27[14,310]	Eight zinc fingers R3H domain	Transcriptional regulation[309,311] Ubiquitination[312]	Composition of domains depends on alternative splicing[310]
NFYA (NF-YA)	dsDNA (CCAAT boxes)[313]	lincRNA PANDA[8]	Binds along with NF-YB and NF-YC, which contain histone-like domains[313]	DNA damage and apoptosis[314]	lincRNA PANDA acts reduces NF-YA occupancy on genomic DNA; RNA and DNA binding may therefore be competitive[8] Crystal structure in complex with DNA has been solved[313] DNA recognition occurs via the minor groove[313,315]
NKRF (NRF)	dsDNA[316]	mRNA (ex: IFN-γ)[317]		Transcriptional control[318] Regulation of inflammation[319]	
NONO (p54nrb)	dsDNA[320]	ssRNA[320] pre-mRNA[320]	2 RRMs and helix-turn helix motif[321]	RNA processing[322] Paraspeckle assembly and maintenance[323] Transcriptional regulation[324]	RNA, DNA bind via the same domain[320] Binds inosine-containing RNAs[321]
NR0B1 (DAX1)	ssDNA (hairpins)[325]	mRNA[326] lincRNA SRA[6]	N-terminal repeats and nuclear receptor ligand binding domain[6]	Transcriptional repression[325] Endocrine function[327]	Lacks the zinc finger DNA-binding domain found in other nuclear receptors (see estrogen receptor, glucocorticoid receptor, RAR, SF-1, and the thyroid receptor)
NR3C1 (Glucocorticoid receptor)	dsDNA (glucocorticoid response elements)[328]	mRNA[1,329] lincRNA Gas5[2] tRNA[330]	Two C4 Zinc fingers[331]	Hormone responses[332] Transcriptional regulation[332] Immunity, metabolism[332]	Crystal structures of monomeric and dimeric GR DNA binding domain in complex have been solved[333,334] AR, PGR and NR3C2 (androgen, progesterone, and mineralocorticoid receptors) also bind Gas5[335] DNA and RNA binding are sequence specific[334,335] and competitive[2]
NR5A1 (SF-1)	dsDNA[336] (ex: LHβ promoter[337])	lincRNA SRA[6]	CCCC zinc fingers and Ftz-F1 box[6]	Transcriptional control[338] Steroidogenesis[338] Development[338]	Solution NMR structure exists in complex with DNA[339] Binds DNA as a monomer, unlike many other members of the nuclear receptor family[340]
NSUN2	Hemimethylated DNA[341]	Ribosomal RNA[342] tRNA[343] ncRNA341 mRNA[344]		RNA methylation[343] Apoptosis and cell division[342] Translational control[345]	Can methylate DNA and RNA[341,343]
NUFIP1	Low-affinity, non-specific binding[346]	Poly(G) RNA[347]	C2H2 Zinc finger[347]	Transcriptional control[346] snoRNA processing[348]	Data for DNA binding activity not shown[346] Zinc finger not required for RNA binding[346]
PARP1	ssDNA and dsDNA[349] damaged DNA[350]	ssRNA[349] mRNA[14]	Two zinc fingers DsDB domain[349]	Genomic maintenance[351] DNA repair[352]	Contains multiple DNA binding sites[349] Contains ADP-ribosyltransferase activity[353] Crystal structure of PARP1 in complex with DNA has been solved[354]
PCBP1	ssDNA[355]	ssRNA[356] mRNA[14]	Three KH domains[355]	Transcriptional regulation[357] RNA processing[357] Translational regulation[357]	Prefers C-rich DNA and RNA[355] Similar to hnRNP K[355] PCBP2, PCBP3 show similar properties, with some binding dsDNA;[356,358] PCBP2 can bind both DNA and RNA via its KH1 domain[358,359] PCBP2-RNA and DNA complex structures have been solved[358,359]

Protein/ gene	DNA bound by protein	RNA bound by protein	Structural domains	General protein function(s)	Notes
PGK1 (Phospho-glycerate kinase)	ssDNA[360]	mRNA[361]		Glycolysis[362]	DNA binding may be indirect[360]
PRNP (Prion protein)	ssDNA[363] dsDNA[364]	HIV-1 RNA[365]	Globular domain mediates DNA binding[366]	Regulation of ionic current[367] Cell adhesion[367]	DNA inhibits aggregation of PRNP[364] RNA may stimulate conversion of PRNP[368]
PTBP1	ssDNA[369]	mRNA[370]	RRMs[369]	RNA splicing[371] Cell proliferation[372]	Binds pyrimidine-rich DNA and RNA Multiple nucleic acid binding sites[369] Structure in complex with RNA exists[373] DNA binding is sequence-specific[369]
PURA (PURα)	ssDNA[374] and dsDNA[375]	ss-RNA[376] mRNA[14,377]	PUR repeats[376]	Transcriptional control[377] Translational control[377]	PURB (PURβ) shows similar properties[375] Binds to purine-rich sequences[377] Unwinds G-rich dsDNA[375]
RAD51AP1	ssDNA and dsDNA[378] D-loop DNA[379]	ss-RNA[378]	Two distinct DNA binding domains[380]	DNA damage response[381] Homologous recombination[379]	Binding to DNA is structure-specific[381]
RARA (RARα)	dsDNA[382]	5' UTRs (ex: GluR1)[383]	CCCC Zinc finger[382] F domain[383]	Transcriptional control[384] Development and cell differentiation[385]	DNA and RNA bind via different domains[383] Binds DNA as a dimer with the retinoid X receptor; a crystal structure of this complex exists[382] Binding to DNA[382] and (apparently) RNA[383] is sequence specific
RBM3	dsDNA (ex: vaccinia virus late promoter)[386]	miRNA[387] mRNA[388]	RRM[389]	Cold response[389] miRNA biogenesis[387] Transcriptional regulation[386] Control of apoptosis[390]	Binds RNA and DNA via the same domain[386,389] RBM5, RBM6, RBM8, RBM12, and RBM19 also bind DNA and RNA[13,14,391,392]
RBMS1	ssDNA and dsDNA (ex: c-myc promoter)[393]	3' UTRs[394]	Two RRMs[395]	Transcriptional regulation[396] DNA replication[393]	RBMS3 also binds to DNA[395] and RNA;[394] RBMS2 may share similar properties
RPL7	dsDNA[397]	rRNA[397] mRNA[397]	bZIP[397]	Translational regulation[398] Cell cycle and apoptosis[399]	Ribosomal constituent Binds RNA with higher affinity than DNA[397] RPL6 also binds DNA and RNA[400] Binds guanine-rich RNA[401]
RTF1	ssDNA and dsDNA[402,403]	mRNA[14] ssRNA[404]	Plus3 domain[402]	Histone modification[405] Transcription elongation[406]	Binds transcription bubble-like DNA[402] Binds ssDNA witch much higher affinity than dsDNA[402]
RUNX1 (AML1)	dsDNA (ex: ALOX2 promoter)[407]	RNA aptamers[408,409]	Runt domain[410]	Hematopoiesis[411] Transcriptional regulation[407,412]	RNA and DNA binding are competitive[408] RNA binding is sequence specific[408] Crystal structure with DNA exists[413]
RUVBL1 (Pontin52)	ssDNA and dsDNA[414]	ss-RNA[414]	Rossmann-like fold[414]	Transcriptional control[415] DNA damage response[415] Stress response[416] Cell proliferation[417]	RUVBL2 also binds DNA[418] and RNA[419] Forms a dodecameric ring with RUVBL2[420]

Protein/ gene	DNA bound by protein	RNA bound by protein	Structural domains	General protein function(s)	Notes
SAFB	Scaffold/matrix attachment region DNA[408]	mRNA[14,421]	SAF-box[422] RRM	Transcriptional regulation[423] RNA metabolism[424] Stress response[424]	Binds AT-rich DNA sequences through its SAF-box[422]
SFPQ (PSF)	ssDNA and dsDNA[425]	mRNA[14,426]	Two RRMs[426]	RNA splicing[426] Transcriptional regulation[427] Viral mRNA synthesis[428]	Binds U-rich RNA sequences[426] Homolog and binding partner of NONO (see above)[429]
SMAD1	dsDNA[430]	miRNA[3]	Two MH domains[3]	miRNA processing[431] Signal transduction[432] Transcriptional regulation[432]	DNA and RNA bind via the same domain and are competitive[3] SMAD3 likely shows similar properties[3] Crystal structure in complex with DNA has been solved[430]
SMN1 (SMN)	ssDNA and dsDNA[433]	ssRNA[433]	HMG box-like[433]	Cell survival[434] RNA turnover[435]	Prefers G-rich RNA sequences[433] DNA and RNA bind via the same domain[433]
SON (NREBP)	dsDNA[436]	mRNA[14,41]	DRBM domain	RNA splicing[437] Pluripotency[438]	
SOX2	dsDNA[439]	lncRNAs (ex: RMST)[440,441]	HMG box[439]	Pluripotency[442]	RMST is required for Sox2-promoter binding[440] Crystal structure with DNA has been solved[439] DNA binding is affected by post-translational modification[443,444]
SPEN (MINT)	ssDNA and dsDNA (ex: osteocalcin promoter)[445]	RNA SRA[446]	Four RRMs	Notch signaling[447] Transcriptional regulation[448]	
SRSF1 (ASF1/SF2)	ssDNA[449] Telomere DNA[449]	mRNA[450]	Two RRMs and RS-rich motif[450]	RNA splicing[450] Apoptosis[451]	Structure in complex with RNA has been solved[452] May prefer purine-rich sequences[453]
SSBP1	ssDNA[454]	mRNA[14,41]	SSB domain (OB-fold)	DNA replication[455]	Binds T-rich DNA[454]
STAT1	dsDNA (ex: IFN-γ activation sites (GAS))[456]	TSU RNA[457]	Immunoglobulin fold (p53 and NF-κB-like)[458]	Transcriptional regulation[459] Immune signaling[460]	RNA binding is mediated by a GAS sequence[457] STAT3 also binds DNA[461] and RNA[462] Structure with DNA has been solved[458] Binds palindromic sequences[463]
SUB1 (PC4)	ssDNA and dsDNA[464,465]	mRNA[14,41]	C-terminal domain[466]	Transcriptional activation[467] DNA repair[467]	Higher affinity for ssDNA than dsDNA[465] Structure with DNA has been solved[468]
SUPV3L1	dsDNA[469]	dsRNA[469]	ATP-dependent helicase domain	mtRNA metabolism[470]	Contains both RNA and DNA helicase activity[469] Crystal structure with RNA has been solved[471]
SURF6	dsDNA[472]	rRNA[472]	No recognizable functional domains[472]	Ribosome biogenesis[473] Development[473]	Higher affinity for RNA than DNA[472]
SUZ12	dsDNA (ex:MYT1 promoter)[474]	lincRNAs, ex: ANRIL[475]	C2H2 Zn finger Polycomb domain (VEFS-box)	Chromatin silencing[476] Histone modification[477] Development[477]	Member of polycomb repressive complexes
TAF15	ssDNA[478]	mRNA[14,41,479]	RRM and RanBP2-type zinc finger	Transcriptional control[478] Cell motility[480] Neuronal mRNA splicing[481]	Binds G-rich DNA sequences[478] ssDNA and RNA bind at RRM[478] Similar to FUS (see above)[482]
TARDBP (TDP-43)	ssDNA[483] dsDNA[484]	mRNA[14,485] miRNA[486] ssRNA[487]	Two RRMs	Neurodegeneration[488] RNA splicing[483] miRNA biogenesis[486]	Shows differing sequence specificities for RNA and DNA[484] Similar to FUS/TLS[143] Structures with DNA and RNA have been solved[489,490]

Protein/ gene	DNA bound by protein	RNA bound by protein	Structural domains	General protein function(s)	Notes
TBX5	dsDNA (ex: Nppa promoter)[491]	ssRNA[492]	T-box	Cardiac development[493] Cell proliferation[493]	Binds pyrimidine-rich RNA sequences[492] Structure with DNA has been solved[494]
TCF7 (TCF-1)	dsDNA[495]	RNA aptamers[496]	HMG box[495]	T cell development[497] Transcriptional regulation[498] Pluripotency[499]	RNA binding also prevents protein-protein interactions[496] Binding to DNA and RNA are competitive and via the HMG box[500] DNA binding is sequence-specific[495]
TERF2 (TRF2)	dsDNA[501]	Telomere-repeat-encoding RNA[55]	GAR and myb domains[55, 502]	Telomere maintenance[503] DNA damage response[504]	TERF1 also binds RNA42 and DNA[501] Both DNA[502] and RNA[55] can bind via the myb domain, but the GAR domain is required for RNA but not DNA binding[55]
THRA (Thyroid hormone receptor α)	dsRNA[505]	ssRNA[506] lincRNA SRA[506]	Hinge (between zinc fingers and ligand binding domain) Two C4 zinc fingers[506]	Hormone responses[507] Transcriptional regulation[507]	Can bind DNA as a heterodimer with the retinoid X receptor,[508] monomer, or homodimer[509]
TIA1	ssDNA and dsDNA[510]	mRNA[510]	Three RRMs RNA metabolism[511]	Stress response[511]	Binds ssDNA with higher affinity than dsDNA[510] Binds to DNA with higher affinity than RNA[510] Binds AU-rich RNA[510] TIAL1 also binds DNA and RNA[512]
TOP1	Telomere DNA[513] dsDNA[514] Quadruplex DNA[515]	Hairpin RNAs[516] mRNA27		DNA repair[517] Transcription[517] DNA replication[517]	Contains DNA type IB topoisomerase activity[518] TOP2B may also bind DNA and RNA[519] Structure with DNA exists[520]
TP53 (p53)	ssDNA[521] dsDNA[522] Damaged DNA[523]	ssRNA[521] 5' and 3' UTRs (ex: PAI-1)[524, 525]	Core DNA-binding domain; β-sandwich containing[526]	DNA damage response[527] Transcriptional activation[527] Apoptosis[527]	RNA binding may regulate oligomerization of p53[528] Affinity for ssRNA higher than for ssDNA[521] Sequence-specificity of RNA binding unclear[529] Structure in complex with DNA has been solved[530]
TSN (Translin)	ssDNA[531] Telomeres and microsatellite repeats[532]	mRNA[533] ssRNA and dsRNA[534] miRNA[535]	Leucine zipper Basic region required for DNA binding[531]	RNA turnover[536] RNA interference[536] RNA trafficking[537]	Binds GT-rich DNA[532] Structure of homologous protein in complex with RNA has been solved[538]
WBP11	ssDNA[539]	ssRNA[539]		mRNA processing[540]	Binds G-rich DNA and RNA[539]
WT1	dsDNA[541]	mRNA[541]	Four C2H2 Zinc fingers[542]	Epithelial-mesenchymal transition542 Development[543]	Zinc fingers are required for DNA[542] and RNA[541] binding NMR and crystal structure of WT1 bound to DNA has been solved[542]
YBX1 (YB-1)	ssDNA and dsDNA[544] Damaged DNA[545]	mRNA[14, 546] ssRNA[547]	Cold shock domain[547]	Cell proliferation[548] Transcriptional regulation[549]	Higher affinity for ssDNA than dsDNA[547] Binds CT-rich DNA[544] YBX2[550] and YBX3[551, 552] also bind DNA and RNA

Protein/ gene	DNA bound by protein	RNA bound by protein	Structural domains	General protein function(s)	Notes
YY1	dsDNA (ex: grp promoter)[553]	lincRNA Xist[554]	Four C2H2 zinc fingers	Transcriptional regulation[555] V(D)J recombination[555]	Binds DNA and RNA via different sequence motifs[554] Structure with DNA has been solved[556]
ZC3H8	dsDNA[557]	mRNA[14,41]	Three CCCH zinc fingers	Transcriptional regulation[557] T cell homeostasis[558]	
ZNF239 (MOK2)	dsDNA (sequence-specific)[559]	poly(G) and poly(U)[559]	Seven Zn fingers	Transcriptional regulation[560]	DNA and RNA binding are not competitive[559]
ZNF638 (ZFML)	dsDNA[561]	mRNA[14,41]	Two RRMs Zinc finger	Adipocyte differentiation[562] Transcriptional regulation562	Binds C-rich DNA[562]

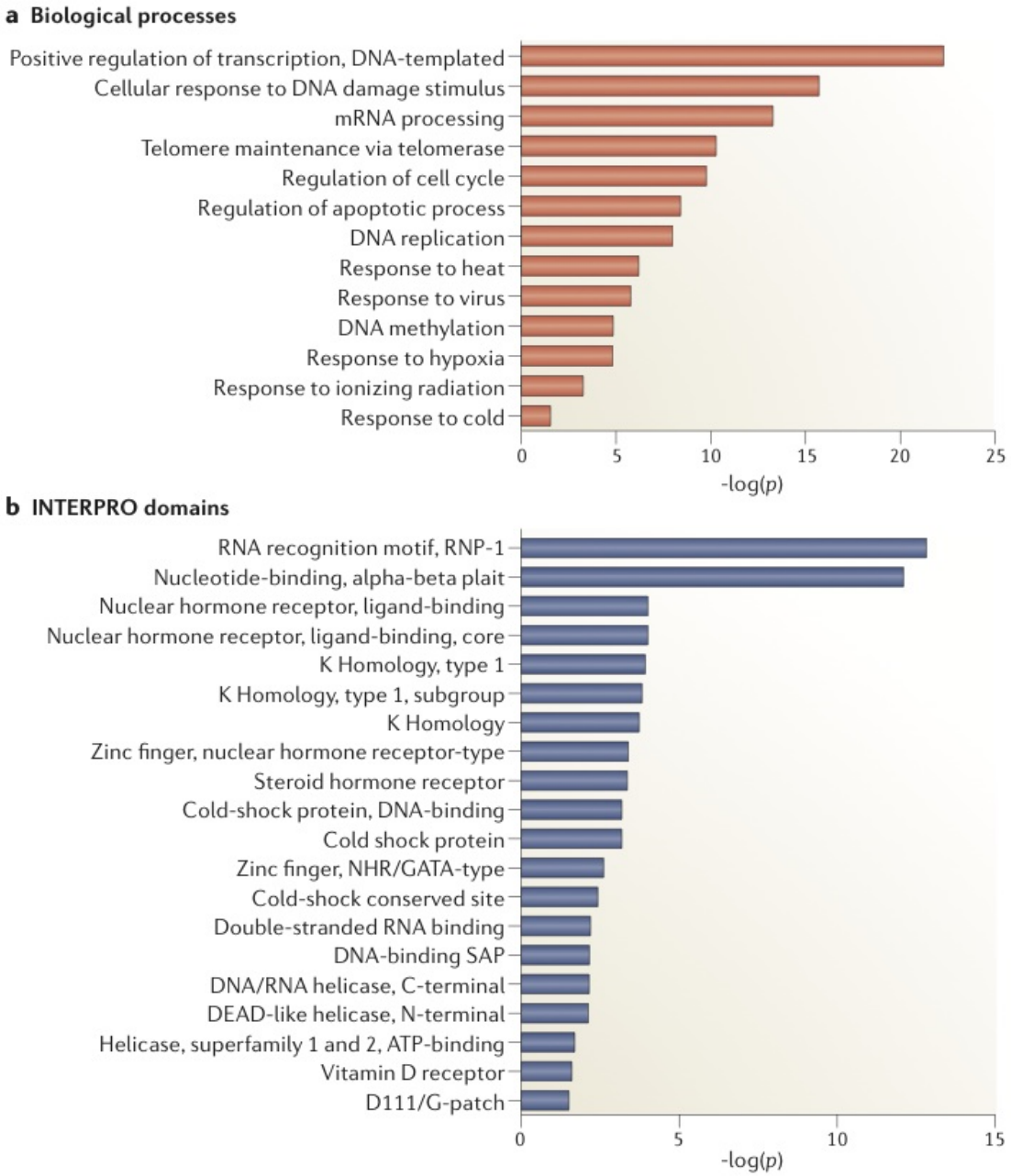


Figure 1.2: Functional and structural properties of DRBPs. The 149 DNA- and RNA-binding proteins (DRBPs; see Table 1.1) were subjected to gene ontology enrichment of biological process (PROTEOME database; BioBase) and to INTERPRO domain enrichment (DAVID ontology[563,564]) in order to explore the biological functions of and protein domains commonly found in DRBPs. (a) In gene ontology analysis, biological processes such as transcriptional regulation and mRNA processing are expectedly prominent terms found to be enriched in DRBPs. However, unexpected functions are also enriched, including response to many cellular stresses (such as heat, viral infection and radiation). For brevity, only selected functions are shown. (b) In domain enrichment analysis, all domains enriched in the set of 149 DRBPs that have P values $\leq 10^{-3}$ are shown. P values in parts (a) and (b) indicate the probability that the over-representation of the stated term in the 149 DRBPs compared with all human genes is due to chance. RNP1, ribonucleoprotein 1.

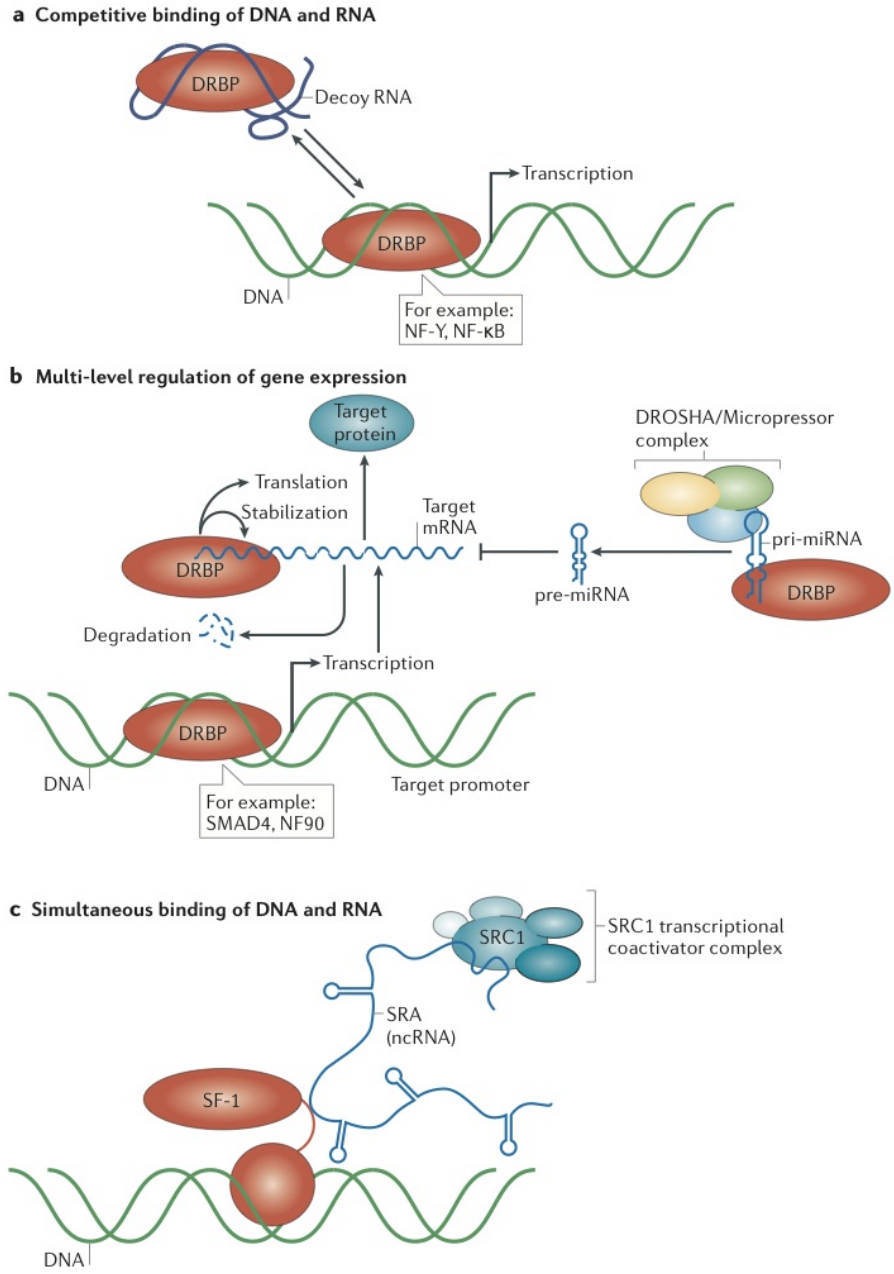


Figure 1.3: Three archetypes of DRBP function. (a) RNA can compete with DNA for binding to DNA- and RNA-binding proteins (DRBPs), typically at the same protein interface. In the case of transcription factors, this can reduce promoter occupancy and the transcription of target genes. (b) DRBPs can regulate gene expression at multiple levels. In addition to binding to the promoters of genes to regulate their transcription, DRBPs can also affect microRNA (miRNA) processing, as well as mRNA stability and translation. (c) DRBPs can bind DNA and RNA simultaneously, allowing the RNA to function as a scaffold to recruit other proteins to a specific DNA locus. Shown here is steroidogenic factor 1 (SF-1) binding to the long non-coding RNA (lincRNA) steroid receptor RNA activator (SRA) to recruit the SRC1 (also known as NCOA1) transcriptional complex in a ligand-independent manner. NF, nuclear factor; NF90, NF of activated T cells 90 kDa; pre-miRNA, precursor miRNA; pri-miRNA, primary miRNA.

1.4.1 Binding DNA or decoy RNAs

The role of certain lincRNAs as decoys of genomic DNA is illustrated by the reduction in promoter occupancy by transcription factors, typically measured by chromatin immunoprecipitation (ChIP), in response to the overexpression of competing decoy RNAs. The glucocorticoid receptor (GR), a steroid hormone receptor, is a classic example of a ligand-activated transcription factor (reviewed in reference 332). In its inactivated state, the GR is kept in the cytoplasm by chaperone proteins. Upon ligand binding, the GR translocates to the nucleus, where it can bind to the promoters, and regulate the transcription, of hundreds of genes.[565] Given the anti-inflammatory role of the GR, much effort has been put into developing modulators of GR-driven transcription.[566] Recently, the lincRNA growth arrest-specific 5 (Gas5) was found to inhibit the transcriptional activity of the GR by competing directly with DNA for protein binding *in vitro* and in cells;[2] overexpression of Gas5 leads to a decrease in ChIP-detected GR occupancy at its target promoters, as well as a decrease in the mRNA levels of glucocorticoid-activated genes.[2,567] As cellular Gas5 levels are regulated by nonsense-mediated decay[567] in response to serum starvation and other stressors,[2] the transcriptional activity of the GR is tuned by titrating the levels of Gas5 against the fixed number of genomic GR-binding sites in response to cellular stress. Three closely related steroid receptors that share the DNA specificity of the GR - the androgen, progesterone and mineralocorticoid receptors - are also susceptible to Gas5-mediated transcriptional repression.[2] Although steroid receptors have traditionally been thought of as DNA-binding proteins, the affinity of the GR for RNA and DNA is similar, as measured *in vitro* by glutathione S-transferase pulldown assays and fluorescence-based competition assays (see also Chapter 2).[2] The most distantly related member of the steroid receptor family, the estrogen receptor, does not share the DNA specificity of the GR and is not susceptible to Gas5-mediated transcriptional repression, indicating that the binding of steroid receptors to RNA is

sequence specific, a topic further explored in Chapter 2.[2]

Additional examples of pairs of transcription factors and decoy RNAs are nuclear factor-Y (NF-Y), which binds the lincRNA P21-associated ncRNA DNA-damage activated (PANDA),[8] and NF-κB, which binds the mouse pseudogene-derived RNA Lethe.[9] The dual nucleic acid-binding activity of NF-κB was demonstrated *in vitro* many years before the discovery of an endogenous RNA target,[10] suggesting that transcription factors that are known DRBPs *in vitro* may have endogenous RNA targets awaiting discovery; an example of such DRBPs is the acute myeloid leukemia 1 (AML1; also known as RUNX1) protein.[408] Although structural information on the interaction of human proteins with their decoy RNAs is lacking, a recent study demonstrated an elaborate mechanism of an analogous bacterial system: the sequestration of ribosomal RNA small subunit methyltransferase E (RsmE) by the non-coding RNA RsmZ.[11] Competitive DNA and RNA binding is a feature not only of transcription factors but also of nucleic acid-modifying enzymes, such as DNA methyltransferases. In humans, DNA methylation is initiated by DNA (cytosine-5)-methyltransferase 3A (DNMT3A) and DNMT3B; DNMT1 maintains this methylation by binding to hemimethylated DNA after replication (reviewed in reference 568). RNA binding can inhibit the DNA-binding and methylation activity of both DNMT3A (reference 110) and DNMT1 (reference 106). *In vitro*, DNMT1 binds RNA with a higher affinity than DNA, as shown in electrophoretic mobility shift assays (EMSAs).[110] In the case of DNMT1, and probably in that of DNMT3A, RNAs bind to the catalytic domain of the methyltransferase to inhibit DNA methylation.[106,110]

It is notable that several metabolic enzymes - such as the glycolytic enzymes lactate dehydrogenase,[106,251–253] glyceraldehyde-3-phosphate dehydrogenase (GAPDH)[148–150] and α-enolase (ENO1)[14,118,119] - are DRBPs with competitive DNA- and RNA-binding capacities. In the case of GAPDH, DNA and RNA compete for binding of the cofactor NAD+ to the enzyme,[148,150] suggesting that Rossmann fold-containing

proteins such as GAPDH may be sensitive to cellular DNA and/or RNA levels. ENO1 binds RNA as a monomer,[119] which inhibits the formation of the catalytically active protein dimer.[119,569] NAD+-specific isocitrate dehydrogenase, which converts isocitrate to α-ketoglutarate, is allosterically inhibited by the 5' untranslated regions of yeast mitochondrial mRNA.[570] Binding of RNA and DNA to metabolic enzymes indicates that nucleic acids can regulate the function of proteins other than transcription factors to modulate cellular metabolism.[571]

1.4.2 DRBPs that regulate gene expression at multiple levels

Approximately half of the DRBPs we identified in our analysis are transcription factors. As discussed above, some such proteins have been shown to be the targets of decoy RNAs. By contrast, several others bind both the DNA and the mRNA of their target genes (Figure 1.3b). Regulating genes at both the DNA and the RNA levels allows powerful, combinatorial control over protein expression and may enable DRBPs to generate both immediate effects (through regulating RNA turnover) and long-lasting effects (through regulating transcription).

When activated, the GR can promote the transcription of anti-inflammatory genes[572] and repress the transcription of pro-inflammatory genes.[572–574] Agonist-bound GR destabilizes the mRNA of pro-inflammatory genes such as the chemokine (C-C motif) ligand 2 (*CCL2*; also known as MCP1) gene through direct RNA binding, perhaps by the recruitment of ribonucleases.[329] The presence of a GR-binding motif in many immunogenic mRNAs has been confirmed using RNA immunoprecipitation (RNA-IP) and suggests that the GR can accelerate the decay of many mRNAs, broadening its known role in the anti-inflammatory reponse.[1] Given that the GR also binds directly to pro-inflammatory transcription factors, such as adaptor protein complex 1 and NF-κB,[575,576] it seems that the GR uses its diverse DNA-, RNA- and protein-binding capacities to regulate inflammatory genes at the transcriptional and

post-transcriptional level.

Transcription factors can also regulate gene expression post-transcriptionally through the regulation of microRNA (miRNA) biogenesis. miRNAs are small RNAs that facilitate gene silencing through sequence-specific pairing to target mRNAs and recruitment of these to the RNA-induced silencing complex (RISC; reviewed in reference 577). Several transcription factors have been shown to regulate Drosha-mediated primary miRNA (pri-miRNA) processing, a key step in the biogenesis of functional miRNAs.[578] SMAD proteins, which are transducers of transforming growth factor-β (TGFβ) signalling, activate transcription by forming a DNA-binding heterodimer (reviewed in reference 579). SMAD proteins also increase the levels of several miRNAs, including miR-21 (reference 431), which has important roles in development and immunity.[580] Surprisingly, the increase in miR-21 levels is due not to increased transcription of pri-miR-21 but to increased Drosha-mediated processing of pri-miR-21 to the precursor miRNA (pre-miRNA) mir-21 (reference 431). Bioinformatic analysis identified a conserved RNA motif in TGFβ-regulated miRNAs, which was shown by RNA-IP and EMSA to bind directly to the MAD homology 1 domain of SMAD to mediate Drosha processing.[3] Interestingly, the RNA sequence motif that is bound by SMAD4 and that mediates the regulation of miRNA expression post-transcriptionally is identical to the DNA sequences that are bound by SMAD4 and that mediate regulation of gene expression transcriptionally.[3]

Nuclear factor of activated T cells 90 kDa (NF90; also known as ILF3) is a particularly versatile DRBP that, along with its partner NF45 (also known as ILF2), has important roles in T cell activation.[581] Through the direct binding of DNA, mRNA and miRNA, NF90 controls transcription,[4,5] regulates mRNA turnover and translation,[233,582] and affects miRNA processing,[583] respectively. These functions assist in its role in T cell activation: NF90 upregulates the mRNA levels of interleukin-2 (IL-2), a critical cytokine in T cell development,[584] by binding its promoter to

activate its transcription and by stabilizing the IL-2 mRNA through direct binding to its 3' untranslated region, as was found by EMSA and ribonucleoprotein (RNP) immunoprecipitation analysis.[4,582] Additionally, using *in vitro* pri-miRNA-processing assays and RNA-IP, NF90 in complex with NF45 was shown to inhibit the processing of the pri-miRNA pri-let-7a by binding it directly.[583] let-7a represses IL-6, a cytokine that is critical for T cell survival and proliferation,[585] which may link inflammation to cancer,[586] and let-7 downregulation following NF90 upregulation reduces survival in several cancer types.[587,588] In summary, these examples illustrate that DRBPs can use both transcriptional and post-transcriptional mechanisms to serve as potent controllers of gene expression.

1.4.3 Simultaneous binding of DNA and RNA

In contrast to DRBPs that target DNA or RNA serially to serve different or related functions, another class of DRBPs binds RNA and DNA simultaneously to perform a single function (Figure 1.3c). Generally, transcription factors not only require DNA binding to target promoters but also bind to co-repressors or co-activators to affect transcriptional regulation. There are several examples of RNA molecules acting as co-activators by simultaneously binding DNA and various transcription factors. The lincRNA rhabdomyosarcoma 2-associated transcript (RMST), in particular, is required for binding of neurogenic gene promoters and subsequent upregulation by SOX2 (reference 440), a transcription factor with important roles in development, pluripotency and cell fate.[589] RNA-IP and RNA pulldown experiments showed that RMST interacts directly with SOX2 (references 440, 441), and DNA occupancy of SOX2 measured by ChIP followed by sequencing (ChIP-seq) was reduced following RMST depletion.[440,441] The lincRNA EVF2 is a transcriptional co-activator of DLX2 (reference 101) and recruits methyl-CpG-binding protein 2 (MECP2) to intergenic enhancers.[590] A direct interaction between DLX2 and EVF2 has been demonstrated

by the immunoprecipitation of DLX2 followed by reverse transcription PCR of the EVF2 lincRNA,[101] and MECP2 also has previously been shown to bind RNA.[282] It should be noted that RNA-mediated recruitment of a protein to a particular DNA locus might not require direct binding of both DNA and RNA by the protein, as lincRNAs could recruit transcription factors to a particular DNA locus to which the lincRNA is bound. Dual nucleic acid recognition also facilitates targeted gene repression through RNA-guided DNA methylation. This phenomenon was first discovered in plants,[591] and some mammalian RNA guides of DNA methylation have since been found,[592,593] although their mechanisms of action are less clear. In mice, DNMT3A forms a complex with Tsix RNA to promote methylation of the X-inactive-specific transcript (Xist) promoter.[594]

Several nuclear receptors - including steroidogenic factor 1 (SF-1), DAX1 (also known as NR0B1) and thyroid receptor-α (TRα) - bind simultaneously to both gene promoters[325,336] and the RNA co-activator SRA (steroid receptor RNA activator) to modulate transcriptional activation[6,7] (Figure 1.3c). Using pulldown experiments, SF-1 and TRα have been shown to bind SRA through their hinges, which are flexible, disordered regions that connect their DNA- and ligand-binding domains.[6,506] Knockdown of SRA decreases the interaction of SF-1 with protein transcriptional activators and the transcription of SF-1-regulated genes.[6] Several other nuclear receptors associate with, but lack direct evidence for direct binding to, SRA, including the androgen, progesterone and estrogen receptors, as well as retinoic acid receptor-α (RARα), which may bind SRA and its target gene promoters simultaneously.[7,595–597] Crosslinking immunoprecipitation has demonstrated that RARα can bind to and regulate the translation of target mRNAs through a unique RNA-binding motif at its carboxyl terminus.[383]

Another example of simultaneous DNA and RNA binding that is required for DRBP function is the role of telomeric repeat-binding factor 2 (TRF2; also known

as TERF2) at telomeres. Deletion of TRF2 leads to an arrest in cell division caused by the formation of chromosome end-fusions.[598] Crystal structures have revealed that TRF2 binds to telomeric DNA in a sequence-specific manner through a C-terminal DNA-binding domain, which resembles a homeodomain.[502] Part of the role of TRF2 at the telomere includes the recruitment, through its positively charged amino-terminal GAR domain, of the origin-recognition complex (ORC, which is a collection of proteins that serves as a scaffold for DNA replication factors, among other functions[599]) so that it can assist in the maintenance of telomere structure.[600] Using biotinylated RNA pulldown experiments, RNA-IP and EMSAs, the GAR domain responsible for ORC recruitment was later shown to bind telomere repeat-encoding RNA (TERRA).[55] Depletion of TERRA hampers ORC recruitment to the telomere without affecting TRF2 binding to the telomere itself, suggesting a model in which TRF2 serves as a mediator between telomere DNA and TERRA, which in turn recruits factors required for telomere maintenance.[55]

1.5 Structural characteristics of DRBPs

For a protein such as TRF2 to coordinate telomeric DNA binding and recruit protein complexes by binding of RNA, it must have multiple nucleic acid-binding motifs. Some DRBPs, such as the GR and NF-κB, have maintained domains that are capable of binding both DNA and RNA, which allows decoy RNAs to evolve and compete with DNA for protein binding. We analyze next the prevalence of structural domains in DRBPs and discuss examples of DRBPs that bind both single-stranded and double-stranded nucleic acids.

1.5.1 DRBP domains that enable DNA and RNA interactions

We carried out InterPro domain enrichment analysis by DAVID[563,564] on our 149 DRBPs to identify domains enriched in proteins that bind both DNA and RNA (Figure 1.2b). The RNA-recognition motif (RRM; also known as the RNP domain or RNA-binding domain) was the most highly enriched domain in DRBPs ($P = 2 \times 10^{-26}$). The RRM is an abundant, short (∼100-amino-acid long) domain that generally recognizes ssRNA and is often present in proteins with other domains, such as zinc-finger domains, WW domains or additional RRMs.[601] Such multidomain DRBPs may bind RNA and DNA simultaneously through separate domains, as occurs with heterogenous nuclear RNA A1, which contains two RRMs.[175] Single RRM-containing proteins are also capable of binding both DNA and RNA, and this function is present, for example, in RNA-binding motif 3 (RBM3), TBP-associated factor 15 (TAF15), and TAR DNA-binding protein 43 (TDP43; also known as TARDBP; see Table 1.1).[485,602,603] Such bivalent domains may not have the same sequence specificity when binding DNA and RNA, highlighting the complexity of recognizing nucleotide bases in a sequence-dependent manner.

Nuclear receptor domains are also enriched in DRBPs: RARα binds mRNA through a unique C-terminal domain,[383] SF-1 binds the RNA co-activator SRA through its hinge and a unique Ftz-F1 domain, and TRα binds SRA through its hinge.[6,506] The majority of nuclear receptors have two highly conserved Cys_4 zinc-fingers through which they bind DNA, and some nuclear receptors, such as the GR, can also bind RNA through these domains.[2] Other types of zinc-fingers are also enriched in DRBPs, such as the RAN-binding protein 2 (RANBP2) type. Other notably enriched domains in DRBPs are the K-homology domain, the dsRNA-binding domain (dsRBD), the cold-shock domain (CSD) and various helicase domains. Each of these domains is capable of binding DNA and RNA, and we focus below on the structural mechanisms

underlying this dual specificity.

1.5.2 General properties of DRBPs

There are only two chemical differences between RNA and DNA. First, RNA (but not DNA) has a 2' hydroxyl (2' OH) group on the ribose sugar, which allows an additional hydrogen bond to be formed and a greater diversity of secondary structures than is possible in DNA. Second, RNA contains uracil rather than thymine, as in DNA; uracil lacks a methyl group at the C5 position. A comparative analysis of known protein-nucleic acid structures revealed that the recognition of DNA occurs largely through electrostatics and direct base-protein interactions. Conversely, RNA recognition by proteins mainly depends on shape complementarity and interaction with the 2' OH group.[602] Given these general differences, one could expect that, during evolution, highly selective protein interfaces would be generated that are optimized for either RNA or DNA, with minimal cross-binding. However, the most energetically favorable associations between proteins and nucleic acids rely on hydrophobic and charge-charge interactions. These interactions are less constrained than interactions with the sugar backbone or with the nucleotide base edge, which is capable of highly-specific Watson-Crick base pairing. Thus, DRBP domains that competitively bind DNA and RNA probably rely on the less specific hydrophobic and charge-charge interactions. For example, ssRNA-binding proteins are more likely to form hydrogen bonds with bases rather than with the phosphate-sugar backbone, compared to those that recognize folded RNA, such as ribosomal proteins and tRNA synthetases.[603] Because ssRNA-binding proteins do not rely heavily on sugar recognition, they are more likely to also bind DNA. This may explain why the RRM is the most enriched domain in the DRBPs included in our analysis (Figure 1.2b).[603]

1.5.3 The RRM

The RRM is an extremely versatile domain that is capable of binding (mainly single-stranded) RNA and DNA, as well as proteins.[601] RRMs preferentially interact with nucleic acid bases rather than with the phosphate-sugar backbone. The structural nature of ssRNA and ssDNA allows much easier access to the exposed aromatic base faces, as opposed to hydrogen bonding to the base edges that occur frequently with double-stranded nucleic acid-binding DRBPs. Additionally, stacking interactions with the faces of bases are more energetically favorable than recognition of the nucleotide edge. Therefore, stacking interactions between aromatic protein side chains and nucleic acid bases are often observed in single-stranded nucleic acid-binding proteins.

TDP43 is a DRBP that has important roles in mRNA splicing and miRNA biogenesis.[485,486] It contains two RRMs, which are separated by a short loop and are both capable of binding DNA and RNA. Crystal structures of the TDP43 RRMs in complex with DNA and RNA have been reported,[489,490] making TDP43 an excellent case study for dual DNA and RNA recognition by RRMs. The DNA- and RNA-bound structures of TDP43 reveal nearly identical modes of nucleic acid recognition. Aromatic side chains, such as Phe149 within the first RRM (RRM1), form stacking interactions with DNA or RNA bases (Figure 1.4a). Trp113, which is part of the more flexible loop 1, is able to shift conformations and base-stack slightly differently when bound to different nucleic acid sequences (Figure 1.4b), whereas Phe149 in the rigid β3 sheet of the RRM1 fold makes similar interactions with DNA and RNA (Figure 1.4a). Relying on the more energetically favorable π-stacking interactions through the planar face of the DNA and RNA bases results in less specificity than that gained from hydrogen bonding with the base edge. Uracil and thymine interact with Phe194 of the second RRM (RRM2) in the RNA- and DNA-bound structures, respectively (Figure 1.4c,d). Despite the additional methyl moiety at position C5 in the DNA,

no TDP43 residues recognize the edge of the nucleotide to interact or clash with the additional carbon (Figure 1.4d). Thus, one of the chemical differences between DNA and RNA, the use of uracil in RNA, plays no part in nucleic acid discrimination in this example.

By contrast, the RRM2 of TDP43 does make RNA-specific contacts with the 2'OH group. The majority of protein-2' OH group interactions are mediated through protein side chains,[602] and both the Lys263 and Arg227 (Figure 1.4e) side chains in RRM2 contact a 2' OH group when bound to RNA. However, when TDP43 is bound to DNA, these same protein side chains contact the DNA backbone phosphates (Figure 1.4f), demonstrating that amino acids are capable of reorienting to allow distinct types of interactions to support RNA and DNA binding. Nevertheless, DNA binding is not a general property of all RRMs. For example, the RRM of poly(A)-binding protein relies on many RNA-specific 2' OH contacts for RNA interaction, and binding to DNA may be at low affinity, if detectable at all.[604]

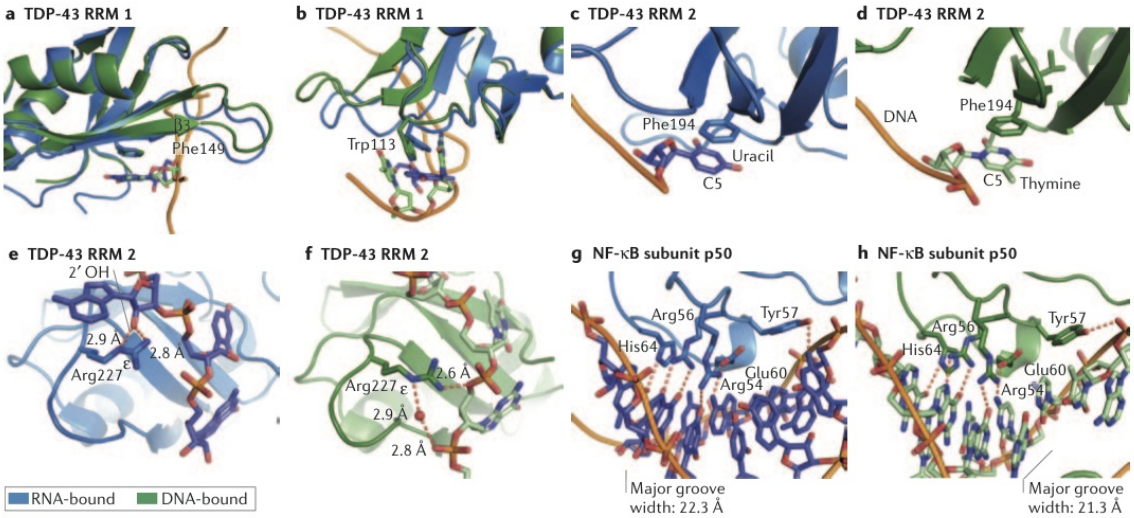


Figure 1.4: The structural basis for dual DNA and RNA recognition by TDP-43 and by the NF-κB subunit p50. Protein-RNA structures are shown in blue and protein-DNA structures in green, with protein in the darker shade. π-stacking interactions play a prominent part in both the single-stranded DNA (ssDNA)- and the ssRNA-binding activities of TAR DNA-binding protein 43 (TDP43). (a,b) Phe149 part (a) and Trp113 part (b) within the first RNA-recognition motif (RRM1) of TDP43 stack with both RNA and DNA bases. (c,d) In the second RRM (RRM2) of TDP43, Phe194 is capable of recognizing both uracil in RNA part (c) and thymine in DNA part (d); the additional methyl group at C5 in thymine does not contribute to nucleic acid specificity. (e) When bound to RNA, both the terminal amine and ϵ-nitrogen of Arg227 in RRM2 of TDP43 contact a 2' hydroxyl (2' OH) group on the RNA backbone. (f). By contrast, these same groups can also make contacts with the DNA backbone, both directly and through water-mediated hydrogen bonding. (g,h). The p50 subunit of nuclear factor-κB (NF-κB) makes strikingly similar base-specific contacts when bound to an RNA aptamer part (g) or to double-stranded DNA part (h). This is due in large part to the similar secondary structure and chemical moieties presented by the RNA and DNA. Major groove width was calculated with 3DNA software using phosphate-phosphate distances.[605]

1.5.4 DRBPs that recognize double stranded nucleic acids

Crystal structures of protein-dsRNA complexes are less common than their single-stranded counterparts, but there are some examples that are instructive for dual nucleic acid recognition. NF-κB is a central transcription factor of immune signalling and is formed of homodimers or heterodimers of Rel family proteins, such as p50 (also known as NFKB1) or p65 (also known as RELA).[306] High-affinity aptamers have been developed for both the p50 subunit and the p65 subunit of NF-κB, with an affinity of RNA binding that approaches that of the transcription factor's affinity for native DNA response elements.[10,305] DNA with an identical sequence to the p50-targeting aptamer will not bind p50 (reference 10); p50 is therefore another DRBP that binds RNA and DNA in a sequence-specific manner, with different sequence specificities for each.

Crystal structures of p50 bound to DNA and to RNA reveal that both bind at the same surface of the p50 immunoglobulin-like domain.[307,308] Although p50 binds to RNA as a monomer and to DNA as a dimer, similar networks of base-specific interactions occur between protein and nucleic acids in each structure (Figure 1.4g,h). Not only do the DNA- and RNA-contacting residues of p50 maintain an equivalent position, but both DNA and RNA also present similar interfaces for p50 recognition in charge distribution and in secondary structure[605] (Figure 1.4g,h). This is a seminal, structurally confirmed example of 'DNA mimicry' by RNA to bind to a transcription factor and, although the RNA in this case was artificial, DNA mimicry has been hypothesized to play a part in the endogenous regulation of several transcription factors.[2,606]

Structures have also been solved of the DRBP dsRNA-specific adenosine deaminase 1 (ADAR1), which binds both double-stranded Z-DNA and double-stranded Z-RNA through its unique Zα domain.[19,22] The ability of the Zα domain to make sequence-independent interactions with the Z-form phosphate backbone of both DNA

and RNA enables ADAR1 to sense nucleic acid secondary structure conformations. Thus, double-stranded nucleotide-binding DRBPs can recognize their DNA and RNA targets through sequence-specific interactions (in the case of NF-κB) or through non-specific interactions with the DNA and RNA backbone (in the case of ADAR1).

1.6 The evolution of DRBPs

The evolutionary forces driving the structure and function of DRBPs are complex, but understanding them will help us to identify new DRBPs and perhaps predict their susceptibility to interactions with lincRNAs. Although the DRBPs identified in our analysis are members of many different structural classes, each with their own evolutionary history, we focus on members of two very dissimilar DRBP families: CSD-containing proteins and eukaryotic DNA methyltransferases. CSD-containing proteins, which are required to protect cells from low temperatures, are members of an ancient DRBP family that uses weak selection criteria to interact with nucleic acids and therefore intrinsically bind to both DNA and RNA. Members of the eukaryotic DNA methyltransferase family are DRBPs that have more recently evolved the ability to recognize both DNA and RNA: they preferentially interact with DNA, and only one family member (DNMT2) acquired the ability to bind and methylate tRNAs.[109] The discovery of a eukaryotic DNA methyltransferase with a predominant tRNA methylation activity and only a modest DNA methylation activity showcases how evolution can modify protein surfaces to create new functions of DRBPs.

1.6.1 The ancient CSD DRBPs

The CSD is one of the most ancient nucleic acid-binding domains found in bacteria, archaea and eukaryotes. All CSD-containing proteins bind DNA and RNA (see Table 1.1). In humans, there are several CSD-containing proteins, such as the three Y

box-binding proteins, LIN28A and LIN28B (which are homologs of the *Caenorhabditis elegans* LIN-28), and CSD-containing protein E1 (CSDE1). Y box 1 (YB-1; also known as YBX1) was originally named for its ability to bind and repress the Y box of major histocompatibility complex class II promoters.[607] YB-1 also binds RNA, with roles in alternative splicing,[546] translational control[608] and RNA stabilization.[609] In addition, YB-1 binds to damaged DNA and is involved in the DNA-damage response[545,610,611] - it translocates to the nucleus following stresses, such as exposure to ultraviolet radiation.[610,612]

In bacteria, CSDs exist in short proteins that contain one CSD with short flanking sequences. In *Escherichia coli*, there are nine such proteins (cold-shock protein A (CspA)-CspI), which are probably products of multiple gene duplication events.[613] Of these, CspA, CspB, CspG and CspI are induced by cold stress, with CspA constituting >10% of all protein synthesized during cold shock.[614–617] Simultaneous deletion of these four genes results in lack of *E. coli* colony formation at temperatures at or lower than 25 °C (reference 618). CspD is induced by nutrient stress,[619] but CspC and CspE are constitutively expressed at normal growth temperature.[620] Many (if not all) of the Csp proteins bind DNA and RNA[621,622] and have similar roles to those of the human CSD-containing proteins, including in maintaining RNA stability,[621] in translational regulation,[623] in transcriptional control,[623,624] in DNA replication and repair,[625,626] and in chromosome folding.[627]

CSD-containing proteins are widespread in plants,[628] in which they have similar cellular functions. The first Csp-like protein found in plants was wheat CSP 1 (WCSP1), which is upregulated specifically by cold stress and binds ssDNA, dsDNA and RNA homopolymers.[629] WCSP1 was found to complement the cold-sensitive phenotype of the *E. coli* four-gene knockouts mentioned above,[630] exhibiting remarkable functional conservation. In addition, WCSP1 showed nucleic acid melting activity in *E. coli*, which is critical to preventing inappropriate nucleic acid secondary structures

that disrupt and terminate transcription. This activity is similar to the endogenous *E. coli* CspA, which also has transcription antitermination activity.[630] *Arabidopsis thaliana* has four CSD-containing proteins, CSP1-CSP4, all of which can also complement the quadruple csp knockout in *E. coli* to varying degrees, suggesting that their DNA and RNA interactions are well conserved during evolution.[631–633]

Unlike their counterparts in bacteria, most plant and animal CSD-containing proteins have additional functional domains (including more CSDs), which expand their functions, protein-protein interactions and/or nucleic acid-binding specificities. For example, the human protein CSDE1 has five CSDs, which increase the protein's affinity for target RNA sequences.[634] YB-1 has both an N-terminal and a C-terminal domain flanking its CSD, which can support homomultimerization and interactions with many other protein partners (reviewed in Reference 635). In addition to its CSD, WCSP1 has three CCHC zinc-fingers, through which most of its dsDNA binding is mediated.[629] Nevertheless, the exceptional sequence and functional conservation between eukaryotic CSD-containing proteins and bacterial Csp proteins demonstrate a conserved, ancient role and origin of the domain. It is likely that a CSD fold that was capable of binding DNA and RNA was present in the last common ancestor of bacteria, archaea and eukaryotes.[636]

1.6.2 The curious case of DNMT2

DNA methylation has important roles in gene expression and repression of transposable elements in eukaryotic cells. There are three eukaryotic proteins in the cytosine-C5 DNA methyltransferase family: DNMT1, DNMT2 and DNMT3. Whereas DNMT1 and DNMT3 play important parts in maintaining genome-wide methylation, DNMT2 has little DNA methylation activity[637] and is instead capable of methylating tRNAs (Figure 1.5a).[109] When this activity was discovered, it was speculated that the three eukaryotic DNMTs might have evolved from an RNA methyltransferase.[109] However,

there is no evidence that DNMT2 is more closely related to the ancestral protein of the family members. In fact, the three eukaryotic DNMTs may not be monophyletic and may have evolved from separate bacterial DNA methylation restriction-modification enzymes.[638] Thus, it seems likely that DNMT2 shifted its nucleic acid specificity from DNA to RNA in the last common eukaryotic ancestor[638] (Figure 1.5b).

Despite the relatively narrow substrate specificity of DNMT2 compared with that of its family members, it is highly conserved and is the only extant DNMT in some species, such as *Schizosaccharomyces pombe* and *Drosophila melanogaster*.[638] This seems to indicate that DNMT2 has an important physiological role; however, Dnmt2$^{-/-}$ mice are viable and fertile, and yield no obvious phenotype.[109] This apparent contradiction was resolved with the recent report that deletion of DNMT2 in addition to another tRNA methyltransferase, NSUN2, in mice is lethal.[345] These mice show defects in tRNA stability, protein synthesis and differentiation,[345] implying that the DNA methylation activity of DNMT2 is dispensable, whereas its tRNA methylation activity is not.

NSUN2 is a member of the nuclear protein 1 (NCL1) family of eukaryotic RNA cytosine-C5 methyltransferases, which are broadly distributed among eukaryotes.[639] Interestingly, NSUN2 itself is a DRBP and is able to bind and methylate both tRNA and hemi-methylated DNA.[341] Crosslinking immunoprecipitation-based analyses showed that NSUN2 also methylates mRNAs and non-coding RNAs.[344] Given the distant evolutionary relationship between DNA and RNA cytosine-C5 methyltransferases,[638] NSUN2 and DNMT2 have most likely undergone convergent evolution from an RNA-binding and a DNA-binding protein family, respectively, to ensure proper tRNA modification. These evolutionary trajectories have bestowed on both proteins the ability (if residual) to bind and modify both DNA and RNA. This indicates not only that proteins with evolutionarily conserved DNA-binding activities are capable of binding RNA (and vice versa) but also that some nucleic acid substrates may be

similar enough in sequence and structure to promote binding promiscuity. As mentioned above, this phenomenon is exploited by RNAs, both endogenous and artificial, that function as decoys to modulate DRBP function.[2,408,640]

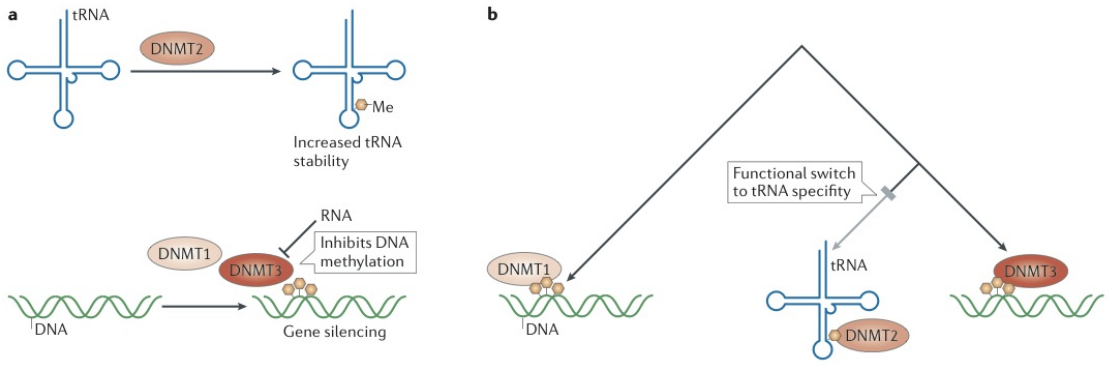


Figure 1.5: DNA methyltransferases target both DNA and RNA. (a) Best known for their role in gene silencing, all DNA methyltransferase family members are able to interact with both RNA and DNA.[106,109,110] DNA (cytosine-5)-methyltransferase 1 (DMNT1) and DMNT3 play a part in initiating and maintaining DNA methylation, whereas DNMT2 methylates tRNAs. This modification is critical for maintaining tRNA stability and cell viability. (b) The evolution of the three major DNMTs is depicted in a cladogram. DNMT2 probably diverged from its ancestral DNA methyltransferase to perform a critical role in methylating tRNAs, a function which it performs redundantly with NSUN2 (reference 639). This radical change in substrate specificity highlights the ability of evolution to reshape a DNA-binding interface into one that preferentially recognizes RNA.

1.7 Conclusion and perspectives

In this analysis, we have demonstrated that DRBPs constitute a significant fraction of cellular proteins - perhaps 2% of the human proteome - and have important cellular roles. Their functions include the control of transcription and translation, DNA repair, mediating responses to stress, splicing and apoptosis. These functions are intimately linked to the structures of DRBPs: orthogonal binding of DNA and RNA provides an opportunity for competitive regulation of transcription by decoy RNAs, whereas simultaneous binding of DNA and RNA permits transcriptional activation by RNA co-activators or allows the recruitment of RNA-containing complexes to specific DNA loci. In turn, the structures underlying DRBP functions are linked to their evolution. Some DRBPs contain ancient domains that have long bound DNA or RNA; others contain multiple domains that separately confer DNA- and RNA-binding abilities and mediate their functional roles.

The majority of RNA-binding proteins have had remarkably similar motifs during evolution,[641] although individual members of protein families, such as the forkhead box transcription factors, can have diverse nucleic acid sequence specificities arising from independent evolutionary events.[642] It is also worth noting that intrinsically disordered protein domains that do not fold into defined secondary structures may also play important parts in mediating nucleic acid binding,[14] as was found for RNA chaperones.[643] In addition to protein evolution, nucleic acid sequence evolution has important roles in the development of DRBP function. The discovery of lincRNAs illuminated new cellular binding targets for proteins that were previously thought of as DNA-specific binding proteins. Tens of thousands of human lincRNAs have been catalogued,[644] and it is likely that many of them have yet-undiscovered functions requiring binding to proteins that are currently considered as DNA-specific binding proteins or that have so far only been shown to bind RNA *in vitro*. For example, the GR and the estrogen receptor were shown to bind DNA and RNA competitively

>20 years before a physiological role for RNA-steroid receptor interactions was established.[128,645,646] Experimental selection techniques, such as systematic evolution of ligands by exponential enrichment (SELEX), have been used to develop inhibitory RNA aptamers for DNA-binding proteins, such as NF-κB. If such inhibitory RNA binding is functionally advantageous, the rapidly evolving sequences of lincRNAs[647] could provide a platform for the evolution of an analogous endogenous function, and many DRBPs may have species-specific RNA targets. For example, the RNA Lethe, which binds NF-κB, exists only in mice and is not present even in the closely related rat genome.[648]

Proteins that bind both DNA and RNA could have several obvious functional advantages. By binding to both mRNAs and their encoding promoters, DRBPs can exert a powerful, amplified effect on gene expression. This also allows greater flexibility in generating cellular responses, as these DRBPs could both produce rapid effects on protein synthesis and impart long-acting changes on gene expression. At a cellular level, using one DRBP rather than two independent DNA-binding and RNA-binding proteins is more efficient, as it requires the transcription and translation of only one gene product. Finally, competitive RNA and DNA binding by some DRBPs allows an additional level of transcription factor regulation through decoy RNAs. These functional advantages, in addition to the rapid pace at which lincRNAs and their functions are being discovered, strongly indicate that more DRBPs and DRBP-mediated functions will be discovered in the coming years.

Chapter 2

Conserved, sequence-specific lincRNA-steroid receptor interactions drive transcriptional repression and direct cell fate

2.1 Abstract

The majority of the eukaryotic genome is transcribed, generating a significant number of long intergenic noncoding RNAs (lincRNAs). Although lincRNAs represent the most poorly understood product of transcription, recent work has shown lincRNAs fulfill important cellular functions. In addition to low sequence conservation, poor understanding of structural mechanisms driving lincRNA biology hinders systematic prediction of their function. Here we report the molecular requirements for the recognition of steroid receptors (SRs) by the lincRNA growth arrest-specific 5 (Gas5), which regulates steroid-mediated transcriptional regulation, growth arrest and apoptosis. We identify the functional Gas5-SR interface and generate point mutations that ablate the SR - Gas5 lincRNA interaction, altering Gas5-driven apoptosis in cancer cell lines. Further, we find that the Gas5 SR-recognition sequence is conserved among haplorhines, with its evolutionary origin as a splice acceptor site. This chapter demonstrates that lincRNAs can recognize protein targets in a conserved, sequence-specific manner in order to affect critical cell functions mediated by SRs.

This chapter is adapted from the manuscript:
Hudson WH, Pickard MR, de Vera IMS, Kuiper EG, Mourtada-Maarabouni M, Conn GL, Kojetin DJ, Williams GT, and Ortlund EA. Conserved sequence-specific lincRNA-steroid receptor interactions drive transcriptional repression and direct cell fate. *Nat Commun.* 2014 Nov; 5:5395.

2.2 Introduction

An emerging phenomenon in biology is the widening gap between the availability of genomic sequences and the functional understanding of the products encoded within.[649] The vast majority of functional research has focused on protein-coding genes, which constitute only 2% of the human genome.[650] Recently, attention has turned to the non-protein-coding elements of the genome and transcriptome. Although some non-coding RNA molecules (ncRNAs) such as small nucleolar and transfer RNAs have well-defined function, the majority of non-ribosomal, nuclear-encoded RNA molecules present in a cell are of unknown function and/or structure.[651] Of these, long intergenic ncRNAs (lincRNAs), defined as intergenic ncRNAs greater than 200 nucleotides in length, remain the most poorly understood product of transcription.[652] lincRNAs have been shown to function in diverse processes such as pluripotency, development, cell cycle, immunity, transcription and apoptosis.[204,653–656] Until recently, conservation of these lincRNAs was considered non-existent; however, emerging studies show higher conservation in lincRNAs than originally perceived, although well below the level of protein-coding exons.[653,657,658] Consistent with this observation, a recent study searched for similarity between mouse and zebrafish lincRNAs and found only 9.0% of annotated mouse lincRNAs contained detectable sequence homology to whole-genome zebrafish alignments, with most of this conservation clustered in small domains separated by longer stretches of divergent sequence.[659] Remarkably, mammalian lincRNA orthologues were still able to rescue developmental defects in zebrafish caused by specific lincRNA knockouts, indicating lincRNA function can be conserved despite divergent primary sequence.[659] However, the mechanisms by which lincRNA functions are conserved in a rapidly changing sequence space are unclear.

The study of such mechanisms requires the identification of functional domains within lincRNAs. Unfortunately, there are currently few lincRNAs with well-characterized function, and still fewer in which functional domains have been identified and char-

acterized. Known lincRNA functional domains include the 5' terminal region of lincRNA-p21, which interacts with hnRNP K to mediate the p53 response.[204] HOTAIR, a regulator of gene expression and chromatin state, uses a 5' domain to bind polycomb repressive complex 2, and a 3' domain to bind the LSD1/CoREST/REST complex.[660] Another lincRNA, growth arrest-specific 5 (Gas5), uses a region in its 12th exon to interact with and repress the steroid hormone receptors (SRs).[2] Without detailed structural information about these protein - lincRNA interactions, the basis of conservation (or lack thereof) within lincRNA functional domains is impossible to determine. By focusing on the interaction between SRs and Gas5, we sought to determine the molecular and structural requirements of a particularly important lincRNA-protein interaction.

The *GAS5* gene is a noncoding 5' terminal oligo-pyrimidine gene consisting of 12 exons in humans and forms two mature fully-spliced and polyadenylated isoforms in addition to numerous expressed sequence tags.[2,661] The 11 introns of the *GAS5* gene contain ten box C/D snoRNA genes, which are critical for ribosome modification and embryonic development.[662] Gas5 is present in all tissue types and is one of the most highly expressed noncoding RNAs in the human genome.[663] Gas5 has been implicated in many important cellular and physiological processes, including p53 signalling,[204] embryogenesis,[654] growth arrest and cell cycle,[664,665] and apoptosis.[2,666–668] In addition, Gas5 is required for rapamycin action in immune cells,[666] and Gas5 knockdown affects mRNA levels for M-Ras, prion protein, parkin and other genes.[654] Gas5 inhibits the transcriptional activity of steroid receptors (SRs) through direct competition for DNA binding.[2] The SR-binding site was previously mapped to a putative stem-loop region (bases 546-566) and described as a 'mimic' of genomic SR-binding sites, or glucocorticoid response elements.[2] This GRE mimic (GREM) within Gas5 represents a compact lincRNA-protein interaction domain, optimal for examining the effects of altered lincRNA sequence and structure on protein-RNA in-

teractions, and the downstream consequences on cellular processes. Here we describe detailed studies of the mechanistic, structural and evolutionary aspects of the SR - Gas5 interaction.

2.3 Results

2.3.1 The Gas5 GREM selectively binds 3-keto SRs

To measure the binding affinity between GR and Gas5, we monitored via fluorescence polarization the ability of full-length Gas5 and shorter GREM-containing constructs (Figure 2.1) to compete with fluorescently-labeled DNA activating glucocorticoid response elements, or (+)GREs, for binding to the glucocorticoid receptor (GR) DNA binding domain (DBD; Figure 2.2a). Full-length Gas5 competed for DNA binding the most strongly, with a K_i of 158 nM - within the range of affinities of GR for (+)GREs (80 nM - 1 μM).[669] Shorter constructs (Figure 2.1) containing only the GREM were sufficient for DNA competition, although affinity was lowered approximately five-fold compared to the full-length Gas5 transcript (Figure 2.2a). In addition, a fluorescently labeled Gas5 GREM construct containing only the stem bound directly to GR, indicating that only the putative double helical region within the GREM is required for GR binding (Figure 2.2b), much like GR - DNA binding requires a short, double helical response element.

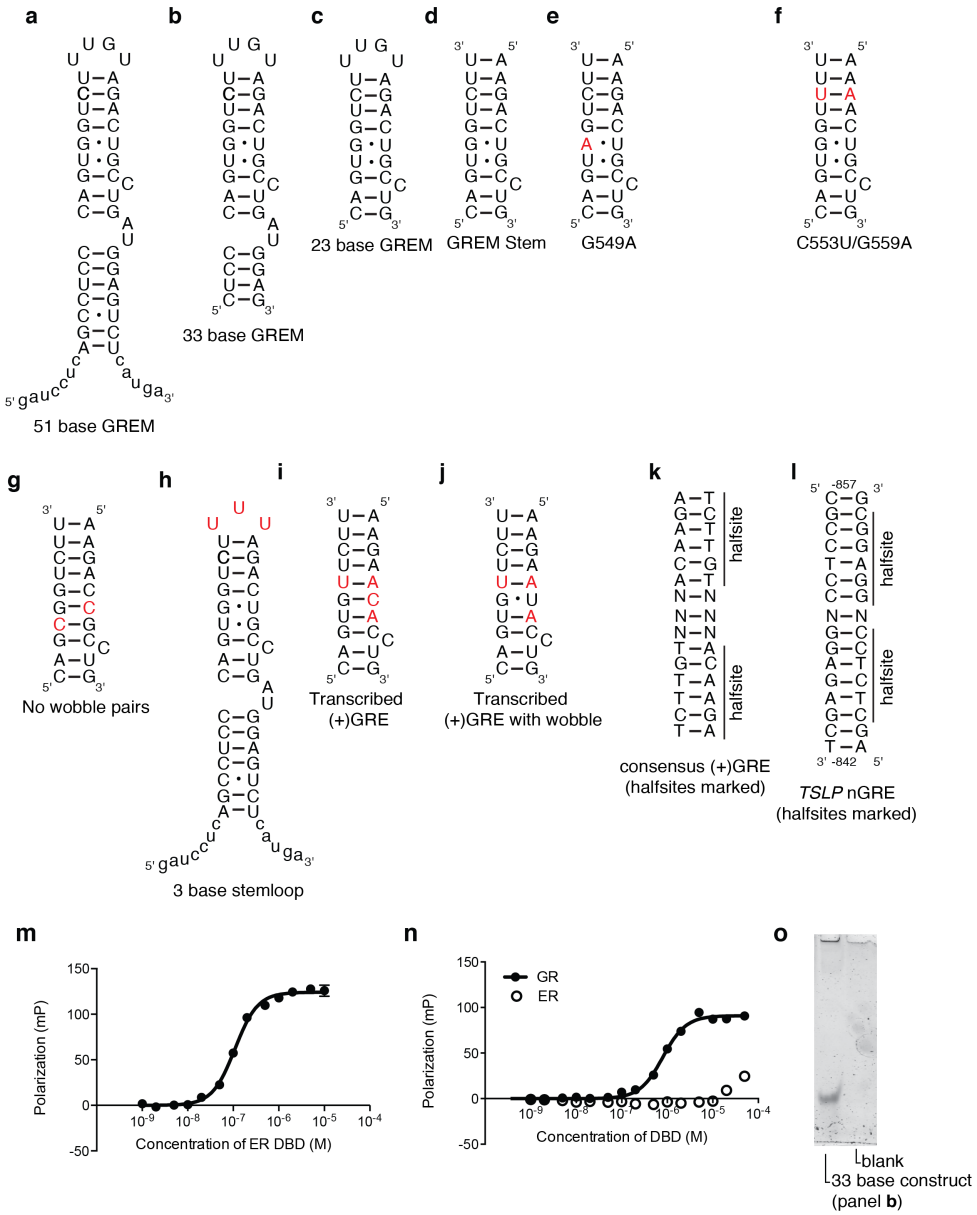


Figure 2.1: Human Gas5 constructs used in this study. (a) The 51 base GREM was used in OH• footprinting and UV melting. (b) Tested for binding in Figure 2.2a and used in NMR studies. (c) Tested for binding in Figure 2.2a. (d) Fluorescently-tagged and used in most WT binding assays, and unlabeled constructs used in crystallization. Mutants (red) of this construct (e-g) were used in Figure 2.6a-c. (h-j) Constructs used in Figure 2.8. The full-length human Gas5 construct used (Figure 2.2a and in all cellular assays) are identical in sequence to the NCBI reference sequence with the poly(A) tail removed (NR_002578.2 bases 1-631). Mutations from the WT sequence are in red. For comparison to the binding constructs, a consensus (+)GRE and the *TSLP* nGRE are shown in (k) and (l). (m) A DNA estrogen response element was used to confirm ER DBD activity; the experimental K_d for ER DBD - ERE was 109 nM, demonstrating that the recombinantly expressed ER DBD is capable of binding DNA. (n) However, the ER DBD does not bind to the Gas5 GREM. (o) Native TBE gel of the 33 base construct generated by IVT (from panel b).

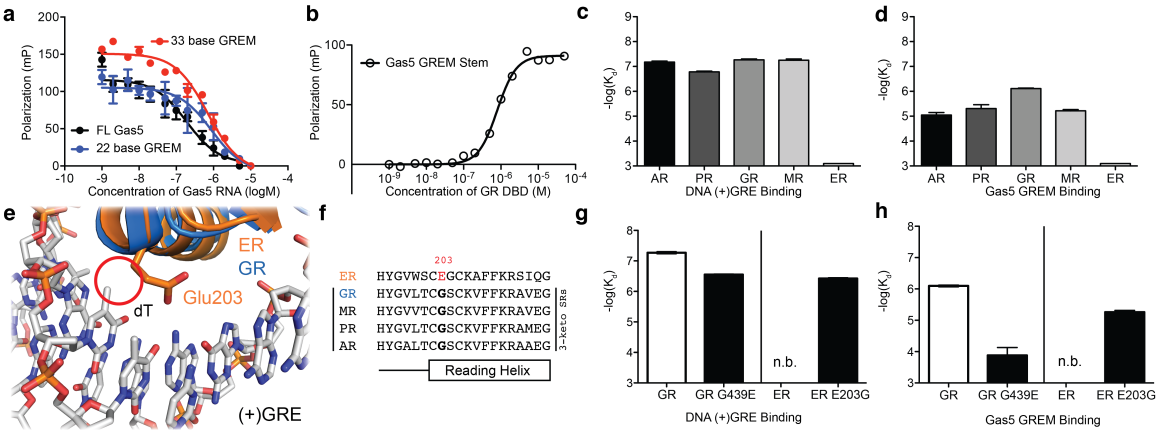


Figure 2.2: Gas5 selectively binds the 3-keto steroid receptors. (a) Full length Gas5 competes strongly for DNA binding to the glucocorticoid receptor (GR; K_i = 158 nM). Shorter Gas5 constructs (Figure 2.1) containing only the GREM stemloop are sufficient for GR binding (K_i for the 23 and 33 base constructs are 825 and 654 nM, respectively). (b) A double helical region containing only the GREM stem is sufficient for direct GR binding (K_d = 801 nM). (c) The 3-keto steroid receptors but not the estrogen receptor are able to bind to an activating DNA glucocorticoid response element ((+)GRE). (d) Similarly, the GREM stem construct shown in (b) binds the 3-keto steroid receptors, but not the estrogen receptor. However, the estrogen receptor is capable of binding its own response element (Figure 2.1m,n). (e) GR and ER recognize their cognate DNA response elements via different sequence-specific contacts. Notably, ER contains a glutamic acid at position 203 (Gly439 in GR), which excludes binding to an activating GR response element, or (+)GRE (f). (g) Mutation of this residue (Glu203) to the homologous GR residue (Glu203Gly) restores ER binding to a (+)GRE (right), and the converse mutation in GR (left, Gly439Glu) reduces affinity for a GRE. n.b., no binding. (h) Remarkably, the ER Glu203Gly mutation confers binding to Gas5 (right), demonstrating that Gas5 exploits differences in GR/ER nucleic acid specificity to selectively target the non-ER steroid receptors.

Previous work demonstrated that Gas5 regulates the transcriptional activity of not only GR, but also the androgen receptor (AR), progesterone receptor (PR), and mineralocorticoid receptor (MR), which share the ability to bind 3-keto steroids as well as DNA (+)GREs.[2] Other nuclear receptors such as the estrogen receptor (ER), PPAR, and RXR, which bind different DNA response elements, are not susceptible to Gas5 mediated repression.[2] To determine the relative affinity of the Gas5 GREM for all SRs, we tested the ability of recombinantly expressed DBDs from all five SRs to bind both (+)GRE DNA and the Gas5 GREM stem (Figure 2.2c,d). All 3-keto SRs bound both to a DNA (+)GRE and to the Gas5 GREM, which is consistent with their transcriptional repression by Gas5.[2] As predicted, the estrogen receptor DBD bound only to an estrogen receptor response element (ERE), and was unable to bind the Gas5 GREM (Figure 2.1m,n and Figure 2.2c,d).

2.3.2 Gas5 exploits sequence-specific differences in related proteins to discriminate among binding partners

Given their similar structure and sequence (>50% within the DBD), it is unclear how Gas5 discriminates among SRs; for example, the ER and GR DBDs share an identical fold and insert an α-helix to read the major groove of DNA (Figure 2.2e). However, key amino acid differences in this helix allow the two receptors to bind disparate response elements. Notably, one ER side chain, Glu203, makes contact with DNA when bound to an ERE, while the 3-keto SRs contain a glycine at this position (Figure 2.2f). When bound to a ERE, ER Glu203 contacts a cytosine, forming an interaction required for DNA binding and subsequent gene activation.[127] The 3-keto SRs do not contact DNA at this position, and substitution of glycine with a bulkier amino acid side chain would introduce steric clashes with (+)GRE DNA bases (Figure 2.2e).

Remarkably, mutation of this single amino acid in ER to glycine (Glu203Gly) confers ER binding not only to a DNA (+)GRE (Figure 2.2g), but also to the Gas5

GREM (Figure 2.2h). Likewise, the converse mutation in GR (Gly439Glu) reduced GR binding to both a DNA (+)GRE and the Gas5 GREM (Figure 2.2g,h). The discrimination among evolutionary-related proteins by Gas5 demonstrates that lincRNAs are capable of sequence-specific interactions with their target proteins. Additionally, the strong correlation between DNA binding and Gas5 RNA binding suggest the presence of sequence-specific, (+)GRE-like SR contacts in the Gas5 major groove.

2.3.3 The GREM forms a double helix with a widened major groove for SR binding

Previous work postulated that a double-helical region of the GREM mediated SR - Gas5 binding.[2] UV melting experiments confirm that the Gas5 GREM stemloop indeed forms a secondary structure in solution (T_m 58 °C; Figure 2.3a). To determine the secondary structure required for Gas5-SR recognition, we determined the x-ray crystal structure of the Gas5 GREM stem to a resolution of 1.9 Å. Two merohedrally twinned crystal forms were obtained in the rhombohedral space group R3, but twin refinement yielded excellent electron density and statistics (Figure 2.3b, Table 2.1 and 2.2). In both crystal forms, the Gas5 GREM adopts an A-form double helix with a widened major groove. The average minor and major grooves widths are greater than 15.0 Å in both crystal structures, wider than most A-form RNA helices found in crystal structures, but within the broader range of major groove widths in solution as determined by NMR33. GR-bound DNA response elements of GR exhibit major groove widths approximately 1.5 - 3.0 Å wider than the free Gas5 GREM structure,[333,669] indicating that only minimal secondary structure perturbation may be necessary for GR - Gas5 binding (Figure 2.3c,d).

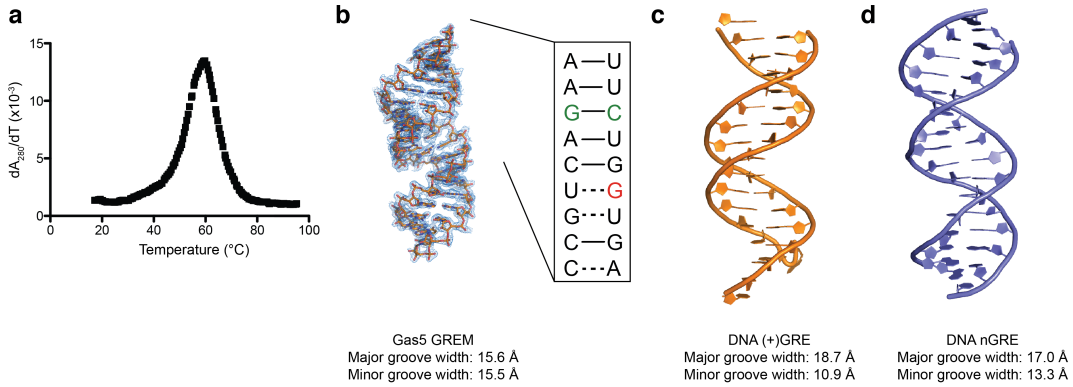


Figure 2.3: Gas5 forms a double helix in the region required for steroid receptor binding. (a) UV melting of the Gas5 GREM demonstrates that a strong secondary structure is present in the Gas5 GREM. (b) The stem of the Gas5 GREM was crystallized to a resolution of 1.9 Å. The blue mesh represents a composite omit map with simulated annealing contoured to 1 σ. The GREM contains a widened major groove compared to previous crystal structures of A-form RNA helices. This widened major groove imitates genomic binding sites for GR, including activating glucocorticoid response elements (panel c) and repressing negative glucocorticoid response elements (panel d). The DNA structures shown in (c) and (d) are GR DBD-bound.

	Gas5 GREM (4 Duplexes)	Gas5 GREM (2 Duplexes)
Data collection		
Space group	R3	R3
Cell dimensions (Å)		
a, b, c (Å)	43.3, 43.3, 304.2	43.5, 43.5, 173.2
α, β, γ(°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Resolution (Å)	1.9 (1.97 - 1.90)	2.2 Å (2.24-2.20)
R_{merge}	5.2 (42.3)	6.9 (25.6)
I / σI	22.5 (2.8)	28.5 (5.5)
Completeness (%)	99.7 (100.0)	99.8 (82.1)
Redundancy	2.9 (2.8)	3.5 (2.5)
Refinement		
Resolution (Å)	1.9	2.2
No. reflections	16794	6140
R_{work} / R_{free}	19.8 / 26.3	17.7 / 25.2
No. atoms		
RNA	1649	857
Water	114	85
B-factors		
RNA	31.6	34.6
Water	24.4	22.2
RMS deviations		
Bond lengths (Å)	0.0053	0.0065
Bond angles (°)	0.77	0.77
Coordinate error (Å)	0.26	0.40

Table 2.1: Data collection and refinement statistics for Gas5 GREM x-ray crystal structures. Data for each structure were collected from a single crystal. Values in parentheses are for highest-resolution shell. Maximum-likelihood coordinate error is shown.

	Gas5 GREM (4 duplexes)	Gas5 GREM (2 duplexes)
Twinning Tests (acentric):		
$< I^2 > / < I >^2$	2.928	2.333
$< F^2 > / < F >^2$	0.708	0.797
$< \|E^2 - 1\| >$	0.868	0.710
$< L >, < L^2 >$	0.515, 0.354	0.430, 0.263
Multivariate Z score L-test	3.370	8.612
Twin law	-h-k, k, -l	-h-k, k, -l
R_{obs}	0.017	0.031
Twin fraction		
Britton Analysis	0.449	0.446
H-test	0.488	0.478
Maximum Likelihood	0.478	0.457

Table 2.2: Twinning and Intensity Statistics for Gas5 GREM crystal structures. For acentric untwinned data, expected values are $< I^2 >= 2.000$, $< F^2 > / < F >^2 = 0.785$, $< |E^2-1| >= 0.736$, $< L >= 0.500$, and $< L^2 >= 0.333$. For acentric perfectly twinned data, expected values are $< I^2 >= 1.500$, $< F >^2 / < F^2 >= 0.885$, $< |E^2 - 1| >= 0.541$, $< L >= 0.375$, and $< L^2 >= 0.200$.

2.3.4 Molecular determinants and model of the GR - Gas5 interaction

To identify bases within the GREM for critical SR - Gas5 binding, we performed hydroxyl radical footprinting and used NMR to structurally map the GR - Gas5 interaction. Increasing concentrations of GR DBD protected several portions of the GREM stem from OH• hydrolysis while deprotecting bases U554-U556 within the loop (Figure 2.4a-c). The pattern of protection indicates GR interaction with the major groove, much like its classical interaction with DNA. To visualize the impact of GR binding on Gas5, we obtained two-dimensional homonuclear [^{1}H,^{1}H] NOESY spectra via ^{1}H NMR of the Gas5 GREM stemloop both in the presence and absence of the GR DBD (Figure 2.4d-f). Nucleotides within the Gas5 GREM stem showed NMR resonances with moderate or strong line broadening, whereas others displayed strongly shifted positions (>0.025 ppm). Broadened resonances correspond to nucleotides that are switching between two states in the intermediate exchange NMR timescale, whereas resonances that are simultaneously shifted and broadened are switching between two states in the slow-to-intermediate or intermediate-to-fast exchange regime. Only C561 and U569 have H1'-H6 resonances that are shifted but not broadened, and therefore, have two states that are in slow or fast exchange. The two states may correspond to nucleotides situated in disparate chemical shift environments, such as (1) free vs. GR DBD-bound Gas5 RNA, (2) two Gas5 RNA conformational states in the presence of GR DBD, or (3) as a result of interactions with a neighboring Gas5 RNA or with another Gas5 RNA region within the same molecule. Both NMR and OH• footprinting demonstrated conformational changes in the four base loop upon GR binding (Figure 2.4c), consistent with previous studies indicating that bulges and loops can widen the RNA major groove, presenting surfaces for protein interaction.[670,671]

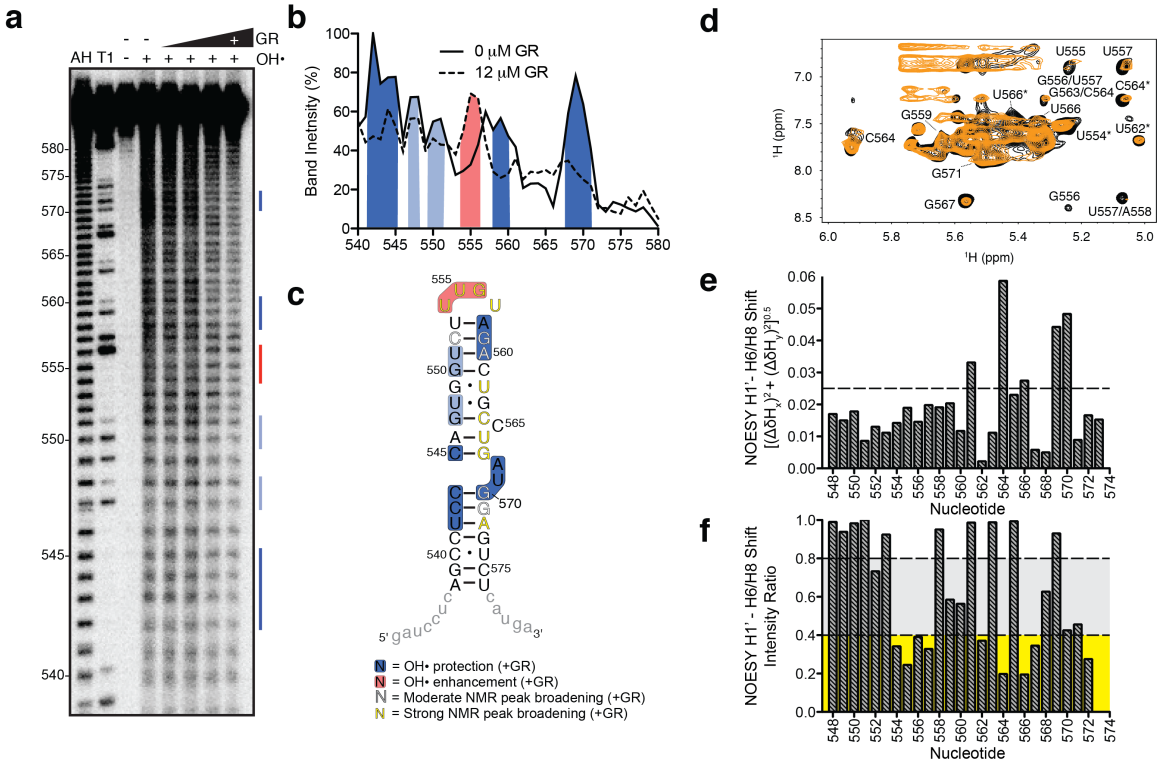


Figure 2.4: Base specific-interactions mediate the glucocorticoid receptor - Gas5 interaction. (a) GR protects the Gas5 GREM from OH• hydrolysis. Increasing amounts of GR DBD were added to *in vitro* transcribed Gas5 GREM RNA and subjected to OH• hydrolysis, followed by denaturing polyacrylamide gel electrophoresis. GR-mediated protection from OH• hydrolysis is quantified in (b). Light and dark blue indicate moderate and strong protection with GR DBD, and red indicates deprotection in the presence of GR DBD. (c) Results from NMR and OH• footprinting are mapped onto the Gas5 GREM. (d) Zoomed in overlay of free Gas5 RNA (black) and 1:1 Gas5 RNA:GR DNA binding domain (DBD; orange) 2D [^{1}H,^{1}H]-NOESY spectra with 200 ms mixing time. Broadened resonances in the presence of GR DBD are labeled. Intra-nucleotide H1'-H6/H8 resonances are labeled according to nucleotide type and position number. H5-H6 peaks are labeled in the same way, followed by an asterisk (*). Peak labels separated by a slash (/) are for inter-nucleotide H1'-H6/H8 NOESY crosspeaks. (e) NMR crosspeak shifting calculated from NOESY spectra for free and GR DBD-bound Gas5 GREM stem reveals strongly shifted resonances (changes >0.025 ppm; horizontal dashed line). (f) NMR crosspeak line broadening determined from NOESY spectra as calculated from the crosspeak intensity ratio for free and GR-DBD bound Gas5 GREM stem (moderately broadened resonances, gray; strongly broadened resonances, yellow).

To determine the effects of Gas5 binding on GR, we collected two-dimensional [^{1}H,^{15}N]-HSQC data on the GR DBD both free and bound to the Gas5 GREM (Figure 2.5a-c). Notably, chemical shift perturbations are observed in helix 1 of the DBD when bound to either DNA[672] or Gas5 RNA, indicating that GR uses its DNA binding interface to bind RNA (Figure 2.5b-e). Additional regions of the GR DBD that contact DNA in a non-specific manner, such as helix 2 and the N-terminus, are also perturbed when GR is bound to DNA and RNA. However, no chemical shift perturbations are observed in the dimerization interface (D-loop) of the DBD upon RNA binding, suggesting that GR binds to Gas5 as a monomer (Figure 2.5c). This is in direct contrast to DNA binding, which induces positive cooperativity and dimerization of the receptor and causes significant chemical shift perturbations in the D-loop upon DNA binding (Figure 2.5e).[672] These GR DBD [^{1}H,^{15}N]-HSQC data, Gas5 GREM [^{1}H,^{1}H] NOESY spectra, and mutagenesis data were used to generate a model of the GR DBD - Gas5 GREM complex using HADDOCK.[673] The top scoring models showed the GR DBD reading the Gas5 GREM by inserting helix 1 into the major groove (Figure 2.6a).

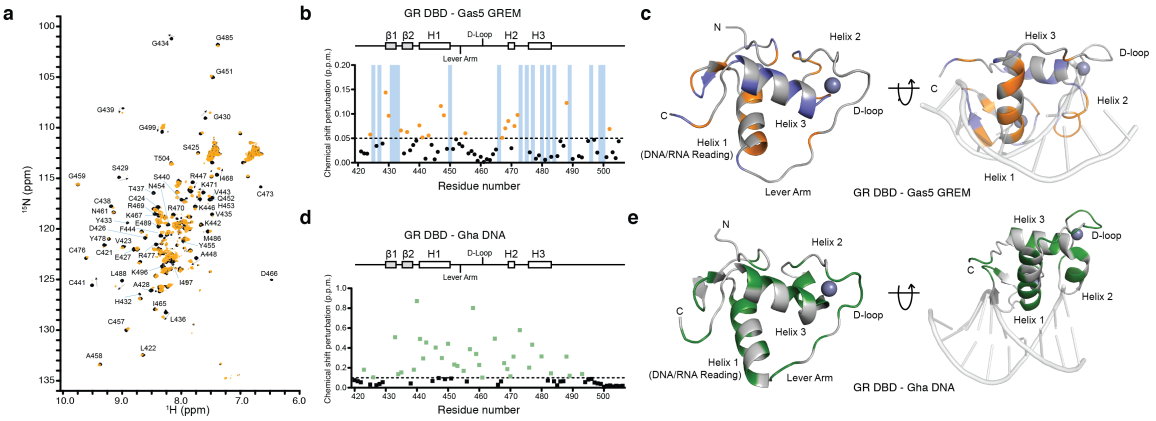


Figure 2.5: NMR reveals Gas5 interacts directly with residues that comprise GR's DNA-binding interface. (a) 2D [^{1}H,^{15}N]-HSQC spectra of GR DBD free (black) and bound to the Gas5 GREM (orange). (b) Chemical shift perturbations induced in GR DBD upon binding Gas5. Perturbations greater than 0.05 ppm are colored orange; blue vertical lines represent resonances that broaden upon binding Gas5 and are no longer observed. (c) GR chemical shift perturbations upon GREM binding mapped onto the HADDOCK model of GR DBD bound to the Gas5 GREM. (d) Chemical shift perturbations greater than 0.1 ppm (green) induced in GR DBD upon binding Gha DNA[672] plotted against GR DBD primary and secondary structure. (e) GR chemical shifts upon Gha DNA binding mapped onto the GR DBD crystal structure (PDB 4HN5).[669]

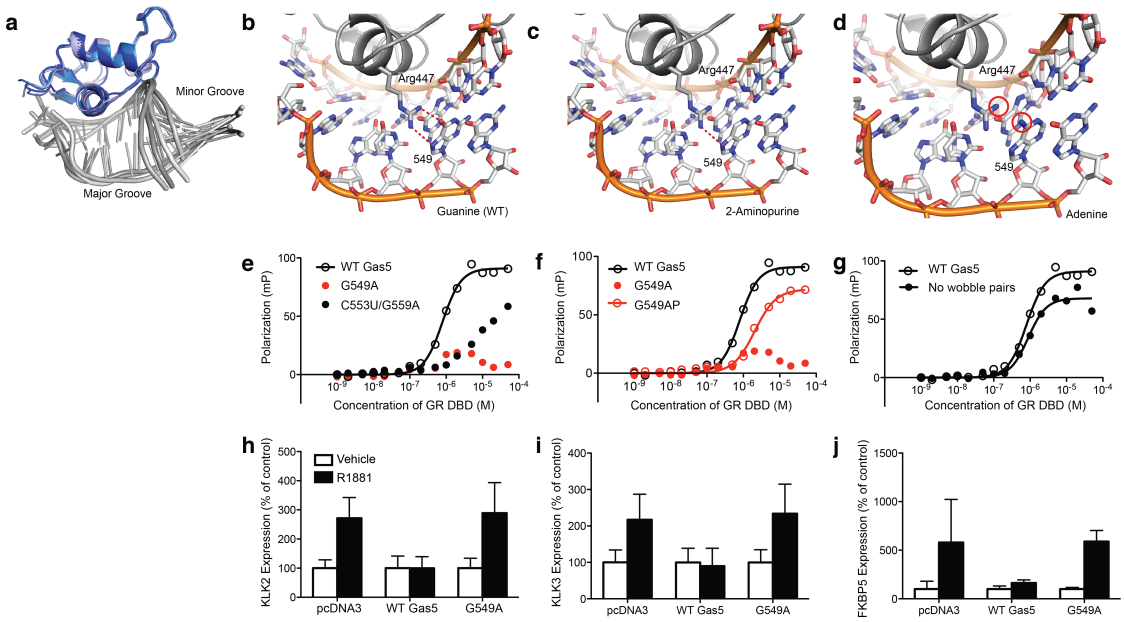


Figure 2.6: A single nucleotide mutation within Gas5 eliminates its ability to bind steroid receptors and repress steroid-driven transcription. (a) Superposition of the highest-scoring cluster of GR - Gas5 GREM models generated by HADDOCK. GR binds in the major groove to the A-form helix of the Gas5 GREM. (b) Guanine 549 presents a carbonyl moiety toward the major groove, which GR may recognize with a basic amino acid such as Arg447 (modeled here by HADDOCK) or Lys442. Mutation of this guanine to 2-aminopurine (c) removes this carbonyl moiety without affecting the minor groove interaction surface. (d) Mutation to adenine introduces an amine in the major groove, which would clash with the basic major-groove reading amino acids of GR. (e) Mutation of the two guanine bases identified by structure probing (Figure 2.5) to adenine ablates binding of GR to the Gas5 GREM stem. (f) The G549A mutation ablates GR binding to the Gas5 GREM stem. However, mutation of G549 to 2-aminopurine reduces GR - Gas5 GREM stem binding only 2.5-fold, indicating that GR contacts the major groove of Gas5 (see panels b-d). (g) Two G-U wobble pairs (including G549) are present in the GREM; however, Watson-Crick G-C pairs are also compatible with GR binding. (h-j) The G549A mutation, which ablates SR - Gas5 binding *in vitro*, also compromises Gas5-mediated repression of steroid-driven transcriptional activation in 22Rv1 cells. Three androgen receptor-responsive genes were measured and normalized to ACTB via RT-PCR in 22Rv1 prostate cancer cells with and without the AR agonist R1881.

2.3.5 A single nucleotide mutation abolishes the SR - Gas5 interaction

Glucocorticoid receptor and 3 keto SR - DNA recognition and gene activation is heavily dictated by the contact of arginine and lysine residues in helix 1 of the GR DBD with two guanine-cytosine (GC) pairs in the (+)GRE major groove. Surprisingly, two equivalently spaced guanine-containing pairs (G549 and G559) are present in the Gas5 RNA GREM and were identified by structure probing as critical for the GR - Gas5 interaction (Figure 2.4). Mutating these two guanines to adenine, which replaces the major groove O6 carbonyl with an amine, ablates GR binding (Figure 2.6b-f). Mutation of G549 to 2-aminopurine, which selectively removes a potential hydrogen bond acceptor (O6) from the major groove while leaving the minor groove unchanged, diminishes but does not abolish binding. Critically, these Gas5 RNA mutations preserve secondary structure as monitored by UV melting (Table 2.3). Along with the structural data above, these experiments establish that GR indeed binds Gas5 in the major groove. Additionally, protein recognition of RNA wobble pairs has been reported in several systems,[674,675] and two such non-Watson-Crick interactions are present in the Gas5 GREM: U548-G563 and G549-U562. To determine if GR-GREM recognition depends on this wobble base pairing, we mutated these G-U pairs to canonical Watson-Crick G-C pairs. This mutation only marginally affected GR binding (Figure 2.6g), indicating that GR does not discriminate between guanine-uracil pairs in Gas5 and guanine-cytosine pairs in DNA (+)GREs. To examine the effect of the G549A mutation on SR-mediated transcription in cells, we transfected AR^+ 22Rv1 prostate cancer cells with WT Gas5, Gas5 with the G549A mutation, or empty vector (pcDNA3.1). In these cells, WT Gas5 inhibited the androgen-driven expression of several AR target genes (Figure 2.6h-j). However, this effect was reversed by the G549A mutation, with expression of the AR target genes returning to vector control levels. These results demonstrate that a single nucleotide mutation is

sufficient to ablate the interaction between SRs and Gas5 both *in vitro* and in cells.

Mutation	Effect on T_m compared to WT Gas5 (°C)
G549A	-2.1
G549AP	-3.8
No wobble pairs	+14.6
Transcribed (+)GRE	-0.5

Table 2.3: Effects of GREM mutations on Gas5 melting temperature. Melting of WT is shown in Figure 2.3a.

2.3.6 Ablation of the Gas5 - SR interaction prevents Gas5-induced apoptosis in steroid-dependent cancer cell lines

Gas5 is downregulated in both breast and prostate cancers, and its reintroduction accelerates apoptosis in both prostate and breast cancer cell lines.[667,668,676] However, the potential role of the Gas5 - SR interaction in cancer cell proliferation and survival has not been studied. To test whether a Gas5 mutant specifically compromised in SR interaction displays aberrant function in SR-driven cancer cell lines, we tested the ability of both the WT and G549A Gas5 to drive apoptosis when expressed in 22Rv1 prostate cancer cells. Transfection of both the WT and mutant constructs increased Gas5 expression compared to the vector control (Figure 2.7a), but only the WT sequence caused caspase activation in these cells, and this was accompanied by a corresponding decrease in cell viability (Figure 2.7b,c). The G549A mutant, in contrast, had no effect on either parameter (Figure 2.7b,c). We repeated these experiments in the PR^+/ER^+ breast cancer cell line MCF-7; again, the G549A mutant completely prevented the pro-apoptotic phenotype induced by transfection of WT Gas5 (Figure 2.7d-f). Thus, in steroid-driven cancer cells such as 22Rv1 cells, Gas5 function appears to be directly related to the SR - Gas5 signaling axis.

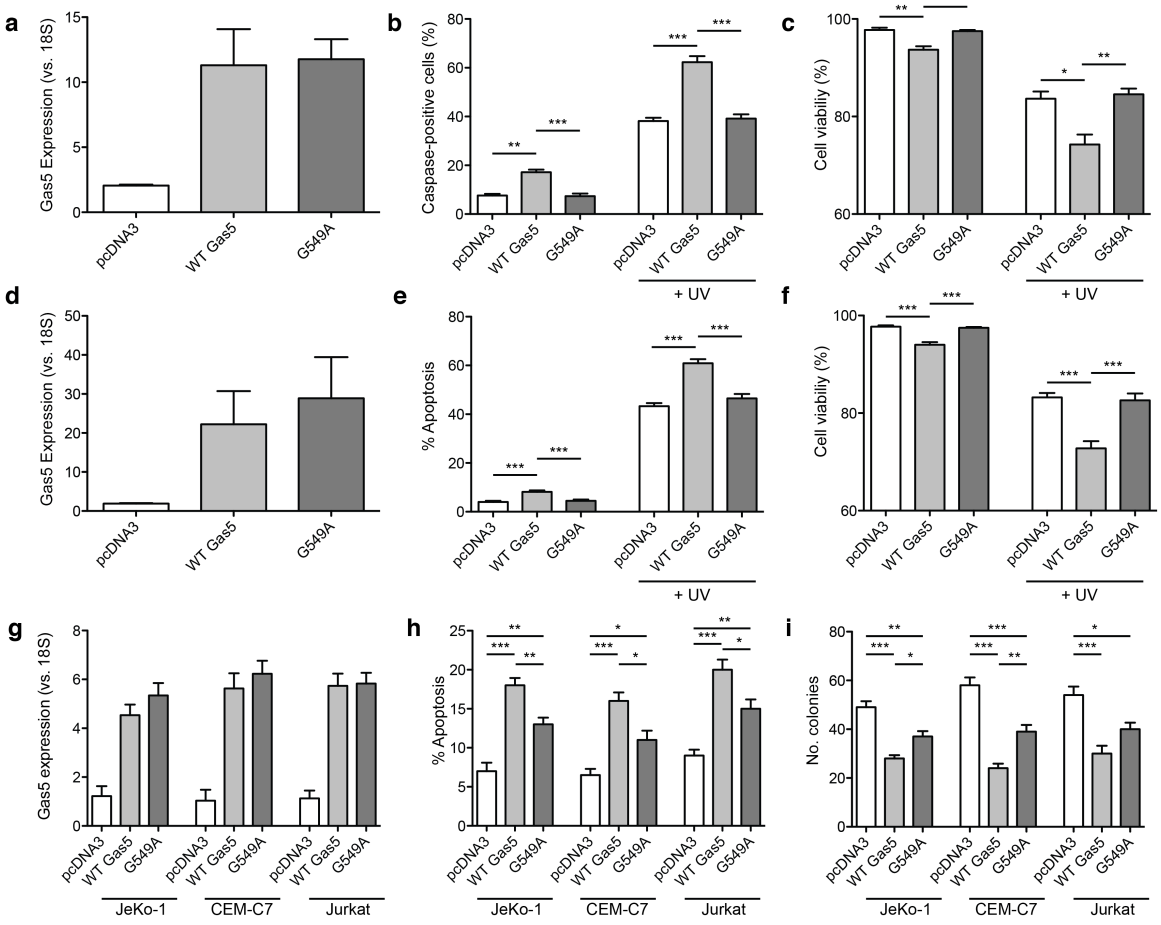


Figure 2.7: Selective ablation of the steroid receptor - Gas5 interaction has dramatic effects on cell fate. (a) Quantification of Gas5 levels normalized to 18S RNA in 22Rv1 cells after transfection with WT Gas5, G549A Gas5, and pcDNA3 constructs. (b) WT Gas5 induces caspase-dependent cell death in 22Rv1 cells, but this effect is completely reversed in the G549A mutant. Similar effects were seen for cell viability (panel c). (d) MCF-7 cells were transfected with identical constructs, and levels quantified via RT-PCR and normalized to 18S RNA. The G549A mutation completely ablated the pro-apoptotic (e) effects and reduction in cell viability (f) caused by transfection of WT Gas5. (g) Likewise, three suspension cell lines were transfected with pcDNA3, WT Gas5, or the G549A mutant and expression were quantified as above. In these cell lines, the G549A mutation only partially reverses the apoptotic effects (h) and colony-forming effects (i) seen by transfection of WT Gas5.

The expression of Gas5 variants has differential effects in immune cells which are less dependent on SR signaling for growth.[665] To determine the role of SR - Gas5 interactions in apoptosis and viability in these cells, we transfected three suspension cell lines, JeKo-1, CEM-C7, and Jurkat, with the WT and G549A Gas5 constructs (Figure 2.7g). Among all three cell lines, transfection of WT Gas5 induced apoptosis and a corresponding decrease in cell viability (Figure 2.7h,i) as previously reported.[665] Interestingly, the G549A mutation only partially reversed this phenotype (Figure h,i), in contrast to the complete reversal of apoptosis induction seen in steroid-driven cancer cell lines (Figure 2.7b-c, e-f). These results indicate that additional functions and regions of the Gas5 transcript play important roles in regulating apoptosis in immune cells.

2.3.7 GR is incapable of binding its own transcribed genomic binding sites

Since 70% of the genome is transcribed, there may be many Gas5-like RNAs capable of binding and modulating the activity of SRs. In particular, there are 2.3×10^6 genomic sequences matching a GR-binding motif,[677] and many of these lie within transcribed regions. Due to their inverted repeat sequence, transcribed (+)GREs would form a stemloop structure much like the Gas5 GREM (Figure 2.8a), which led us to inquire whether an RNA GREM could readily evolve from a DNA GR binding site. However, a (+)GRE-derived stemloop differs from the Gas5 GREM in two ways. First, the loop of the (+)GRE structure would contain a 3-base loop, whereas the GREM contains a 4-base loop. Second, the (+)GRE sequence differs slightly from the bases found in the Gas5 GREM (Figure 2.8a).

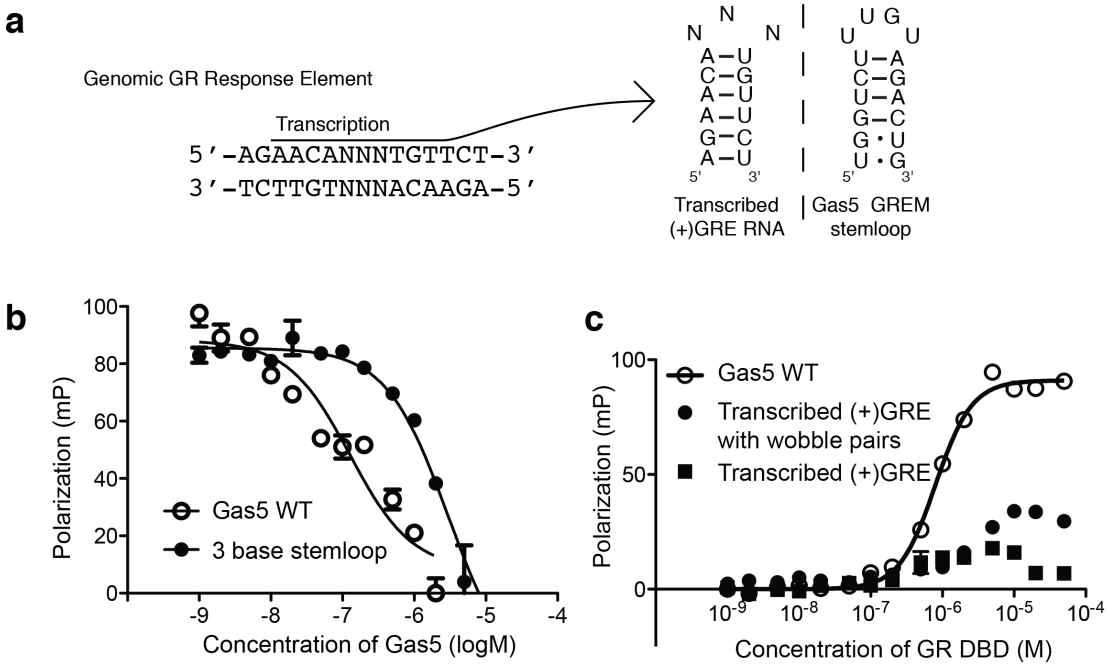


Figure 2.8: The glucocorticoid receptor will not bind its transcribed genomic response elements. (a) Within genomic DNA, GR binds (+)GREs, which are inverted repeats separated by three nucleotides. Such inverted repeat sequences would form a stemloop with similar sequence to the Gas5 GREM when transcribed (on right). However, this sequence is incapable of binding GR due to both secondary structure constraints caused by a 3 base stemloop (b) as well as a different primary sequence (c). (b) Competition for a fluorescently-labeled (+)GRE DNA construct was performed with a GREM-containing 50 nt construct of WT or the same construct with a 3-base stemloop (ΔG556). (c) A fluorescently-labeled transcribed (+)GRE sequence with Gas5 GREM stem flanking sequences was tested for binding to the GR DBD.

To test the effects of the decreased-length loop on GR binding, which would occur upon transcription of a (+)GRE, we generated the WT Gas5 GREM with a 3-base loop UUU loop rather than the 4-base UUGU loop found in WT Gas5. Although 3-base loops are often less stable than 4-base loops, a UUU loop with a closing AU pair demonstrates adequate stability.[678] Nonetheless, the 3-base loop Gas5 GREM showed dramatically lower affinity for GR than the WT 4-loop structure, indicating that inverted repeat sequences with 3-base spacers are unlikely to be active GR-binding elements following transcription (Figure 2.8b).

However, it is possible that a SR-binding RNA sequence such as the GREM could form via a single nucleotide insertion into a (+)GRE, creating a four base stemloop for SR binding. To eliminate effects of the loop length, we placed the consensus (+)GRE halfsite within the Gas5 GREM stem, which is sufficient for GR binding (Figure 2.2b). GR could not bind this consensus sequence with either Watson-Crick or Gas5-like wobble pairs (Figure 2.8c), but required additional sequence changes to achieve GR-binding ability (Figure 2.9). These results indicate that Gas5 independently and convergently evolved a SR-binding surface, and we sought to establish the evolutionary origin of the Gas5 GREM.

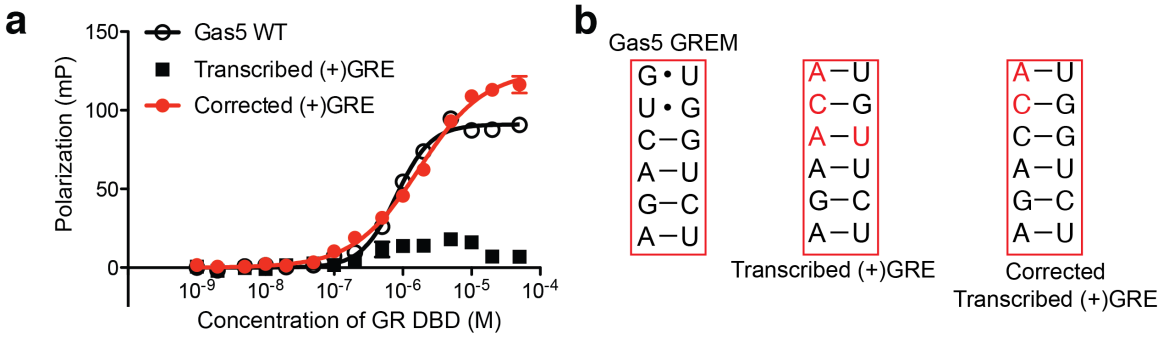


Figure 2.9: Additional primary sequence changes are required to convert a transcribed DNA (+)GRE into a GR binding site. (a) Binding of Gas5 GREM stems to the GR DBD. (b) Sequences showing the WT Gas5 GREM (left), a transcribed (+)GRE (middle), and the additional change required to correct a transcribed (+)GRE to achieve GR binding (right; panel a).

2.3.8 The Gas5 GREM evolved from a splice acceptor site

The *GAS5* gene is annotated in only three genomes in the *Ensembl* database, and the low sequence conservation of lincRNAs makes sequence alignment across multiple species challenging. Fortunately, Gas5 hosts ten snoRNAs in humans that are highly conserved throughout vertebrates. Intron 11, just upstream of the Gas5 GREM, contains the snoRNA gene *U81*, which is positionally conserved from reptiles to mammals. The length of intronic sequence from snoRNAs to the downstream 3' acceptor splice site is restricted for proper processing of box C/D snoRNAs, with lengths of 65-75 nucleotides from a box C/D snoRNA to the downstream exon ideal for efficient snoRNA excision.[679] Since Gas5-containing snoRNAs are critical for embryonic development,[662] we expected the length of intronic sequence downstream of these snoRNAs to be well-conserved. Using this anticipated high conservation of snoRNA sequence and intronic length as a guide, we were able to align the Gas5 GREM sequences from over 40 vertebrates. Gas5 does not appear to be present in jawless fish. As expected, the distance from *U81* to the downstream splice acceptor site was tightly clustered across vertebrates (Figure 2.10a).

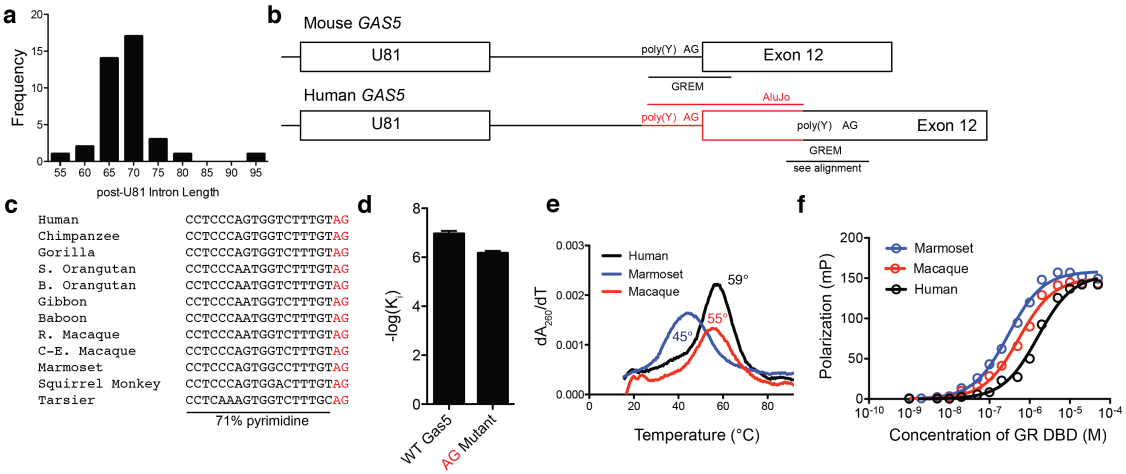


Figure 2.10: (a) The downstream intron length from the snoRNA *U81* is highly conserved due to excision constraints, illustrating the importance of sequence and positional conservation of the splice acceptor site at this locus. (b) The Gas5 GREM was interrupted by an AluJo insertion (red) in haplorhines, moving the ancestral splice acceptor site into exon 12, where splicing no longer occurs. (c) However, despite the lack of splicing at the haplorhine GREM, the haplorhine GREM maintains a splice-site like sequence with a polypyrimidine tract and AG sequence, which is required for high-affinity steroid receptor - Gas5 binding (mutated to GA in panel d). Other haplorhine species contain Gas5 GREM stemloops that exhibit similar secondary structure (e) and GR-binding affinity to the human Gas5 GREM (f).

The most striking feature of the *GAS5* GREM alignment is the presence of an Alu insertion in the haplorhine lineage, belonging to the AluJo family (Figure 2.10b). This insertion occurred directly at the *GAS5* intron 11/exon 12 boundary, lengthening the *GAS5* exon 12 by 100 nucleotides compared to mice. In haplorhines, an Alu-contained splice acceptor site is used for Gas5 transcript processing, maintaining the strict requirement for intron length downstream of *U81*. However, this insertion demonstrates that the GREM, while exonic in humans and other haplorhines, evolved from an ancestral splice acceptor site still present in other vertebrates. Multiple previously unassigned cDNAs from GenBank support both the alignment as well as the presence of a splice site at the GREM (Table 2.4). The AluJo insertion repositioned the ancestral splice acceptor site to be fully contained within exon 12, and it is this now-defunct splice site through which the SRs bind Gas5 (Figure 2.10b).

Species	GenBank Accession	Spliced at GREM?
Baboon	GE873351	No
Chicken	XM_001234629	No
Chicken	CR387333	Yes
Chimpanzee	XR_126956	No
Cow	GW424764	No
Dog	CO680308	Yes
Dog	CF411110	No
Gibbon	XR_122856	No
Gibbon	XM_003259120	No
Horse	DN505807	No
Horse	DN509669	Yes
Human*	NR_002578	No
Macaque (fascicularis)	DW525031	No
Macaque (mulatta)	CB312793	No
Marmoset	EH380662	No
Mouse*	X59728	Yes
Mouse*	BB732963	No
Pig	EW253363	No
Pig	EW573340	Yes
Platypus	ENSOANG00000020492 (Ensembl)	Yes
Rabbit	CU465388	Yes
Rat*	NR_002704.1	Yes
Wallaby	FY571474	Yes
Zebra finch	DQ214391	Yes
Zebrafish*	CT722790	No
Zebrafish*	DT877503	Yes

Table 2.4: Previously unassigned Gas5 transcripts from vertebrates confirm the presence of an alternatively spliced acceptor site at the Gas5 GREM in non-haplorhines. Entries marked with an asterisk indicate previous identification as a Gas5 transcript.

Due to snoRNA excision restraints, the haplorhine Gas5 GREM, over 150 bases downstream of *U81*, would be under negative selection to lose the ancestral splice site features located inside exon 12 incorporated with the AluJo insertion. However, if this locus was selected for both its splice site sequence and SR binding ability before the Alu insertion, it is possible that splice-site like sequences required for SR binding would be conserved. Remarkably, when haplorhine GREM sequences are aligned, splice site-like features are readily apparent: a polypyrimidine tract and acceptor AG sequence are conserved throughout the haplorhine clade, despite a lack of splicing at this locus (Fig 2.10c). The splice-site like sequence shows higher sequence identity among 14 haplorhines (93.8% versus 84.5% for the remainder of exon 12). Additionally, relative rate analysis of each position within the ancestral splice site reveals a lower rate of nucleotide substitution within the GREM versus the remainder of exon 12, supporting sequence conservation of the splice-site like sequences required for GR binding to the GREM (Figure 2.11). Mutation of these splice-site like features, such as the AG within the ancestral splice acceptor sequence, reduces the GREM's affinity for GR binding (Fig 2.10d), suggesting that other haplorhines contain conserved, functional GREMs with both a primary sequence and secondary structure capable of binding SRs. Indeed, the GREM sequences of two other haplorhines formed secondary structures in solution (Fig 2.10e) and were capable of binding the GR DBD (Fig 2.10f).

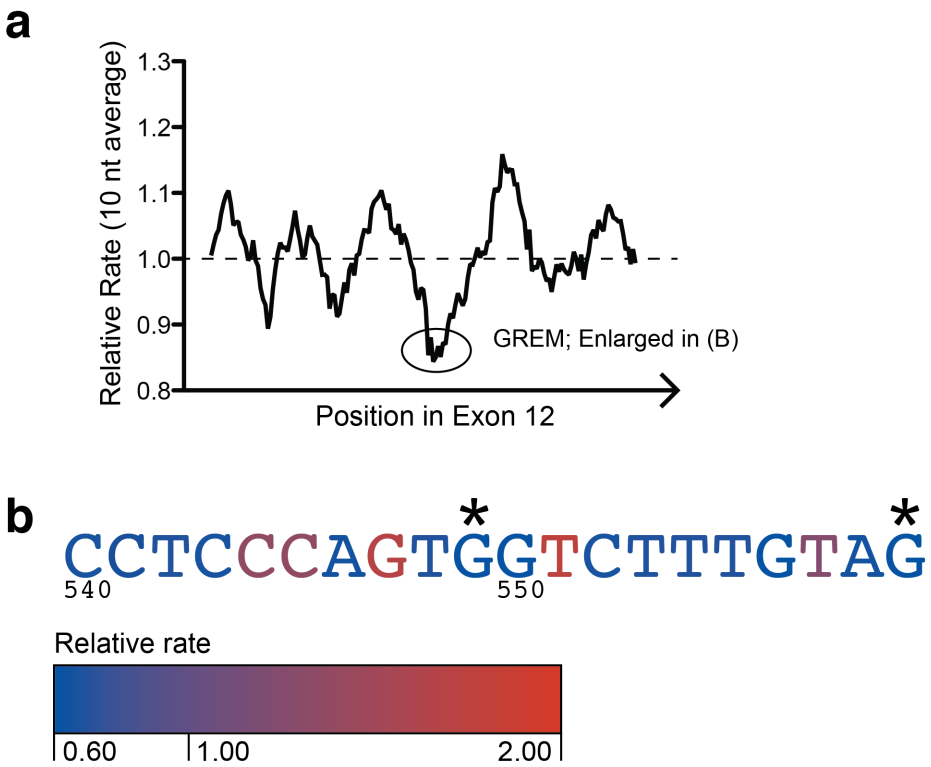


Figure 2.11: Relative rate analysis supports conservation of the ancestral splice site at the haplorhine Gas5 GREM. (a) Maximum likelihood relative rate analysis was performed on fourteen haplorhine genomic Gas5 exon 12 sequences (exon 12 average = 1.00). Plotted is relative rate (10-neighbor average) versus position in exon 12; the region of exon 12 with lowest sequence turnover is the GREM splice-site like sequence, shown in part (b). Guanine bases required for GR binding are marked with an asterisk. The average relative rate for each nucleotide within the ancestral splice site (20 nt) is 0.85, compared to 1.00 across exon 12 as a whole. The median relative rate for each nucleotide within the ancestral splice site is 0.68, compared to 1.00 for exon 12 nucleotides outside the ancestral splice site. Genomic sequences used for analysis were *Callithrix jacchus*, *Chlorocebus aethiops*, *Gorilla gorilla*, *Homo sapiens*, *Macaca fasicularis*, *Macaca mulatta*, *Nomascus leucogenys*, *Pan paniscus*, *Pan trogolodytes*, *Papio anubis*, *Pongo abelii*, *Pongo pygmaeus*, *Samiri boliviensis*, and *Tarsius syrichta*.

Given the evolutionary origin of the human Gas5 GREM as a splice acceptor site, it is possible that alternate splicing of the *GAS5* gene may control its ability to bind and repress GR in non-haplorhines. The mouse splice isoforms with exon 11 excised or retained bound GR *in vitro* (Figure 2.12a), and the GREM derived from both the splice or unspliced Gas5 transcripts was sufficient for secondary structure formation and GR binding (Fig 2.12b,c). When we tested for GR - Gas5 binding across the vertebrate clade, eutherian mammals demonstrated the highest affinity interaction (Fig 2.12d), suggesting that the Gas5 GREM was perhaps capable of binding SRs at the time of the AluJo insertion, roughly 55 Mya.[680] However, further study of the mouse and other mammalian Gas5 transcripts is required to definitively determine if GR regulation is a conserved property of Gas5 outside of the haplorhine clade. This is especially true given the short lifetime of unspliced Gas5 and the relative shortening of exon 12 compared to haplorhine Gas5 (Figure 2.10b).

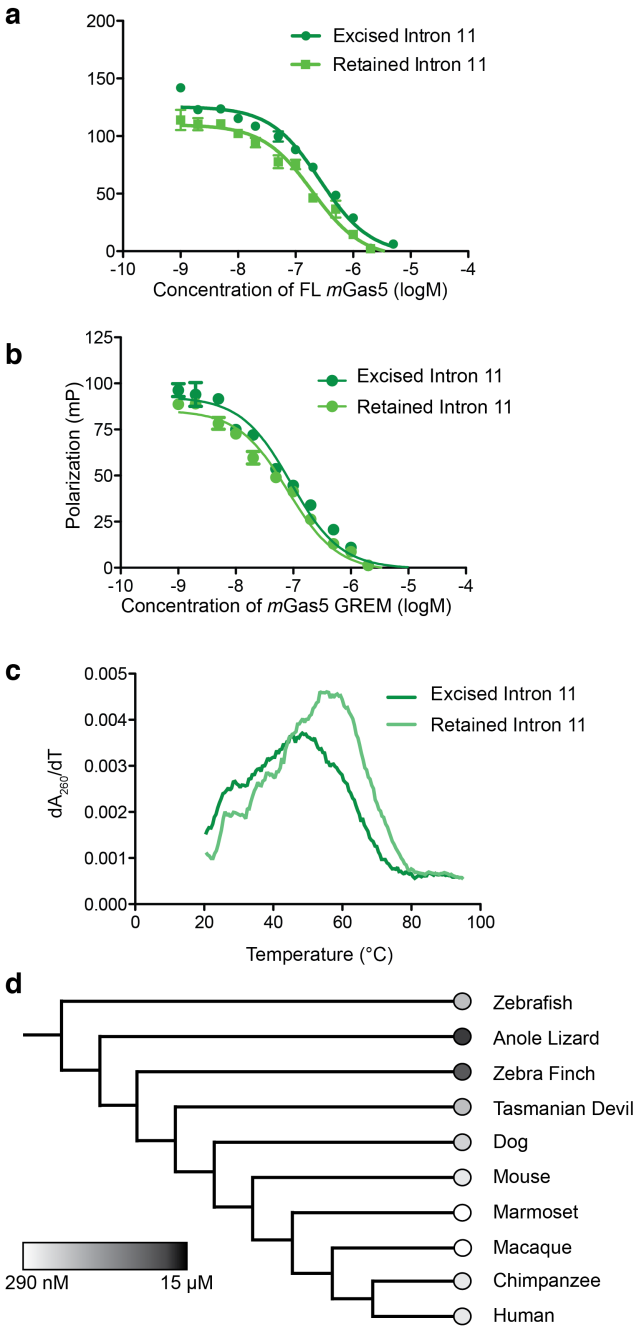


Figure 2.12: Non-haplorhine *GAS5* gene products retain the ability to bind GR. (a) Full length constructs of mouse Gas5 are able to directly compete for GR binding with (+)GREs. K_i for the excised and retained intron 11 full length constructs are 253 and 185 nM, respectively. (b) Mouse Gas5 GREMs (60 base constructs) are sufficient to compete with (+)GREs for GR binding. K_i for the excised and retained intron 11 GREM constructs are 86 and 76 nM, respectively. (c) Both the spliced and unspliced mouse Gas5 GREMs form a secondary structure in solution. (d) Binding of the human GR DBD, which is 100% identical from reptiles to mammals, to fluorescently-labeled Gas5 GREMs from various species.

2.3.9 *GAS5* is a retrotransposed gene within multiple mammalian lineages

Finally, given the unique ability of the GREM to bind SRs, we searched for additional GREM-like sequences within the human genome which could serve as SR-repressing elements. Remarkably, we found processed copies of Gas5 in at least five locations in the human genome, including within the introns of four genes: *GABRB3*, *NAPE-PLD1*, *MDN1*, and *SV2C* (Figure 2.13). At least 29 splice variants of the *GAS5* gene are known,[681] and *GAS5* retrotransposition events mirror this complexity. Two human copies are fully spliced (splice variant *GAS5*-001; Figure 2.13), whereas the sequence within the *GABRB3* gene is a copy of the *GAS5*-008 splice variant (Figure 2.13b). Additionally, some retrotransposed copies of Gas5 are from unreported, species-specific, or ancestral splice variants, such as a copy in gibbons that contains exons 1-4, 6, 7, 11, and 12 (Figure 2.13). Such a splice variant is not currently described in humans.

GAS5 appears to be highly susceptible to retrotransposition, with either full or partial copies of the processed Gas5 present in multiple mammalian lineages including, mice, dogs and more distal members of the mammalian clade, such as sloths (Table 2.5). At least 15 independent copies appear in the haplorhine lineage with many of these copies located within transcribed loci (Table 2.5). Given the limited species distribution and high sequence identity between genomic *GAS5* sequences and these copies, the retrotransposition events described here are relatively recent. It is likely that additional, ancient copies of *GAS5* are present in mammalian genomes but have diverged beyond recognition.

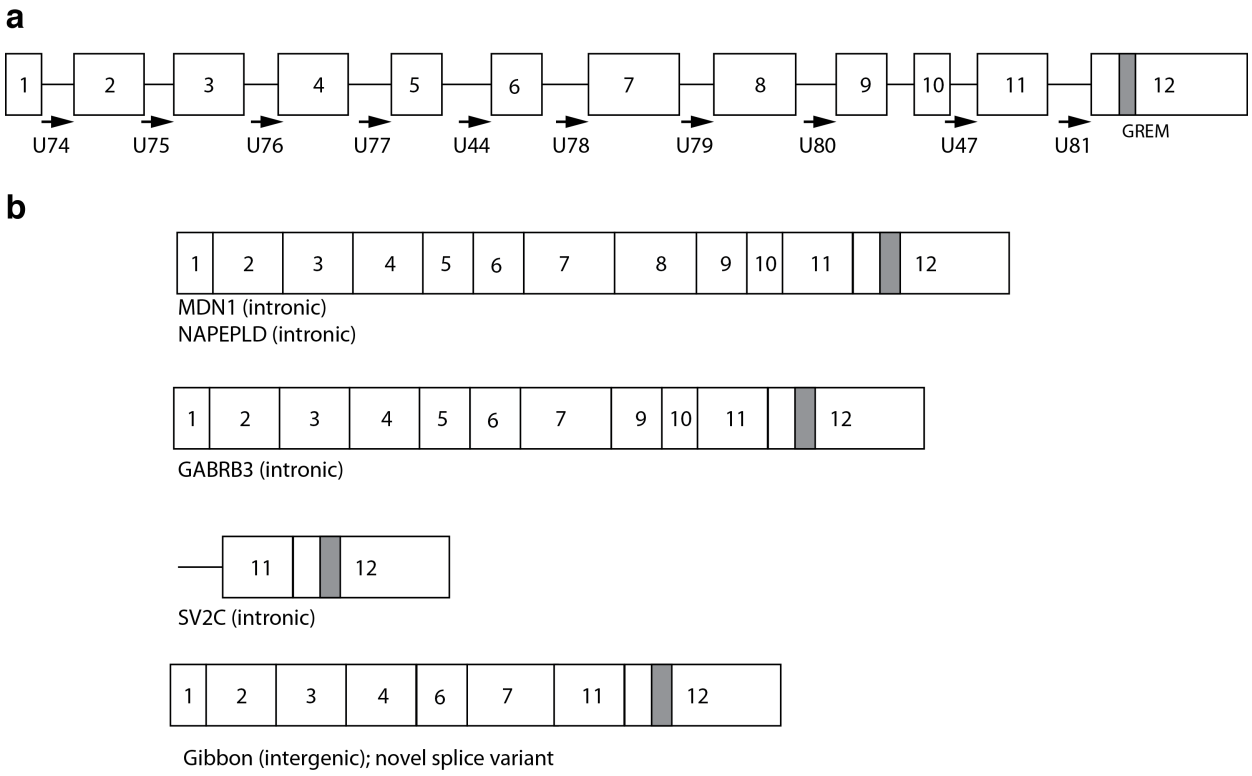


Figure 2.13: Processed Gas5 transcripts are present across mammalian genomes. (a) The structure of the *GAS5* gene, showing the complicated 12-exon structure that gives rise to at least 29 splice variants. (b) Selected retrotransposed copies of Gas5 in primate genomes. In humans, copies of fully spliced Gas5 are present in the introns of the *MDN1* and *NAPEPLD* genes. Additional variants are found in the *GABRB2* and *SV2C* genes. A currently undescribed splice variant is present as a retrotransposed copy in the gibbon genome. A more comprehensive list of Gas5 copies in mammalian genomes can be found in Table 2.5.

Species	Host gene	Chromosome or contig	Position	Notes
Human	Intergenic	1	91515150-91515512	Not present in marmoset or macaque
Human	*SV2C*	5	75503325-75504199	Intronic; In marmoset through human
Human	*MDN1*	6	90430849-90431533	Intronic; Not present in marmoset or macaque
Human	*NAPELD1*	7	102766413-102767023	Intronic; Not present in marmoset, present in macaque
Human	*GABRB3*	15	26890549-26891121	Intronic; Not present in marmoset or macaque
Gibbon		GL397277.1	15903584-15904571	Not present in human
Gibbon		GL397291.1	13299130-13300294	Novel splice variant; not present in human
Gibbon		GL397418.1	2010802-2011783	
Macaque	*SFRS16*	5	109342131-109342386	In the host gene promoter; Not present in human
Macaque	Intergenic	19	51464559-51465515	Not present in human
Marmoset		2	170511210-170511852	Not present in human
Marmoset		4	36004499-36005456	Not present in human
Marmoset	Intergenic	4	42481974-42483397	Not present in human
Marmoset	Intergenic	9	58792598-58794226	Not present in human

Marmoset	Intergenic	12	11735326-11736078	Not present in human
Marmoset	*TM7SF4*	16	90230274-90230600	In the host gene 3' UTR; Not present in human
Dog		8	16442238-16442450	
Dolphin		scaffold_86663	8686-9353	
Mouse	*USP13*	3	32864146-32864604	Intronic
Mouse	*Wee1*	5	79638027-79638538	Intronic
Mouse	Intergenic	8	75431299-75431482	
Rat	*F1LZG6_RAT*	2	178388169-178388429	Intronic (LRBA alternate gene name)
Rat	Intergenic	2	184798750-184799252	
Rat	Intergenic	8	16893702-16893937	
Rat	*Kpnb1*	10	85892834-85893047	Intronic
Rat	*Flt1*	12	7878928-7879481	Intronic
Sloth		scaffold_294758	615-1099	
Sloth		scaffold_436	32904-33393	
Sloth		scaffold_72748	4883-5310	
Tenrec		Genescaffold_5531	164947-165316	
Tenrec		scaffold_257391	3053-3462	

Table 2.5: Retrotransposed, processed copies of Gas5 pervade both transcribed and intergenic regions of mammalian genomes. Species, host gene (if applicable), location, and various notes are given. Genomic location may not be the exact ends of the retrotransposon. If genome annotation does not allow determination of a host gene (or intergenic position), no host gene is given. This may not be a comprehensive list of all processed Gas5 retrotransposons in the listed species. Genome builds used: Human, GRCh37; Gibbon, Nleu1.0; Macaque, MMUL_1; Marmoset, C_jacchus3.2.1; Dog, CanFam3.1; Dolphin, truTru1; Mouse, GRCm38; Rat, RGSC3.4; Sloth, choHof1; Tenrec, TENREC. Genomes were accessed via Ensembl.[648]

2.4 Discussion

Using a range of structural and biochemical approaches, we offer a mechanism to explain the potent ability of Gas5 to suppress SR-mediated transactivation. This repression is mediated through sequence-specific protein RNA-contacts within an A-form double helical structure with a widened major groove that facilitates SR binding. We show that full-length, fully spliced Gas5 binds strongly to the GR DBD, with affinity comparable to physiological DNA (+)GREs.[669] In addition, GR exhibits stronger binding to full-length Gas5 than single (+)GRE half-sites or the newly characterized negative GREs (Chapter 3 and reference 574), illustrating the strength of the GR - Gas5 interaction as well as the versatility of GR as a diverse nucleic acid-binding protein.[1,330,669] As discussed in Chapter 1, GR is one of a growing number of proteins that bind DNA and RNA, a class that includes ADAR, TFIIIA and the p50 subunit of nuclear factor-κB, among many others. p50 has been crystallized in complex with its DNA response element as well as a competing RNA aptamer.[307] The RNA aptamer adopts a similar conformation as nuclear factor-κB response elements, and p50 makes nearly identical contacts with RNA as it does with DNA;[307,308] we expect a similar mechanism of DNA mimicry occurs with GR and Gas5. Here we show that GR - Gas5 interactions are mediated through helix 1 of GR's DBD, demonstrating that GR uses the same surface to bind both DNA and Gas5 RNA. However, the dimerization loop of the GR DBD is not affected upon Gas5 binding, although more study is required to definitively determine GR's dimerization status when bound to RNA, given recent reports demonstrating the importance of the ligand-binding domain in receptor dimerization.[682] Although the exact Gas5 GREM sequence does not appear in any human mRNAs, we have demonstrated that this sequence evolved from a splice acceptor site; therefore, GR may be able to recognize RNA transcripts similar to Gas5 at splice site sequences; GR has been previously shown to affect the alternate splicing of several genes.[683–685]

The potent repression of SR signalling by Gas5 is likely explained by both the strong affinity of the SR - Gas5 interaction and by the large discrepancy between Gas5 copy number and GR DNA-binding sites in a cell during periods of Gas5 upregulation. In HeLa cells, Gas5 copy number increases by an order of magnitude to roughly 75,000 during serum starvation,[2] which is on par with or greater than the increase in Gas5 expression caused by exogenous transfection in this study (Figure 5a,d,g). In contrast, available genomic GR-binding sites are roughly fixed in number, ranging from 3,000 to 8,000 depending on cell type.[677] This suggests that Gas5 regulates SR activity by simply titrating its own GR-binding site against genomic GR-binding sites. However, it is unclear whether Gas5 interferes with DNA binding- independent, SR-mediated transrepression such as tethered' GR transrepression.[686] Future studies are needed to determine if Gas5 modulates these DNA-independent effects.

The *GAS5* gene is unique in its high level of alternative splicing: at least 29 non-redundant, alternatively-spliced transcripts originate from the *GAS5* locus, making it one of the most highly alternatively spliced genes in the genome.[661] Although it is not clear whether Gas5 splicing patterns are tissue-specific, the majority of Gas5 splice variants include the GREM, which is sufficient for SR binding. Notably, it was recently shown that exon 12-containing Gas5 transcripts are downregulated in breast ductal carcinoma compared with adjacent normal tissue.[667] Nevertheless, 13 Gas5 splice variants either exclude exon 12 or shorten exon 12 such that the GREM is not included in the final transcript,[681] indicating that alternative splicing may present an additional level of SR regulation by Gas5. Such control may be required for tissue-specific homeostasis and may be disrupted by single-nucleotide polymorphisms (SNPs) that affect the splicing of GREM-containing exon 12 (rs11537772, rs111755386). An additional SNP that alters the GREM sequence (G559U) has also been described (rs186249529). Mutation of G559 in this study, even when combined with a compensatory mutation to preserve secondary structure, dramatically weakens

the GR - Gas5 interaction (Figure 2.6e), potentially attenuating riborepression of SR signalling. Thus, it is possible these and other Gas5 SNPs may predispose individuals to steroid-driven cancers despite apparently normal tissue levels of Gas5.

Given the paucity of identified sequence-specific lincRNA function, we used the detailed biochemical and structural data here to examine the evolutionary origin of the Gas5 GREM. The SR DBDs are highly conserved throughout vertebrates; in contrast, two recent studies with lincRNA functional domains show mixed levels of conservation: megamind and cyrano show conserved function between zebrafish and mammals,[659] whereas key functional human domains of the lincRNA HOTAIR are missing in mice.[687] Given the comparatively faster sequence turnover within lincRNAs compared with their protein targets, we propose that Gas5 co-opted a splice acceptor site into its current SR-binding role. Such a mechanism allows the primary sequence required for SR binding to be conserved because of the presence of a second, stronger pressure for conservation. The relatively loose sequence requirements of an acceptor splice site - a polypyrimidine tract followed by AG - may have allowed this locus to gradually acquire SR-binding ability in eutherian mammals (Figure 2.12d). However, it is clear that once Gas5 - SR binding was established, this function was conserved, given the retention of splice site-like sequences required for SR binding at the GREM (Figure 2.10c). The development of this functional domain within the translationally regulated 5' terminal oligo-pyrimidine Gas5 transcript may have provided an important link between translational downregulation and other critical events involved in growth arrest.[665]

Given the large number - tens of thousands - of lincRNAs, it is likely that hundreds of lincRNAs have precise, sequence-specific functions analogous to that of Gas5. Furthermore, given their large size, it is likely that some lincRNAs contain multiple protein-recognition sites and may serve as scaffolds for protein complexes. Here we have demonstrated that the interaction between Gas5 and SRs depends on

unique structural elements and precise, sequence-specific interactions to differentiate among target proteins. Subtle disruptions of these interactions have dramatic effects on cell fate. Further, we show that the lincRNA Gas5 co-opted splicing regulatory sequences into a SR-binding role, explaining functional and sequence conservation at the GREM. Together, these results demonstrate that lincRNAs can contain conserved, sequence-specific protein interaction surfaces that ensure a high level of specificity to bind to target proteins and control cell fate.

2.5 Methods

2.5.1 Protein expression and purification

The DBDs of the human glucocorticoid receptor (GenBank ADP91253.1; residues L381-E541), androgen receptor (UniProtKB P10275.2; residues D550-K630), progesterone receptor (UniProtKB P06401.4; residues E560-F642), mineralocorticoid receptor (GenBank AAA59571.1; residues S593-G671), and estrogen receptor (UniProtKB P03372.2; residues A175-R260) were cloned as 6X His fusions into the pMCSG7 vector. Mutations were made with the QuikChange site-directed mutagenesis kit (Stratagene). The DBDs were expressed in BL-21(DE3)pLysS *E. coli* and induced with 0.3 mM IPTG for four hours at 30 °C. Cells were lysed in 1 M NaCl, 20 mM Tris-HCl pH 7.4, 25 mM imidazole, 5% glycerol via sonication, and protein was purified via affinity chromatography (HisTrap) followed by TEV protease cleavage and dialysis to 100 mM NaCl, 20 mM Tris-HCl pH 7.4, 25 mM imidazole, and 5% glycerol. The DBDs and tags were separated by affinity chromatography (HisTrap), and further purified by gel filtration chromatography. For storage, protein was concentrated to 4 mg/ml, flash frozen in liquid N_2, and stored at -80 °C. For ^{15}N labeling, the GR DBD was expressed in M9 minimal media with $^{15}NH_4Cl$ as the nitrogen source. Cultures were inoculated at an OD_{600} of 0.3-0.4, and induced for 14 h with 1 mM IPTG at an

OD_{600} of 1.2. The GR DBD was then purified via affinity chromatography as above.

2.5.2 Nucleic acid synthesis and purification

Constructs for *in vitro* transcription were ligated into the puc57 vector using a forward primer with the T7 promoter sequence and a reverse primer using the HindIII restriction site. Plasmid was cut overnight with HindIII (Fermentas) and used as template for *in vitro* transcription in 50 mM Tris-HCl pH 7.5, 15 mM $MgCl_2$, 5 mM DTT and 2mM spermidine with 2 mM each of individual NTPs. Recombinant T7 RNA polymerase (0.1 mg/ml reaction) and inorganic pyrophosphatase (1 U/ml reaction, Fermentas) were added and the reaction was incubated for 5 hours at 37 °C. RNA was purified by anion exchange chromatography as described previously[688] and ethanol precipitated. Both *in vitro* transcribed and synthetic nucleic acid duplexes (Integrated DNA Technologies) were annealed in 10 mM NaCl, 20 mM Tris-HCl pH 8.0 by heating to 85 °C and slow cooling to room temperature.

2.5.3 Nucleic acid binding assays

Fluorescence polarization (FP) was used to detect the formation of DBD-nucleic acid complexes. Indicated amounts of DBD were added to wells containing 10 nM of 5' 6-FAM-labeled nucleic acids. For competition assays, labeled DNA was used at a concentration 1.2 times higher than the GR - (+)GRE K_d. Reactions were performed in 100 mM NaCl, 20 mM Tris-HCl pH 7.4, 5% glycerol and measured with a Biotek Synergy plate-reader at an excitation/emission wavelength of 485/528 nm. FP data were analyzed and graphed using GraphPad Prism 5 (Graphpad Software, Inc.). Error is shown as ±SEM, and was normalized to the mean for figures showing -logK_d. Sequences used in binding experiments can be found in Figure 2.1. All binding experiments were performed at least in duplicate, with three internal replicates.

2.5.4 UV melting assays

UV melting assays were conducted as described previously.[689] Briefly, samples containing 25 μg RNA in a solution of 10 mM Tris pH 7.4 and 20 mM NaCl were melted and UV absorption collected on a Varian Cary 400 UV/Vis spectrophotometer. Sample temperatures were measured with an in-cell temperature probe. First derivatives of the melting curves were calculated in GraphPad Prism 5.

2.5.5 Hydroxyl radical probing

In vitro transcribed Gas5 GREM RNA was treated with FastAP (Fermentas) and 5'-end labeled using [γ-^{32}P]ATP and T4 polynucleotide kinase. Labeled RNA was urea denaturing gel purified and annealed by heating at 65 °C for 10 minutes and slowly cooling to room temperature prior to use. GR was dialyzed against 20 mM Tris pH 7.5, 100 mM NaCl. In a total volume of 25 μl, labeled RNA (50,000 cpm) was incubated without or with GR (0.2, 0.8, 3 or 12 μM) and hydroxyl radicals produced by adding 1 μl each of 50 mM $Fe(SO_4)_2$, 100 mM EDTA, 250 mM ascorbic acid and 3% (v/v) hydrogen peroxide. The reaction was allowed to proceed on ice for 5 minutes before quenching by ethanol precipitation in the presence of yeast tRNA (0.4 mg/ml) as a carrier and resolving the RNA products by 12% urea denaturing-PAGE for 1.5 hours at 55 W. Band quantification was performed with ImageQuant applying the Rubber Band background subtraction. For normalization, the intensity of each band was measured and normalized to the most intense band in 0 μM GR. The normalized intensities were compared between 0 μM and 12 μM GR. A nucleotide was considered protected/enhanced if the difference in normalized intensity was greater than ($\pm$) 20%.

2.5.6 Crystallization and structure analysis

Bipyramidal crystals of the Gas5 GREM duplex appeared in two conditions. Crystals grown in 0.05 M sodium succinate pH 5.5, 0.5 mM spermine, 0.02 M $MgCl_2$, and 3 M ammonium sulfate were frozen directly in liquid N_2. Data from these crystals were collected at 100K at the Emory University home source to a resolution of 2.2 Å. Crystals that formed in 0.1 M Citric Acid pH 4, and 3 M ammonium sulfate were cryoprotected in 0.1 M citric acid pH 4, 3 M ammonium sulfate, and 10% glycerol and flash-frozen in liquid N_2. Data from these crystals were collected at the Advanced Photon Source BM-22, and data from both crystals were integrated and indexed in HKL-2000. Phases were determined using molecular replacement from an ideal RNA helix modeled in Coot,[690] using PHASER in the PHENIX suite.[691] Crystal twinning was detected using phenix.xtriage, and twin refinement was performed in phenix.refine.[691] Molprobity indicated no incorrect sugar puckers, bad bonds, or bad angles in both crystal structures. The composite omit map utility within PHENIX was used to generate the composite omit map, with simulated annealing enabled.[691] RNA and DNA structure analysis was performed in 3DNA,[605] using refined P-P distances to determine major and minor groove widths. Figures of models and crystal structures were generated in the PyMOL Molecular Graphics System (Schrödinger LLC). The Gas5 GREM duplex crystal structures have been deposited to the PDB under accession numbers 4MCE and 4MCF.

2.5.7 NMR

NMR data were collected at 25 °C on a Bruker 700 MHz ^1H frequency equipped with a QCI cryoprobe. For RNA NMR experiments, the 33 nt Gas5 RNA (Figure 2.1) was reconstituted after ethanol precipitation in 20 mM phosphate, 50 mM KCl, 10% D_2O buffer, pH 7.4 to a final concentration of 10 mg/ml. The RNA sample was annealed by denaturing at 95 °C for 5 minutes and equilibrating to room temperature (20-23 °C)

overnight. Two-dimensional ^{1}H-detected NOESY and TOCSY spectra were collected using 100 ms and 200 ms mixing times for Gas5 RNA and after adding equimolar concentration of GR DBD. The proton carrier frequency was set coincident with the water resonance for all experiments. For protein NMR experiments, 2D [^{1}H,^{15}N]-HSQC spectra were collected for free ^{15}N-labeled GR DBD protein or complexed with 33 nt Gas5 RNA in the same buffer as above. Chemical shift perturbations were assigned using previously published GR DBD NMR chemical shifts[672] and calculated using the minimum chemical shift perturbation procedure.[692] Data were analyzed with NMRViewJ (OneMoon Scientific, Inc.).[693]

2.5.8 Modeling the GR DBD - Gas5 RNA complex

NMR chemical shift perturbation and mutagenesis data was used as input to HADDOCK[673] to model the GR DBD-Gas5 RNA complex. Prior to docking, the AMBER 12 package[694] was used to perform molecular dynamics on the crystal structure of the Gas5 GREM duplex in order to widen the major groove. Briefly, the tleap module was used to solvate the Gas5 GREM using a TIP3P octahedral box water model[695] and neutralize the system. The ff12SB force field was employed in the calculations with the sander module used for the equilibration period and the pmemd.cuda code[696,697] used on a graphics processor unit (GPU) for the production runs. The system was equilibrated in a multi-step process,[698] starting with 10000 steps of energy minimization (5000 steps of steepest descent, followed by 5000 steps of conjugate gradient); all non-water and non-hydrogen atoms were restrained during this minimization step. The systems were then equilibrated for 200 ps in the canonical NVT ensemble[699] during which time the temperature was ramped from 10 to 295 K during the first 100 ps and the restraints on the solute atoms were retained. Particle-MESH-Ewald periodic boundary conditions and a cutoff of 1 nm for both electrostatic and Lennard-Jones interactions were employed; the SHAKE algorithm[700] was applied to constrain

bonds to hydrogen atoms, and a time step of 1 fs was used. After HADDOCK docking, the lowest energy docked model from the best scoring cluster was used as the representative model.

2.5.9 Cell culture

Human adherent cell lines, 22Rv1 (prostate carcinoma cells) and MCF-7 (breast carcinoma cells), and suspension cell lines, Jurkat, CEM-C7 (both T-lymphoblastic leukemia cells) and JeKo-1 (Mantle cell B lymphomacells) were routinely cultured in R-10 medium [RPMI-1640 medium supplemented with L-glutamine (2 mM), fetal bovine serum (10%) and gentamicin (50 μg/ml)] at 37 °C in a humidified incubator with 5% CO_2. For experiments to determine the influence of Gas5 on androgen signaling, 22Rv1 cells were switched to prfR-10DCSS [phenol red-free RPMI-1640 medium supplemented with L-glutamine (2 mM), dextran charcoal-stripped fetal bovine serum (10%) and gentamicin (50 μg/ml) at 24 h prior to transfection. All experiments were carried out using cells in the logarithmic phase of growth.

2.5.10 Plasmid DNA transfection

WT Gas5 (NR_002578.2 bases 1-631) was cloned into the pcDNA3.1 vector and the G549A mutation made with the Quikchange site-directed mutagenesis kit (Stratagene). 22Rv1, MCF7 cells (each at 2 x 10^6 cells in 0.1 ml Nucleofector solution V) and JeKo-1 cells (5 x 10^6 cells in 0.1 ml Nucleofector solution R) were nucleofected with 2 μg plasmid, using programs X-001, E-014 and X-001, respectively. Cells were diluted to 2 ml with R-10 medium (or $_{\text{prf}}$R-10_{DCSS} medium, as appropriate) and cultured in 6-well plates.

2.5.11 Determination of apoptosis and cell survival

Adherent cell lines were trypsinized at 24 h post-nucleofection, samples were collected for RNA isolation, and cells were re-plated in 12-well plates (8 x 10^4 cells/well). For 22Rv1 and MCF-7 cells, additional samples were irradiated with ultraviolet (UV) light to induce apoptosis before re-plating: cells (10^5/ml medium) in plastic petri dishes (lids removed) were exposed to a dose of 40 J/m^2 UV-C light, then immediately centrifuged and resuspended in fresh medium. After culture for 24 h (22Rv1) or 48 h (MCF-7), cells were trypsinized, and adherent and non-adherent cells were combined. Suspension cell lines were also collected at 24 h following transfection for total RNA isolation and for re-plating at 2 x 10^5 cells/well in 12-well plates. To determine viability, cells were stained with 0.1% nigrosin (w/v) and counted using a hemocytometer; cells which excluded the dye were considered to be viable. To determine apoptosis, cells were stained with acridine orange (25 μg/ml), and the proportion of cells containing condensed or fragmented chromatin was scored by fluorescence microscopy ($\geq$200 cells/treatment were scored). In some experiments, the proportion of cells containing activated caspase activity was determined using a commercial Caspatag assay (CaspaTag Pan-Caspase In Situ Assay Kit; Millipore, Watford, UK) and fluorescence microscopy ($\geq$200 cells/treatment were scored), according to the supplied instructions.

2.5.12 Clonogenic assay

Long-term survival and proliferation of transfected JeKo-1, CEM-C7, and Jurkat cells were assessed by the ability of the cells to form colonies in soft agar. An equal proportion of the culture from each experimental condition was diluted in 5 ml Iscove's medium (Sigma) containing 20% heat inactivated fetal calf serum, 10% cell conditioned medium and 0.3% Noble agar (Difco) and plated in 60 mm dishes. Dishes were also overlaid with 2.5 ml Iscove's complete medium containing 10% cell con-

ditioned medium. The number of colonies formed was counted following 2-3 weeks incubation at 37 °C in 5% CO_2.

2.5.13 Determination of androgen sensitivity

22Rv1 cells were trypsinized at 24 h post-nucleofection and cells were re-plated in 12-well plates (8 x 10^4 cells/well) in $_{prf}$R-10_{DCSS}. Cells were allowed to attach to plates for 6 h, then treated with 10 nM R1881 in $_{prf}$R-10_{DCSS}; controls received 0.05% DMSO. After 4 h, cells were collected in TRIZOL reagent (Invitrogen, Paisley, UK) for total RNA isolation.

2.5.14 Real time PCR analysis of gene expression

Total RNA was isolated using TRIZOL reagent (Invitrogen, Paisley, UK), treated with RQ1 RNase-free DNase (Promega, Southampton, UK) then reverse transcribed using random hexamer priming and SuperScript II Reverse Transcriptase (Invitrogen, Paisley, UK), according to the supplied protocols. For determination of Gas5 levels, real time PCR was conducted using SensiFast Probe Hi-ROX kit (Bioline, London, UK) and Taq Man Gene Expression Assays (Applied Biosystems, Warrington, UK; assay codes Hs03464472_m1 for *GAS5* and Hs99999901_m1 for 18S). For determination of expression levels of androgen-responsive genes, real time PCR was conducted using SensiFast SYBR Hi-ROX kit (Bioline, London, UK) according to the supplied instructions and primer sets specific for: *KLK2* (forward: 5'-ATGTGTGCTAGAGCTTACTC-3'; reverse:
5'-AAGTGGACCCCCAGAATCAC-3'); *KLK3* (forward:
5'-CCAAGTTCATGRGTGTGR-3'; reverse: 5'-CCCATGACGTGATACCTTGA-3'); *FKBP5* (forward:
5'-CGCAGGATATACGCCAACAT-3'; reverse: 5'-GAAGTCTTCTTGCCCATTGC-3'); and *ACTB* (forward: 5'-GTTTGAGACCTTCAACACCC-3'; reverse:

5'-ATGTCACGCACGATTTCCC-3'). Assays contained 10 ng sample cDNA in a final volume of 25 μl and were run on an ABI Prism Sequence Detection System model 7000. For adherent cell lines, standard curves, comprising 0.2 - 60 ng cDNA (prepared from parental cells), were included with each assay to allow relative quantitation; standard curves of threshold cycle (C_t) value versus log input standard cDNA were constructed by linear regression, and the equation of the line was used to calculate input amounts of samples from their respective C_t values. For suspension cell lines, quantitation of Gas5 was determined using the comparative C_t method, using parental cells as calibrators. Gas5 levels were expressed relative to 18S rRNA, and levels of androgen-responsive genes were expressed relative to ACTB.

2.5.15 Statistical analysis

Data are presented as the mean and standard error of the mean (SEM). Data shown represents at least three independent transfections, each being conducted on a separate cell culture. Data were usually analyzed by a one-way analysis of variance; culture growth of suspension cell lines was analyzed by a two-way analysis of variance. Post-hoc analysis was by Tukey's multiple comparison test. *, **, and *** indicate $p < 0.05$, 0.01, and 0.001, respectively by ANOVA with Tukey's post-hoc test.

2.5.16 Evolutionary analysis

Genomic sequences of the *GAS5* locus were downloaded from the *Ensembl* database.[681] For species where Gas5 was not annotated, the exon 12 sequence was determined using splice prediction downstream of *SNORD81*.[701] Sequences were aligned in Geneious (Biomatters), and evolutionary rate analysis was conducted with the haplorhine sequences in MEGA 5 with the Tamura-Nei model, using all sites and 3 gamma rate categories.[702]

2.6 Acknowledgements

We thank Jo Linpinsel for assistance with the UV melting experiments, Christine Dunham for the T7 polymerase expression construct, and Keith Yamamoto and Lisa Watson for the NMR restraints for the GR DBD and GR DBD - Gha complex. W.H.H. was supported by the Emory-NIH Graduate Training in Pharmacological Studies Grant 5T32GM008602-14, an American Heart Association predoctoral fellowship (13PRE16920012), and E.A.O. and W.H.H. were supported by the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health under award number RO1DK095750. M.R.P, M.M-M, and G.T.W are grateful to the UK charities Breast Cancer Campaign and Leukemia and Lymphoma Research for financial support. This research was also supported by startup funds to E.A.O. from Emory University and startup funds to D.J.K from The Scripps Research Institute. X-ray data were collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-BM beamline at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at www.ser-cat.org/members.html. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38.

Chapter 3

The structural basis of glucocorticoid-mediated transrepression

3.1 Abstract

A newly discovered negative glucocorticoid response element (nGRE) mediates DNA-dependent transrepression by the glucocorticoid receptor (GR) across the genome and plays a major role in immunosuppressive therapy. The nGRE differs dramatically from activating response elements and the mechanism driving GR binding and transrepression is unknown. To unravel the mechanism of nGRE-mediated transrepression by the glucocorticoid receptor, we characterize the interaction between GR and a nGRE in the thymic stromal lymphopoetin (*TSLP*) promoter. We show using structural and mechanistic approaches that nGRE binding represents a new mode of sequence recognition by human GR and that nGREs prevent receptor dimerization through a unique GR-binding orientation and strong negative cooperativity, ensuring the presence of monomeric GR at repressive elements.

This chapter is adapted from the manuscript:
Hudson WH, Youn C, Ortlund EA. The structural basis of direct glucocorticoid-mediated transrepression. *Nat Struct Mol Biol.* 2013 Jan;20(1):53-8.

3.2 Introduction

The glucocorticoid receptor (GR) is a ubiquitously expressed vertebrate nuclear receptor that controls the transcription of genes critical for metabolism, immunity, development, and responses to stress.[703–705] Glucocorticoids, widely prescribed for their powerful immunosuppressive and anti-inflammatory properties,[706] drive the both the transactivation and transrepression of GR target genes, with transactivation of target genes slightly more prevalent than repression.[565] Therapeutically beneficial GR-mediated immunosuppression is thought to occur primarily through indirect or tethered DNA-independent interactions of GR with other transcription factors such as NF-κB[572] and Stat3[707] to repress pro-inflammatory genes.[708] In contrast, side effects of glucocorticoids are often attributed to direct gene activation.[709] Recently, a new role for direct, DNA-dependent transrepression by GR was discovered through the identification of widely-prevalent negative glucocorticoid response elements (nGREs).[574] These elements differ in sequence from activating glucocorticoid response elements ((+)GREs) and selectively recruit the corepressors NCoR and SMRT to the promoters of nGRE-containing genes upon GR binding.[574] Functional nGREs have been identified within hundreds of promoters, including many key inflammatory and metabolic genes.[574] Furthermore, nGRE-containing genes such as insulin, the insulin receptor, and Bcl-2 are implicated in side effects associated with glucocorticoid therapy.[574] Identifying the repressive mechanism of GR at nGRE-containing genes may support the quest for dissociated GR ligands that separate beneficial effects of glucocorticoid agonists from their side effects.

Numerous mechanistic studies have shown that GR transactivation requires the presence of (+)GREs which allosterically mediate GR binding, recruitment of coactivators, and transcription.[669,710] These elements contain two inverted repeat AGAACA sequences separated by three nucleotides, with bold residues critical for GR binding.[677] The three nucleotide spacing between half-sites is strictly required to

preserve dimerization potential of GR on the element.[334] In contrast, the nGRE consensus sequence, CTCC$(n)_{0\text{-}2}$GGAGA, differs dramatically from activating sequences. The spacing required in the nGRE is variable, ranging from 0-2 nucleotides, suggesting that GR dimerization may not be necessary for nGRE-mediated transrepression.

Given the radically different sequence and organization of nGREs, it is unclear how GR binds to this element to repress the vast array of nGRE containing genes. To unravel the mechanism of nGRE-mediated transrepression by the glucocorticoid receptor, we characterize the interaction between GR and a nGRE in the thymic stromal lymphopoetin (*TSLP*) promoter. This nGRE, 850 base pairs (bp) upstream of the *TSLP* transcription start site, mediates the reduction of *TSLP* mRNA levels by 50% in response to GR agonists.[574] TSLP regulates many critical immune processes[711–713] and is implicated in disorders such as atopic dermatitis, asthma, irritable bowel syndrome, and arthritis.[714–718] Using this prototypical nGRE, we employ structural, biochemical, and cellular approaches to demonstrate that two GR monomers bind nGREs in an everted repeat orientation with strong negative cooperativity. When combined, the unique GR conformation and negative cooperativity, ensures the presence of monomeric GR at nGREs. This interaction mechanism represents a new mode of GR - DNA binding and a new paradigm for GR transrepression.

3.3 Results

3.3.1 nGRE binding displays negative cooperativity

To initially characterize the affinity of GR for nGREs, we compared binding of recombinantly expressed glucocorticoid receptor DNA binding domain (DBD) to fluorescently-labeled activating and repressive GR response elements. The canonical (+)GRE contains two nearly-identical inverted GR binding sites separated by 3 bp and enables cooperative DNA binding and dimerization. We found that GR binds

these elements with a K_d of 73 nM and a Hill slope of 1.4, indicating the expected positive cooperativity (Figure 3.1a, Table 3.1). However, when testing GR binding to the *TSLP* nGRE, we observed a dramatically different binding curve, which qualitatively appeared as a two-site binding event (Figure 3.1a).

To test the superiority of a two-site binding model for GR - nGRE interactions, we performed an extra-sum-of-squares F-test comparing a two-site binding event with a cooperative one-site binding event (Table 3.1). We found that the *TSLP* nGRE contains two distinct binding sites ($p < .0001$) with K_ds of 363 nM and 63 μM. To establish this as a property of all nGREs, we confirmed this result on other nGREs from promoters of genes such as insulin and *FGFR3* (Figure 3.1b, Table 3.1). We found that relatively weaker binding of GR to nGREs compared to (+)GREs appears to be a general feature of nGREs and mirrors the affinity of GR for a canonical (+)GRE half-site (Table 3.1). While high-affinity site binding affinity was relatively constant among nGREs, affinity of the second site varied considerably, suggesting that flanking sequence of the low-affinity site may affect its ability to recruit GR.

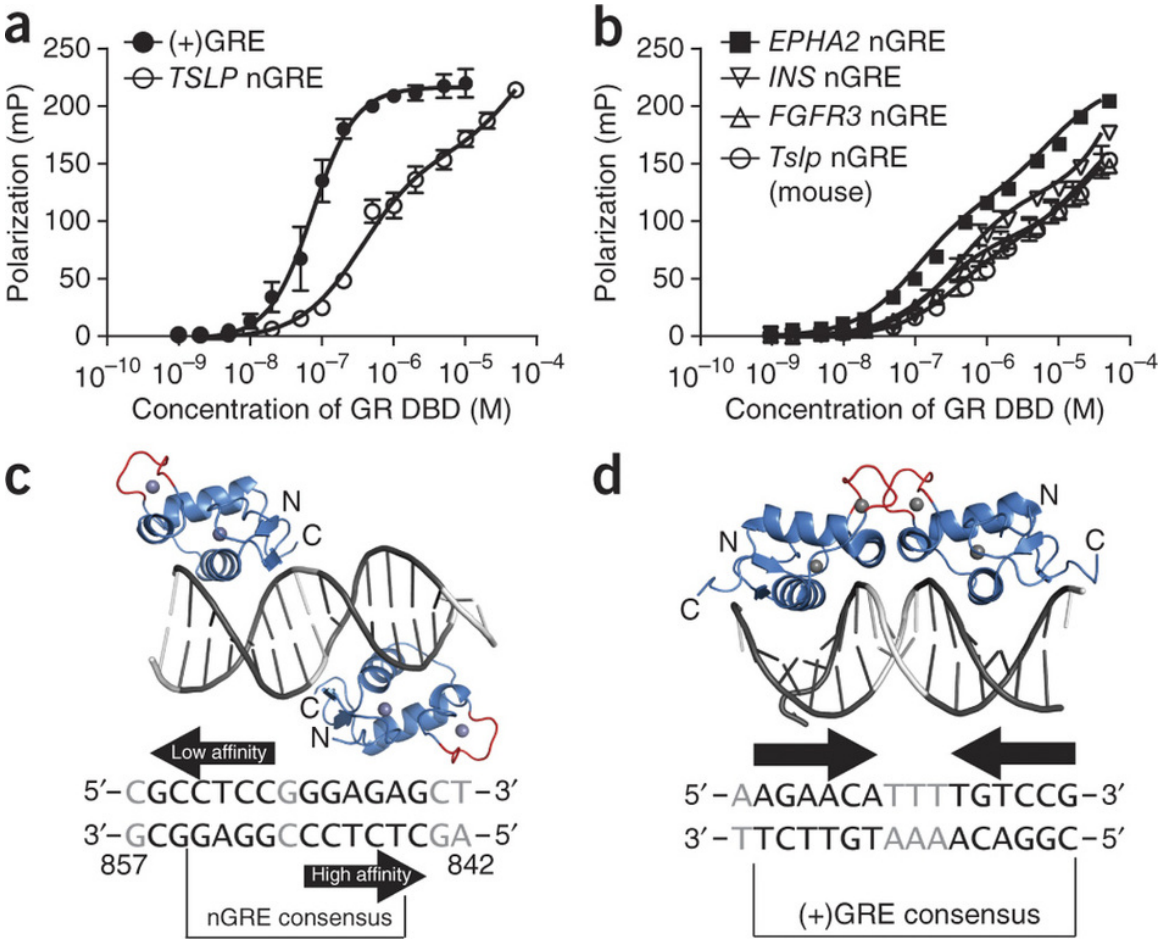


Figure 3.1: GR interacts with nGREs in a unique orientation, preventing receptor dimerization. (a) GR-DNA binding monitored by fluorescence polarization. mP, millipolarization. (b) Additional human nGREs as well as the mouse *Tslp* nGRE exhibit a similar binding profile with two GR binding sites and affinities comparable to that of the *TSLP* promoter. Binding data are represented as mean ± SEM from three replicates from at least two independent fluorescence polarization experiments. (c) Overall structure of GR DBD (blue) in complex with the *TSLP* nGRE (gray). Bases comprising the GR binding sites are in black, and Zn^{2+} is depicted as gray spheres. The GR dimerization interface (red) of each GR monomer is oriented away from the second monomer in an everted repeat conformation. (d) Structure showing how GR binds to a (+)GRE as a dimer in an inverted repeat conformation, enabling contact between dimerization loops (PDB 3FYL).[669]

	F (DF_n, DF_d) for two-site binding	p-value for two-site binding	$K_{d,\ high}$ ± SEM (μM)	$K_{d,\ low}$ ± SEM (μM)	r^2
(+)GRE	NC	NC	0.07 ± .007	NC	0.92
TSLP nGRE	21.84 (1,92)	<0.0001	0.36 ± 0.06	63 ± 86	0.97
FGFR3 nGRE	8.386 (1,92)	0.0047	0.28 ± 0.08	42 ± 42	0.92
INS nGRE	43.26 (1,92)	<0.0001	0.50 ± 0.07	444 ± 1976	0.98
EPHA2 nGRE	39.85 (1,43)	<0.0001	0.14 ± 0.01	9.8 ± 2.1	0.99
Tslp nGRE (mouse)	3.457 (1,92)	0.0662	0.52 ± 0.19	36 ± 37	0.92
GR DBD Arg460Asp / Asp462Arg - *TSLP* nGRE	10.16 (1,140)	0.0018	0.14 ± 0.06	5.6 ± 1.5	0.95
GR DBD Lys442Ala - *TSLP* nGRE	NC	NC	5.14 ± 0.51	NC	0.97
TSLP nGRE, high site mutation	2.579 (1,92)	0.1117	0.83 ± 0.2	14 ± 9.3	0.98
TSLP nGRE, low site mutation	16.83 (1,44)	0.0002	0.26 ± 0.06	11 ± 1.2	0.99
TSLP nGRE (mouse), IR0	8.968 (1,92)	0.0035	0.79 ± 0.18	43 ± 57	0.97
(+)GRE Half-site only	NC	NC	0.38 ± 0.09	NC	0.92
TSLP, low site only	NC	NC	12.7 ± 0.8	NC	0.99
GR DBD Ala458Thr - (+)GRE	26.54 (2,44)	<0.0001	0.37 ± 0.06	10.9 ± 6.0	0.99
GR DBD Ala458Thr - *TSLP* nGRE	NC	NC	1.1 ± 0.08	NC	0.99
GR DBD Ala458Thr - *TSLP* nGRE (mouse)	NC	NC	1.5 ± 0.08	NC	0.99
GR DBD Ala458Thr - *INS* nGRE	NC	NC	1.1 ± 0.05	NC	0.99
GR DBD Arg460Asp / Asp462Arg - (+)GRE	NC	NC	0.08 ± 0.007	NC	0.97
GR DBD Ala458Thr - low site only	NC	NC	2.8 ± 0.3	NC	0.99

Table 3.1: Interaction between GR DBD and nGREs, monitored by fluorescence polarization. NC, no convergence of two-site binding model. All binding experiments are with WT GR DBD unless otherwise noted.

3.3.2 Structure of the repressive GR DBD - nGRE complex

To discover the structural basis for this unusual mechanism of binding, we solved the crystal structure of the GR DNA binding domain (DBD) in complex with the *TSLP* nGRE to a resolution of 1.9 Å (Table 3.2). Surprisingly, the crystal structure showed two GR monomers bound to non-identical everted sites in a head-to-tail fashion, separated by 1 bp as predicted (Figure 3.1c).[574] In this orientation, the dimerization loop (or D-loop) of each GR monomer is directed away from the other monomer and rotated by 180° around the DNA axis (Figure 3.1c), abrogating the opportunity for DBD-mediated GR dimerization. In contrast, GR binds (+)GREs in a head-to-head orientation on the same side of DNA, allowing cooperative binding and dimerization (Figure 3.1d). The everted repeat conformation found in the nGRE ensures monomeric binding by preventing DNA-mediated GR dimerization and may explain the element's repressive character since monomeric GR is associated with gene repression.[719] To our knowledge, this unexpected everted repeat nuclear receptor-DNA binding geometry has been previously described only in the thyroid and retinoic acid receptors.[720,721]

	GR DBD-*TSLP* nGRE	GR DBD Arg460Asp / Asp462Arg - *TSLP* nGRE
Data collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
a, b, c (Å)	39.3, 96.6, 104.0	38.7, 87.9, 103.2
Resolution (Å)	1.9 (1.97-1.90)	2.55 (2.64-2.55)
R_{sym}	8.5 (54.5)	9.8 (37.7)
I / σI	26.7 (2.2)	15.7 (2.8)
Completeness (%)	99.7 (98.4)	96.9 (83.3)
Redundancy	6.3 (4.2)	3.7 (2.8)
Refinement		
Resolution (Å)	1.9	2.55
No. reflections	31,815	11,685
R_{work} / R_{free}	20.7 / 23.5	19.5 / 24.8
No. atoms		
Protein	1,115	1,110
DNA	650	650
Water	204	33
B-factors		
Protein	39.7	32.6
DNA	56.5	45.5
Water	44.7	30.2
RMS deviations		
Bond lengths (Å)	0.007	0.013
Bond angles (Å)	1.15	1.78

Table 3.2: Data collection and refinement statistics. Both data sets obtained from a single crystal.

3.3.3 GR binds to nGREs as two monomers at nonequivalent sites

Based on the identification of low- and high-affinity sites in our GR - nGRE binding data, we hypothesized that each nGRE-bound GR monomer may make different contacts with DNA, resulting in differing affinities of each monomer for its binding site. Indeed, each of the two bound monomers uses different amino acid side chains to make base-specific contacts. One GR monomer makes three base-specific contacts, whereas a second monomer contacts only one base in a specific fashion (Fig 3.2). To assist in the determination of the high- and low-affinity sites, we used the PISA server[722] to identify free energy gains from GR monomer-DNA interactions. The first DNA-monomer interaction, with three specific contacts, has a very favorable free energy change upon formation of the interface (ΔG = -9.5 kcal/mol). The second monomer shows a ΔG of only -5.9 kcal/mol, identifying the former site as the likely high-affinity site.

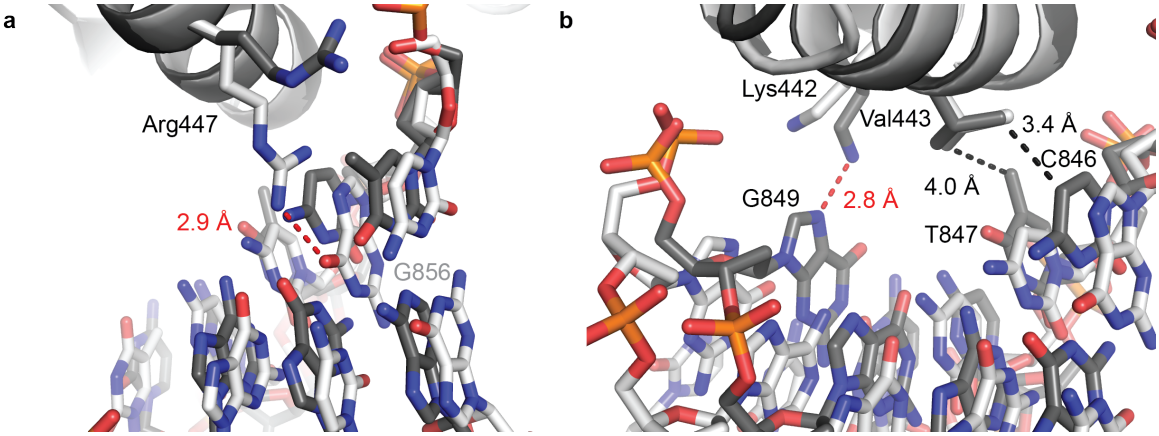


Figure 3.2: Comparison of GR receptor conformation bound to the low- and high-affinity nGRE sites. Close up view of the high- and low-affinity GR - *TSLP* nGRE interaction with side chains and nucleotide depicted as sticks (O, red; N, blue). (a) In the low affinity nGRE binding site (white), Arg447 contacts a guanine base outside the nGRE consensus sequence. However, in the high-affinity site (black), Arg447 makes no base-specific contacts. Rather, Arg447 makes contact with the C844 backbone phosphate. (b) In the low-affinity site (white) Lys442 and Val443, which are critical for recognition of the high affinity nGRE site (black), make no base-specific contacts.

This suspected high-affinity GR DBD - nGRE DNA interaction involves three base specific contacts within the major groove (Figure 3.3a): Val443 makes hydrophobic contacts with cytosine 846 and thymine 847 and Lys442 donates a hydrogen bond to N7 of guanine 849. Mutation of this guanine to adenine increases the K_d of GR for the high affinity site, confirming the identity of the high-affinity GR binding site (Table 3.1). In a previous study, the identical mutation ablates the repressive ability of the mouse *TSLP* nGRE.[574] Likewise, mutation of Lys442 significantly diminishes nGRE binding (Table 3.1). Unlike its DNA-reading function in (+)GRE structures, the Arg447 side chain is prevented from making base-specific contacts due to a steric clash with thymine 845 (Figure 3.3b). The repositioned Arg447 instead make hydrophobic interactions with this base and ionic interactions with the cytosine 844 backbone phosphate. Mutation of thymine 845 to guanine, which would permit the active conformation of Arg447, abrogates transrepression.[574] The low-affinity GR DBD - DNA interaction involves only one sequence-specific contact: Arg447 contacts guanine 856, outside the nGRE consensus sequence (Figure 3.2a). Mutation of guanine 856 does not affect GR binding to the high-affinity site (Table 3.1), and Lys442 and Val443 do not sufficiently penetrate the major groove to facilitate sequence-specific DNA contacts (Figure 3.2b). As a result, the DNA major groove at the low-affinity site contains more waters than the high-affinity site. Recognition of the nGRE high-affinity site requires a more specific contacts and a greater hydrophobic interaction surface than either the low-affinity nGRE site or (+)GRE sequences, as confirmed by PISA.[574]

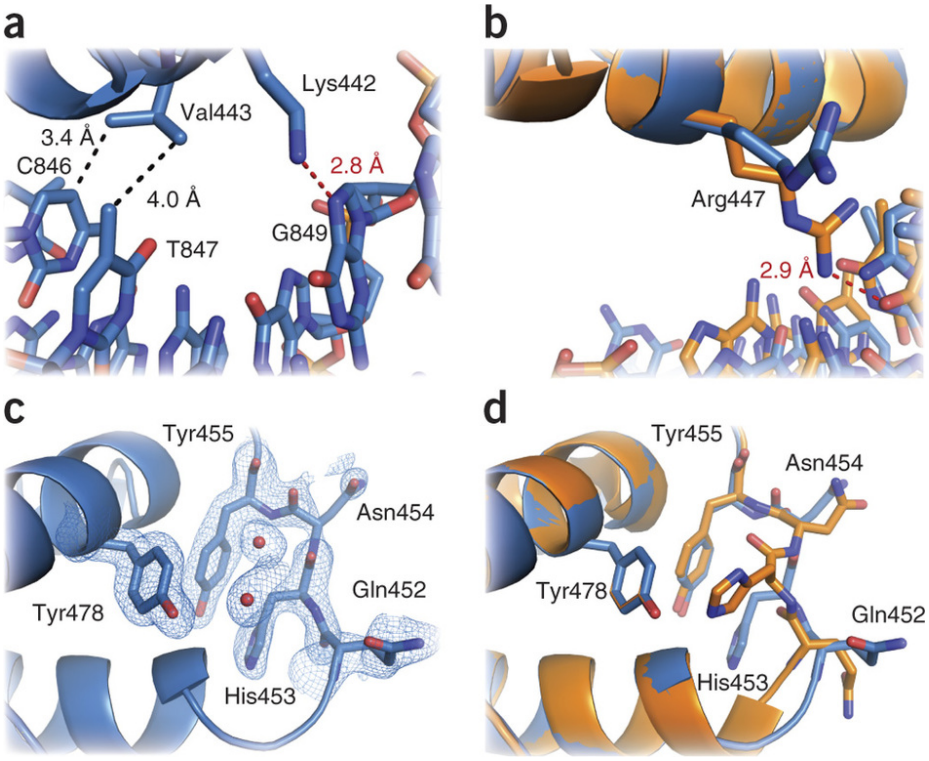


Figure 3.3: GR employs unique interactions to recognize the high-affinity site within nGREs. (a) Close up view of the high-affinity GR - *TSLP* nGRE interaction with side chains and nucleotide depicted as sticks (O, red; N, blue). Hydrogen bonds and van der Waals interactions are represented by red and black dashed lines, respectively. Three base-specific contacts are present between GR and the high-affinity nGRE binding site. Val443 makes two hydrophobic contacts, and Lys442 donates a hydrogen bond to guanine 849. (b) Arg447 makes unique non-specific interactions with DNA at the high-affinity nGRE binding site. In contrast, Arg447 makes base-specific contacts with a guanine base when bound to a (+)GRE (orange; PDB 3FYL). c) $2F_o$-F_c electron density (blue mesh) contoured at 1 σ showing the conformation of the lever arm residues in *TSLP* nGRE-bound GR alone and (d) superimposed on (+)GRE-bound GR (orange).

Taken together, these data demonstrate the mechanism by which GR recognizes the GGAG within the high-affinity nGRE binding site (Figure 3.1c), and explains the strict conservation of one of these GGAG motifs present in the nGRE consensus.[574] The role of the low affinity GR site within the nGRE remains unclear. Despite an identical GGAG sequence present at the low-affinity site, GR binds this site very weakly (Table 3.1). The low affinity site is far more resistant to mutation than the high-affinity site, yet spacing between the low- and high-affinity sites affects both GR binding and transrepression (reference 574, Table 3.1).

3.3.4 DNA-mediated allostery differs between activating and repressive GR response elements

Recent work comparing several GR - (+)GRE crystal structures demonstrated that DNA serves as an allosteric modulator of GR activity where the binding of the first GR monomer relays conformational information through DNA to promote the second binding event, ultimately driving transactivation by favoring coactivator recruitment.[669,723] This positive cooperativity is so strong that detection of the intermediate state (monomeric GR on DNA) is often difficult.[724] In contrast, we found that GR binding to the *TSLP* nGRE exhibits unusually strong negative cooperativity, where binding of the first GR monomer impedes binding of the second (Figure 3.4). The GR nGRE complex also exhibits a different DNA shape than (+)GREs, with a narrow major groove compared to the average of 11 GR DBD structures solved[669] in complex with 16 bp (+)GRE DNA constructs (Figure 3.5). B-factor analysis also reveals that nGRE and (+)GRE DNA undergo dramatically different structural perturbations upon GR binding (Figure 3.5). On (+)GREs, GR binding drives a constriction of the minor groove to facilitate direct protein protein contacts. In contrast, the GR - nGRE interaction forces a narrower major groove and wider minor groove, which opposes the binding of a second GR monomer (Figure 3.5). Since monomeric GR

is linked with transcriptional repression,[719] negative cooperativity may reinforce the repressive character of the nGRE. Alternatively, since recruitment of coactivators by steroid receptors may depend on cooperative binding on DNA response elements,[723] non-cooperative mechanisms of DNA binding may allow DNA-sensitive domains of GR to adopt alternate, repressive confirmations.

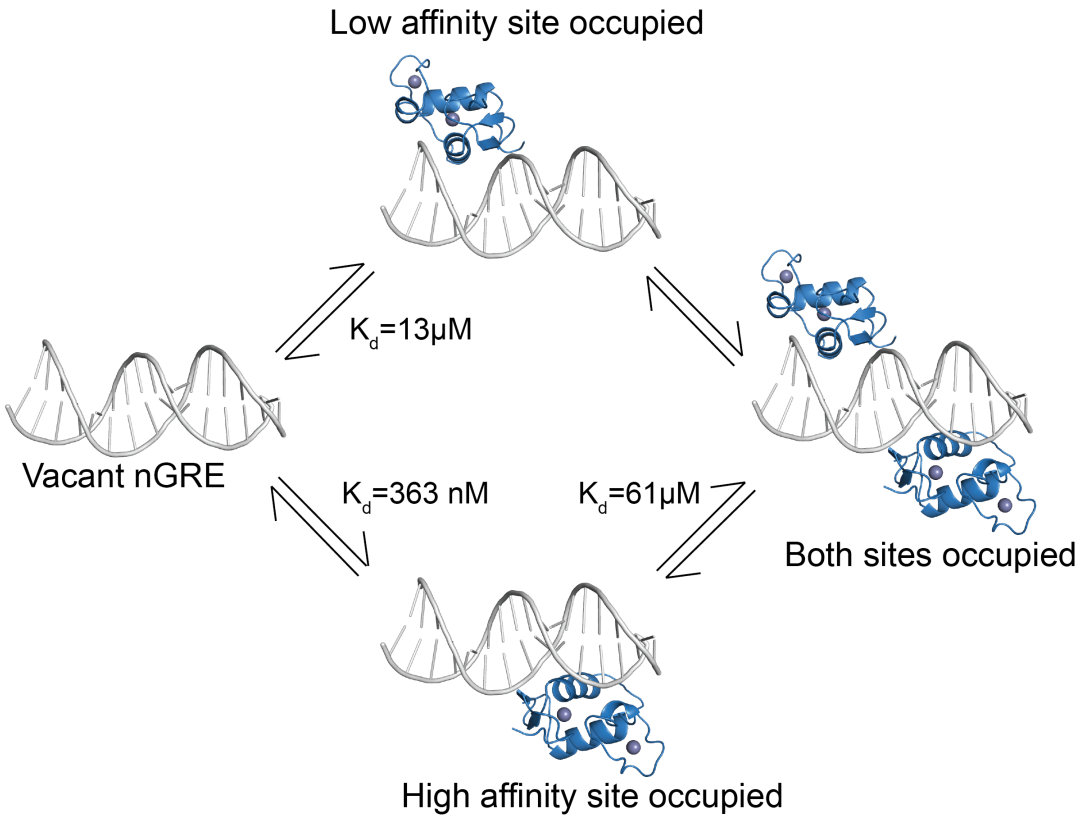


Figure 3.4: Binding model of GR to the *TSLP* nGRE. The *TSLP* nGRE contains two non-identical, interacting binding sites. Binding of GR to the high- and low affinity sites is defined by K_1 and K_2, respectively. However, binding of GR to the low-affinity site is reduced when the high affinity site is occupied ($K_2 > K_3$), demonstrating negative cooperativity.

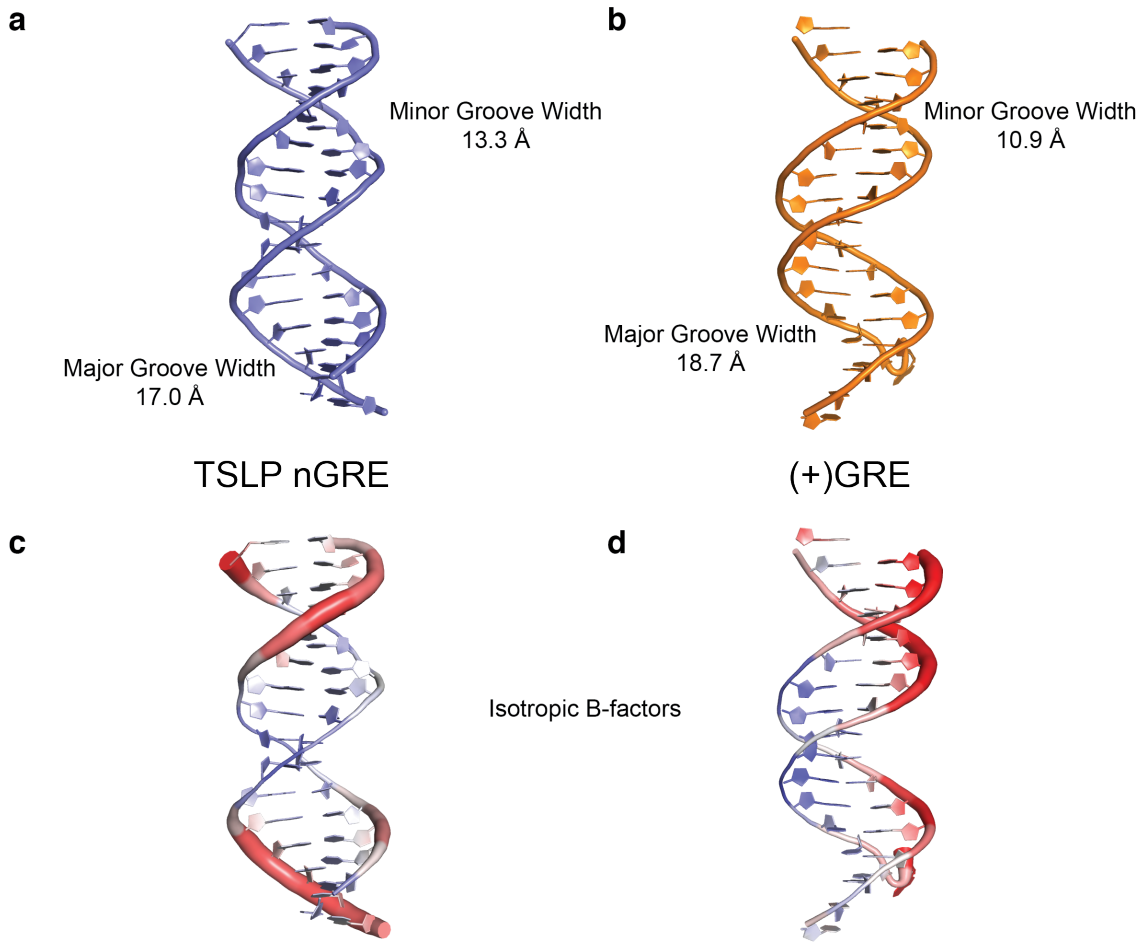


Figure 3.5: GR-bound nGRE DNA exhibits a different conformation than GR-bound (+)GRE elements. (a, b) The DNA shape from the GR - nGRE crystal structure reported in this article (left) and a previously solved GR - (+)GRE structure (PDB 3G9P). Groove measurements for (+)GRE DNA are an average of 11 GR - (+)GRE structures solved with 16 bp DNA constructs. (c, d) Isotropic B-factors are shown for the DNA in the GR-*TSLP* nGRE structure and a GR - (+)GRE structure.

For example, the lever arm, which immediately follows the DNA reading helix, has been identified as being the critical structural motif sensitive to sequence-dependent conformational changes on (+)GREs.[669] When bound to a nGRE, these lever arm residues adopt a distinct conformation compared to (+)GRE-bound GR (Figure 3.3c, 3.3d). Specifically, His453 adopts a flipped conformation in both monomers, interacting with Arg447 and Tyr455 rather than the packed conformation critical for activation from (+)GRE-containing promoter elements (Figure 3.3d, Figure 3.6). The loss of sequence-specific contacts by Arg447 in the nGRE allows His453 to be stabilized in a flipped conformation by both a hydrogen bond and van der Waals contact from the repositioned Arg447 (Figure 3.6). Repositioning of Arg447 also eliminates half of a helical turn of the DNA reading helix, supporting the flipped conformation of the lever arm (Figure 3.3d). The lever arm is the most dynamic portion of the GR DBD, with B-factors significantly higher than other regions of the protein, yet these residues display good electron density. This modulation of the lever arm by sequence-specific contacts illustrates the pivotal role of the lever arm in receptor activation and confirms the allosteric ability of DNA to drive receptor activation and repression.

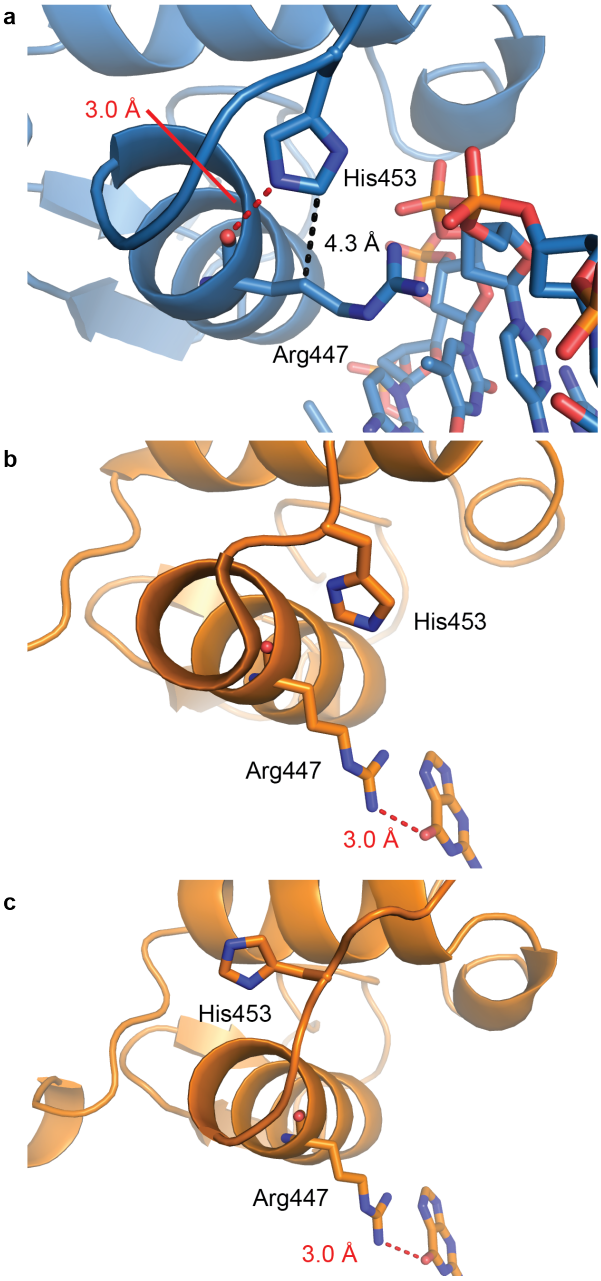


Figure 3.6: Repositioning of Arg447 in nGRE-bound GR stabilizes a flipped conformation of His453. A close up view of Arg447 and His453 interactions in the GR - *TSLP* nGRE complex (blue) and GR - (+)GRE complex (orange) is shown. (a) When bound to the high-affinity nGRE site, Arg447 of nGRE makes no sequence specific contacts (see also Figure 3.3b). However, Arg447 is able to stabilize His453 in the flipped conformation using a hydrogen bond (red dashed line) and van der Waals contact (black dashed line) not present in (+)GRE-bound GR (PDB 3G6P, panels b and c). On (+)GREs, His453 may be flipped (b) or packed (c). In both cases, Arg447 makes a sequence specific contact with DNA and is therefore not positioned to control the conformation of His453.

3.3.5 Dimerization competes with nGRE binding and transrepression

Our structure and model of nGRE action predict that receptor dimerization opposes nGRE binding and therefore interferes with direct transrepression. Recent work has indicated that GR is unique among steroid receptors in that it exhibits no reversible self-association[725] and is dependent on receptor-DNA interactions for dimerization.[726,727] To examine the effects of altered dimerization surfaces on nGRE binding and repression, we used two well-characterized GR mutants: Ala458Thr, often called the GRdim mutant, which is unable to support most glucocorticoid mediated gene activation, (+)GRE binding, or direct DNA repression *in vivo*,[728] and a double mutant containing the Arg460Asp and Asp462Arg mutations (Arg460Asp / Asp462Arg), which has been shown to reduce GR dimerization and decrease activation of multiple (+)GREs.[729] Notably, this mutation was previously found to potentiate repression of Bcl-2,[730] which was recently shown to harbor a consensus nGRE within its promoter.[574]

The Ala458Thr mutant bound to a (+)GRE in a clear two-site binding event (Table 3.1), indicating a loss of cooperativity on this element. In this way, binding of the Ala458Thr mutant to (+)GREs strongly resembles binding of WT GR to nGREs. The Arg460Asp / Asp462Arg mutant showed similar DNA binding as WT to (+)GRE sequences. Next, we tested each of these mutants for binding to nGREs. The Ala458Thr mutation differentially affected binding to each of the GR binding sites on the *TSLP* nGRE, improving low-affinity site binding but decreasing high-affinity site binding (Table 3.1). The net effect of this mutation is to decrease the affinity of GR for nGREs by 500%, to nearly 3 μM. In contrast, the Arg460Asp / Asp462Arg mutation improved binding at both sites on the *TSLP* nGRE. We then tested the ability of each variant to repress a reporter containing a constitutively active luciferase gene preceded by the nGRE-containing region of the *TSLP* promoter, as performed

previously.[574] In line with our *in vitro* binding data, the Ala458Thr showed a modest ability to repress luciferase expression (Figure 3.8). Strikingly, the Arg460Asp / Asp462Arg mutation was a more potent repressor of luciferase activity than WT GR (Figure 3.7a). To observe the effects of this mutant on the GR dimerization interface, we solved the crystal structure of the GR Arg460Asp / Asp462Arg mutant bound to the *TSLP* nGRE (Figure 3.7b, Table 3.2). The structure of the Arg460Asp / Asp462Arg mutant shows a less favorable dimerization interface (Figure 3.7c-e), suggesting that the superior binding and repressive potential of the GR Arg460Asp / Asp462Arg mutant is indeed due to decreased dimerization efficiency.

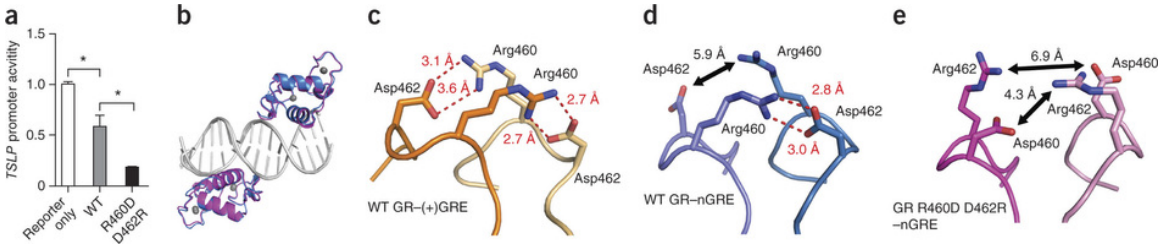


Figure 3.7: The GR Arg460Asp / Asp462Arg mutant reduces receptor dimerization, enhancing GR-mediated transrepression at the *TSLP* nGRE. (a) Transfection of the plasmid encoding the GR Arg460Asp / Asp462Arg mutant in HeLa cells potentiates downregulation of a constitutively active *TSLP* promoter compared to the wild-type (WT) GR (5 ng each). Data are presented as the mean ± SEM of two independent experiments with five internal replicates each. (b) Superposition of GR (blue) and the GR Arg460Asp / Asp462Arg mutant (magenta) bound to the *TSLP* nGRE (gray). (c) When bound to a (+)GRE element, Arg460 and Asp462 form two intermolecular salt bridges (red dashes) across the homodimer interface (PDB 3FYL). (d) In the GR - nGRE structure, crystal-packing interactions require the formation of a pseudocontinuous DNA helix and promote the formation of a pseudo-GR dimer across a twofold symmetry axis. These interactions are necessary for crystal formation but are not possible in solution-based binding assays. (e) Struture showing that the GR Arg460Asp / Asp462Arg mutant lacks key dimerization contacts between GR monomers, disrupting symmetry-imposed dimerization contacts.

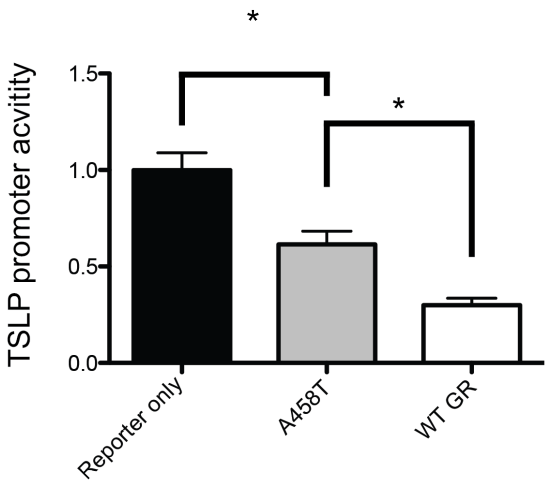


Figure 3.8: The GRdim mutant, Ala458Thr, represses a constitutively active *TSLP* nGRE- containing promoter. 10 ng of indicated receptor was transfected into HeLa cells, and the data represents two independent experiments with five internal replicates each, and mean ± SEM is shown.

3.4 Discussion

The glucocorticoid receptor controls the transcriptional activation and repression of thousands of genes. Multiple regulatory levels are required to achieve a coordinated response, including epigenetic and mRNA regulation, posttranslational modification, circadian rhythms, ligand availability, and target DNA sequence accessibility and binding.[677,731–734] Here, we demonstrate that DNA-binding orientation and sequence-specific contacts control repression of negative GR response element-containing genes. GR binds to these nGREs in a head-to-tail, rotated conformation that prevents DNA-mediated dimerization, in contrast to the DNA-mediated dimerization found on activating GR-binding sites. These unique nGRE sequences alter the conformation of GR residues critical for transcriptional activation, further illustrating the importance of DNA as an allosteric modulator of receptor activity.

A similar mechanism of allosteric modulation between repressive and activating response elements has been demonstrated with the transcription factor Pit-1. Like GR, Pit-1 is monomeric in solution and dimerizes in a DNA-dependent manner.[735] Pit-1 differentially represses and activates transcription of target genes based on spacing between DNA response elements, and this difference in DNA sequence allows recruitment of NCoR to repressive Pit-1 elements.[736] However, Pit-1 maintains similar protein-DNA contacts at both repressive and activating elements; repressive elements differ in that they contain two additional, conserved bases between half-sites.[736] Further, Pit-1 homodimerizes in both the transactivating and transrepressive complexes. In contrast, we demonstrate that nGREs have evolved to recognize GR using a new set of sequence specific criteria favoring monomeric binding over the cooperative binding observed in (+)GREs. This altered sequence generates a novel high-affinity GR binding site and affects the conformation of GR residues, such as His472, which are critical for receptor activation.[669] The comparison between Pit-1 and GR gives an excellent example of how different transcription factors adopt activating and repressive confor-

mations via contrasting mechanisms. It is possible that other transcription factors have alternate DNA response elements that may differentially affect their function. Notably, the other 3-keto steroid receptors (the androgen, mineralocorticoid, and progesterone receptors) can recognize (+)GREs, but it is currently unknown whether these receptors can bind or mediate repression from nGREs.

In general, GR-dependent activation requires DNA-mediated receptor dimerization. We confirm that the GRdim mutation, Ala458Thr, ablates DNA-mediated cooperative binding to (+)GREs. Despite this, the GRdim mutation does not actively repress (+)GRE containing genes (e.g. it does not convert a (+)GRE into a repressive element), suggesting that the GRdim mutant is either incapable of stably binding (+)GREs as a monomer *in vivo* or that the presence of monomeric GR at (+)GRE elements is not sufficient for corepressor recruitment. This indicates that the nGRE sequence may be specific not only for monomeric binding of GR but also for arranging the receptor into a repressive conformation. The lever arm, previously implicated in receptor activation status,[669] adopts a distinct conformation in the nGRE-bound structures reported here, suggesting that it plays a critical role mediating not only GR transactivation but also transrepression.

Widespread clinical use of glucocorticoids has fueled the search for dissociated compounds, capable of minimizing side effects without compromising their anti-inflammatory function. One such GR ligand, Compound A (CpdA), has been shown to inhibit GR dimerization and consequently transactivation from (+)GRE containing genes,[737] yet still supports the transrepression of nGRE-containing genes such as *POMC*.[738,739] Thus, if the major effect of CpdA is to disrupt dimer formation, it is now clear why CpdA permits transrepression from nGREs while preventing transactivation, suggesting that the opposing effects of direct, DNA-dependent transrepression and transactivation are mediated by the dimerization status of the receptor.

3.5 Methods

3.5.1 Protein expression and purification

The DNA binding domain (DBD) of human glucocorticoid receptor (GR) α (residues 417-506, accession ADP91252) was cloned with a 6X-Histidine tag into the pMCSG7 vector. The DBD was expressed in BL-21(DE3) pLysS *E. coli* and induced with 0.3 mM IPTG for four hours at 30 °C. Cells were lysed in 1 M NaCl, 20 mM Tris-HCl (pH 7.4), 25 mM imidazole, and 5% glycerol via sonication. Protein was purified via affinity chromatography (HisTrap) followed by TEV protease cleavage and dialysis to 100 mM NaCl, 20 mM Tris-HCl (pH 7.4), and 5% glycerol. The DBD and affinity tag were separated by affinity chromatography (HisTrap), and further purified by gel filtration. For storage, protein was concentrated to 4 mg/ml, flash frozen in liquid N_2, and stored at -80 °C. Mutations were made using the QuikChange Site-directed Mutagenesis kit (Stratagene).

3.5.2 Nucleic acid binding assays

Synthesized nucleic acid duplexes (Integrated DNA Technologies) were annealed in 10 mM NaCl, 20 mM Tris-HCl pH 8.0 by heating to 90 °C and slow cooling to room temperature. Fluorescence polarization was used to detect the formation of DBD - nucleic acid complexes. Indicated amounts of DBD were added to wells containing 10 nM of 6-FAM-labeled nucleic acids (Table 3.1). Reactions were performed in 100 mM NaCl, 20 mM Tris-HCl (pH 7.4), 5% glycerol and measured with a Biotek Synergy plate-reader at an excitation/emission wavelength of 485/528 nm.

For each binding experiment, an F-test was used to compare a two-site binding event to a one-site binding event with Hill slope, generating an F-statistic and p-value for a two-site binding model. In Table 3.1, these F-statistics (with numerator and denominator degrees of freedom) and p-values are shown in addition to K_d values

for the low and high affinity DNA binding sites, and the coefficient of determination (r^2) of the applicable fit. Complexes are with WT GR DBD unless otherwise noted, and Graphpad Prism 5 was used for binding data analysis and graph generation. Nucleotide sequences used in binding experiments are shown in Table 3.3.

Construct	Sequence
TSLP nGRE (crystallization)	5' - CGCCTCCGGGAGAGCT - 3' 5' - AGCTCTCCCGGAGGCG - 3'
TSLP nGRE (binding)	5' - (FAM)CCGCCTCCGGGAGAGCTG - 3' 5' - CAGCTCTCCCGGAGGCGG - 3'
FGFR3 nGRE	5' - (FAM)CTGCCTCCCCGGAGATGG - 3' 5' - CCATCTCCGGGGAGGCAG - 3'
INS nGRE	5' - (FAM)TGCTCTCCTGGAGACATT - 3' 5' - AATGTCTCCAGGAGAGCA - 3'
EPHA2 nGRE	5' - (FAM)TGAGCTCCAGGAGAAGGG - 3' 5' - CCCTTCTCCTGGAGCTCA - 3'
Tslp nGRE (mouse)	5' - (FAM)TGAGCTCCAGGAGAGTAG - 3' 5' - CTACTCTCCTGGAGCTCA - 3';
TSLP (high site mutation)	5' - (FAM)CCGCCTCCGAGAGAGCTG - 3' 5' - CAGCTCTCTCGGAGGCGG - 3'
TSLP (low site mutation)	5' - (FAM)CCGCCTCGGGGAGAGCTG - 3' 5' - CAGCTCTCCCCGAGGCGG - 3'
Tslp, IR0 (mouse, spacer removed)	5' - (FAM)TGAGCTCCGGAGAGTA - 3' 5' - TACTCTCCGGAGCTCA - 3'
(+)GRE	5' - (FAM)CCAGAACAGAGTGTTCTGA - 3' 5' - TCAGAACACTCTGTTCTGG - 3'
TSLP, low site only	5' - (FAM)CCGCCTCCGG - 3' 5' - CCGGAGGCGG - 3'

Table 3.3: Oligonucleotides used in crystallization and binding studies. FAM indicates the position of 6-carboxyfluorescein.

3.5.3 Reporter gene assays

A 400 bp region of the *TSLP* promoter surrounding the nGRE (chr5:110,406,332-110,406,745; GRCh37) was cloned between an SV40 enhancer and promoter upstream of firefly luciferase, similar to the construct described previously.[574] 50 ng of this construct, indicated amounts of receptor, and 1 ng of constitutively active renilla luciferase were transfected with FuGene HD (Promega) in OptiMEM (Invitrogen) into HeLa cells cultured in AlphaMEM (Invitrogen) supplemented with 10% charcoal stripped FBS (PAA). 24 hours after transfection, cells were treated with 1 μM dexamethasone, and after 18 hours, firefly and renilla luciferase were measured with the Dual-Glo assay system (Promega) on a Biotek Synergy plate-reader. Firefly luciferase activity was normalized to renilla luciferase for each well, and levels of all treatments were normalized to cells transfected only with the constitutively active nGRE construct and not treated with dexamethasone. An asterisk indicates $p < 0.01$ by ANOVA with Tukey's multiple comparison test.

3.5.4 Structure determination

Crystals of the GR DBD - *TSLP* nGRE complex were grown by hanging-drop vapor diffusion in 15% PEG 20000, 6% glycerol, 7.5% ethanol, and 0.1M HEPES (pH 7.5) with a protein concentration of 3.5 mg/ml and a 2:1 protein:DNA molar ratio. Crystals were cryoprotected in crystallant with 20% PEG 20000 and 20% glycerol and flash-frozen in liquid N_2. Crystals of the GR DBD Arg460Asp / Asp462Arg - *TSLP* nGRE complex were grown by hanging-drop vapor diffusion in 15% PEG 2000 MME, 6% glycerol, and 0.1M HEPES (pH 7.5), with a protein concentration of 3.5 mg/ml and a 2:1 protein:DNA molar ratio. These crystals were cryoprotected in crystallant with 25% PEG 2000 MME and 20% glycerol, and flash-frozen in liquid N_2. Data were collected at 100 K and a wavelength of 1.00 Å at Southeast Regional Collaborative Access Team (SER-CAT) at the Advanced Photon Source (Argonne,

IL) and processed using the HKL-2000 software. The structures were phased via molecular replacement using previously-solved structures[669] of the GR-GRE complex in Phenix.[691] The structure was refined with phenix.refine[691] and model building was performed in COOT.[690] 99% of residues are Ramachandran favored or allowed regions for both the WT GR and Arg460Asp / Asp462Arg structures, respectively, with 1% outliers in both structures. Pymol was used to visualize the structure and generate figures. 3DNA was used to analyze nucleic acid groove widths.[605] Amino acids are numbered according to the human GR sequence (ADP91252). Bases are numbered by position upstream of the *TSLP* (NM_033035) transcription start site, which is an additional 199 nt upstream of the translation start site (CCDS4101).

3.5.5 Accession codes

Coordinates and structure factors of the human GR DBD and Arg460Asp / Asp462Arg mutant bound to the human *TSLP* promoter have been deposited in the Protein Data Bank under accession numbers 4HN5 and 4HN6, respectively.

3.6 Statistical Note

In a model with two identical interacting binding sites, cooperativity (β) is positive when the affinity for binding of a second GR monomer is higher than the affinity for the first and is defined by the ratio of two affinities, yielding $\beta > 1$. Negative cooperativity is defined by $\beta < 1$. The value for perfect negative cooperativity is zero where the binding of one monomer completely prevents the binding of second monomer.

However, since the *TSLP* nGRE contains two non-identical interacting GR binding sites, we used a two independent interacting binding site model to determine cooperativity (Figure 3.4). Binding of GR DBD to the fluorescently labeled WT

TSLP nGRE construct provided association constants ($K = 1/K_d$) for high affinity site binding and low affinity site binding following high affinity site binding (K_1 and K_3, respectively). To determine the affinity of low affinity site binding (K_2), we used a fluorescently labeled *TSLP* nGRE construct containing only the low affinity site (Table 3.3).

In this model, cooperativity (β') is given by equation 3.1:

$$\beta = \frac{\frac{2K_1K_3}{K_1+K_2}}{\frac{1}{2}(K_1+K_2)} \tag{3.1}$$

which gives a value indicating strongly negative cooperativity, $\beta' = 0.02$. This indicates that binding of the first GR monomer at the high affinity site strongly hinders binding of a second GR monomer.

Statistical model adapted from notes by A. van Oudenaarden: Equilibrium binding and cooperativity, Accessed May 31, 2012. `http://web.mit.edu/biophysics/sbio/PDFs/L3_notes.pdf`.

3.7 Acknowledgements

We thank N.M. Choi, O. Laur and members of the Emory Custom Cloning Core Facility for assistance with cloning the *TSLP* nGRE reporter construct. This work was supported with start-up funds from Emory University. W.H.H. was supported by an Emory-National Institute of Health Pharmacological Sciences graduate training grant (5T32GM008602).

Chapter 4

Crystal structure of the mineralocorticoid receptor DNA binding domain in complex with DNA

4.1 Abstract

The steroid hormone receptors regulate important physiological functions such as reproduction, metabolism, immunity, and electrolyte balance. Mutations within steroid receptors often drive cancers or cause endocrine disorders. Despite the conserved structure of proteins within the steroid receptor family, activation of individual steroid receptors can often have differing effects on gene expression. Here, we present the first structure of the human mineralocorticoid receptor DNA binding domain, in complex with a canonical DNA response element. The overall structure is similar to the glucocorticoid receptor DNA binding domain, but small changes in the mode of DNA binding and lever arm conformation may begin to explain the differential effects on gene regulation by the mineralocorticoid and glucocorticoid receptors. In addition, we explore the structural effects of mineralocorticoid receptor DNA binding domain mutations found in type I pseudohypoaldosteronism and multiple types of cancer.

This chapter is adapted from the manuscript:
Hudson WH, Youn C, Ortlund EA. Crystal structure of the mineralocorticoid receptor DNA binding domain in complex with DNA. *PLoS ONE*. 2014 Sept 4; 9(9):e107000.

4.2 Introduction

Steroid hormones are powerful regulators of homeostatic functions such as cell growth, immunity, reproduction, and metabolism.[740] Steroid hormones exert their effects by binding to steroid hormone receptors (SRs), which include the estrogen receptors as well as members of the NR3C subfamily (i.e. the mineralocorticoid, glucocorticoid, androgen, and progesterone receptors). Upon ligand binding, SRs translocate from the cytoplasm, where they are bound to heat shock proteins, to the nucleus where they bind their DNA response elements and regulate the transcription of hundreds of genes.[741] The potent transcriptional activity of SRs combined with a very high affinity for their endogenous ligands allows small concentrations of steroid hormones to coordinate diverse cellular processes across a single organism.

Protein domain structure is conserved throughout the NR3C family. The SRs contain an N-terminal transactivation domain of variable length, a DNA binding domain (DBD) containing two Cys_4 zinc fingers, and a flexible hinge connecting the DBD with the ligand binding domain (LBD).[742] The N-terminal transactivation domain and the hinge are different lengths and not well conserved among the NR3C receptors. While hormone preference differs among NR3C receptors due to sequence differences in the LBD, the DBDs are highly conserved, conferring overlapping DNA binding preferences for all members of this subfamily. However, the mineralocorticoid receptor (MR) and glucocorticoid receptor (GR), which diverged after a gene duplication event of the ancient corticoid receptor, also show overlap in hormone preference:[743] MR responds to aldosterone, 11-deoxycorticosterone and cortisol, while GR is selective for cortisol only.

Expression of MR is tissue-specific, with highest concentrations found in the kidney, brain, and heart.[744–747] MR is involved in responses to stress and is basally activated in the brain,[748,749] but is most commonly studied for its role in vascular health and salt and water balance. MR knockout mice develop normally, but die near

postnatal day 10 from renal sodium and water loss.[750] These mice exhibit extreme hyperactivation of the renin-angiotensin system, with elevated renin, angiotensin II, and aldosterone levels.[750] The endogenous mineralocorticoid aldosterone promotes atherosclerotic plaque formation,[751] and aldosterone levels are predictors of acute ischemic events and death in patients with coronary artery disease.[752] These findings have led to the use of MR antagonists to treat heart failure.[753]

Both MR and GR bind glucocorticoid response elements ((+)GREs) on genomic DNA to control target gene expression.[565,754] Administration of glucocorticoids prolongs the survival of $MR^{-/-}$ mice, suggesting overlapping - but not fully compensatory - functions of GR and MR.[750] Additionally, GR and MR differentially regulate cellular functions such as inflammation, with MR often acting as a pro-inflammatory factor and GR acting as an anti-inflammatory factor.[755–758] These differences may be due to opposing effects on gene regulation. For example, MR upregulates the expression of the pro-inflammatory gene *ICAM1*,[759] while GR acts to transrepress its expression.[760]

The mechanisms that underlie such differential gene regulation by receptors with similar sequence and overlapping preferences for ligand and DNA binding are unknown. In this chapter, we determine the first crystal structure of the MR DBD in order to provide a framework for elucidating the subtle differences between the corticosteroid receptors and interpreting the biology of disease-associated mutations.

4.3 Results

4.3.1 Crystal structure of the MR DBD - (+)GRE complex

After purification of the MR DBD, we tested the ability of both the MR DBD and GR DBD to bind to a fluorescently labeled (+)GRE via fluorescence polarization (Figure 4.1a). Both proteins showed similar affinity for the element, at 55 nM and 53 nM for MR and GR, respectively. These are consistent with previous reports of

MR - DNA binding on the order of 10 nM.[761] Both proteins showed similar, slight positive cooperativity in DNA binding, which would likely be augmented by lower salt concentrations.[762]

We then crystallized the MR - (+)GRE complex, obtaining small crystals that anisotropically diffracted to 2.4 Å (Table 4.1). The crystal structure of the MR - (+)GRE complex reveals a canonical steroid receptor DBD dimer bound to the (+)GRE sequence via two half sites (Figure 4.1b). As expected, the structure is very similar to structure of GR DBD - (+)GRE complexes (RMSD $<$ 1.0 Å; Figure 4.1c). As multiple GR - (+)GRE complexes have previously been solved, we compared our novel MR-GRE structure to PDB 3G6P, which contains GR in complex with a (+)GRE from the FKBP5 promoter.[669] This (+)GRE is nearly identical to the sequence contained in the crystal structure reported here (Figure 4.1b,c). One GR DBD monomer (monomer B in Figure 4.1c) contains a C-terminal α-helix when bound to the FKBP5 (+)GRE (Figure 4.1c); the MR DBD structure reported here does not exhibit such a structure. However, not all GR DBDs form this helix when bound to DNA, including monomer A in the GR DBD - FKBP5 (+)GRE structure (Figure 4.1c).

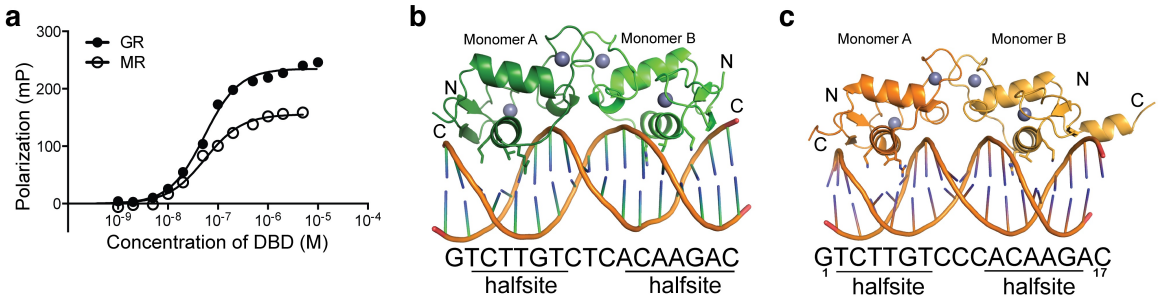


Figure 4.1: Structure of the human mineralocorticoid DNA binding domain in complex with a glucocorticoid response element. (a) The MR DBD binds to a (+)GRE with approximately the same affinity as the GR DBD. (b) Overall structure of the MR DBD (green) bound to a 17 base pair (+)GRE. The sequence of the element, along with the two bound half sites, is shown below the structure. In both panels (b) and (c), the structure shown depicts the asymmetric unit of the crystal structure and separate GR monomers are differentially colored. (c) Structure of the GR DBD (orange) bound to a similar (+)GRE, with sequence and half sites indicated below. Panel (c) is derived from the structure of the GR DBD bound to the FKBP5 (+)GRE, PDB 3G6P.[669]

Data collection	
Space group	$C222_1$
Cell dimensions	
a, b, c (Å)	74.1, 115.1, 81.4
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å)	2.39 (2.48 - 2.39)*
R_{merge}	7.8 (45.4)
I / σI	25.3 (2.4)
Completeness (%)	92.7 (66.7)**
Redundancy	4.1 (3.0)

Refinement	
Resolution (Å)	2.39
No. reflections	13291
R_{work} / R_{free}	21.9 / 25.8
No. atoms	
Protein	1084
DNA	691
Water	4
B-factors	
Protein	71.0
DNA	86.0
Water	60.1
RMS deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.23

Table 4.1: Data collection and refinement statistics. *Data for highest resolution shell are in parentheses. **Data are 94.1% complete to 2.59 Å. Data were collected from a single crystal. The estimated isotropic and anisotropic Wilson B for the data is 52.94 and 86.19 $Å^2$, respectively.

4.3.2 Sequence-specific contacts between MR and (+)GREs

Inspection of the MR - DNA interface reveals three amino acids that make sequence-specific contacts with (+)GRE bases (Figure 4.2a). The terminal nitrogen of Lys624 forms a hydrogen bond with the N7 position of guanine 3. On the opposite DNA strand, Val625 makes van der Waals contacts with C7 of thymine 13, and Arg629 makes two interactions with guanine 12 at the O6 and N7 positions. These interactions are supported by excellent electron density (Figure 4.2a).

These interactions are conserved in the glucocorticoid receptor, which contacts DNA in an identical fashion (Figure 4.2b). Lys442, valine 443, and arginine 447 in GR make contacts with a guanine, thymine, and guanine, respectively. These DNA-reading amino acids are strictly conserved in the four steroid receptors in the GR/MR-like subfamily (Figure 4.2c), and their mutation in GR leads to deficiencies in both DNA and RNA binding in GR.[2,763]

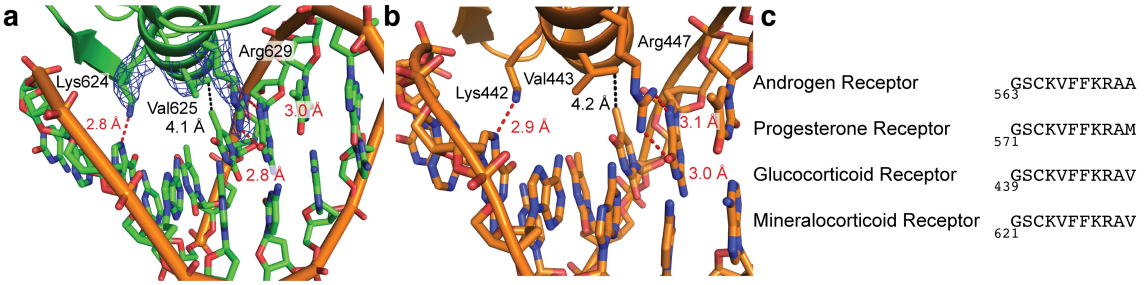


Figure 4.2: Sequence-specific DNA recognition by the MR DBD. (a) Three residues mediate sequence-specific contacts by the MR DBD DNA reading helix. Lys624 makes a hydrogen bond with a guanine base, Val625 makes van der Waals contacts with a thymine base, and Arg629 makes two interactions with a guanine base. Electron density (composite omit $2F_o$-F_c map with simulated annealing, contoured to 1 σ) is shown for the three protein side chains. (b) GR recognizes (+)GREs in an identical manner as the MR DBD, using Lys442, Val443, and Arg447 to contact analogous bases. (c) Sequence alignment showing conservation of the DNA reading helix among the NR3C receptors.

4.3.3 Lever arm conformation of MR

Previous studies have proposed that DNA sequence allosterically modulates GR's structure, in turn affecting transcriptional activation.[669,672] One possible mechanism for such allosteric modulation may be structural changes in the lever arm of steroid receptors, which connects the DNA reading helix of the receptor to its dimerization loop (Figure 4.3a). Mutation of lever arm residues affects transcriptional activation,[669] and one GR splice variant, GRγ, contains a single arginine insertion into the lever arm. This insertion has the ability to affect both GR's binding to target DNA as well as its transcriptional activity.[764]

The lever arm sequence of MR is identical to that of GR and the progesterone receptor (PR), although the androgen receptor (AR) contains three amino acid changes in this region (Figure 4.3a). One key structural element of the lever arm is the position of His453 in GR, whose homologous residue is His635 in MR. In GR, the side chain of His453 can assume a flipped conformation, where it occupies a position between the DNA and the reading helix (Figure 4.3b). This conformation can also be seen, with minor variations, in the crystal structure of MR bound to a (+)GRE (Figure 4.3b). However, His453 can also assume a packed conformation in GR, wherein the side chain rests between GR helices and stacks against a tyrosine residue in helix 3 of the DBD fold (Figure 4.3c). This tyrosine, residue 478, is unique to GR and is conserved as a leucine in the other NR3C family receptors (Figure 4.3a). This likely reduces the stability of the packed conformation, and such a conformation is not found in the MR - GRE structure (Figure 4.3c). This single amino acid change may cause MR to respond differently than GR to identical sequence elements and alter any potential protein-DNA mediated allostery. However, conclusions regarding the structure and function of the lever arm may be confounded by crystal packing contacts found in the lever arm in many GR DBD-DNA structures as well as in chain B (but not chain A) of the structure reported here.

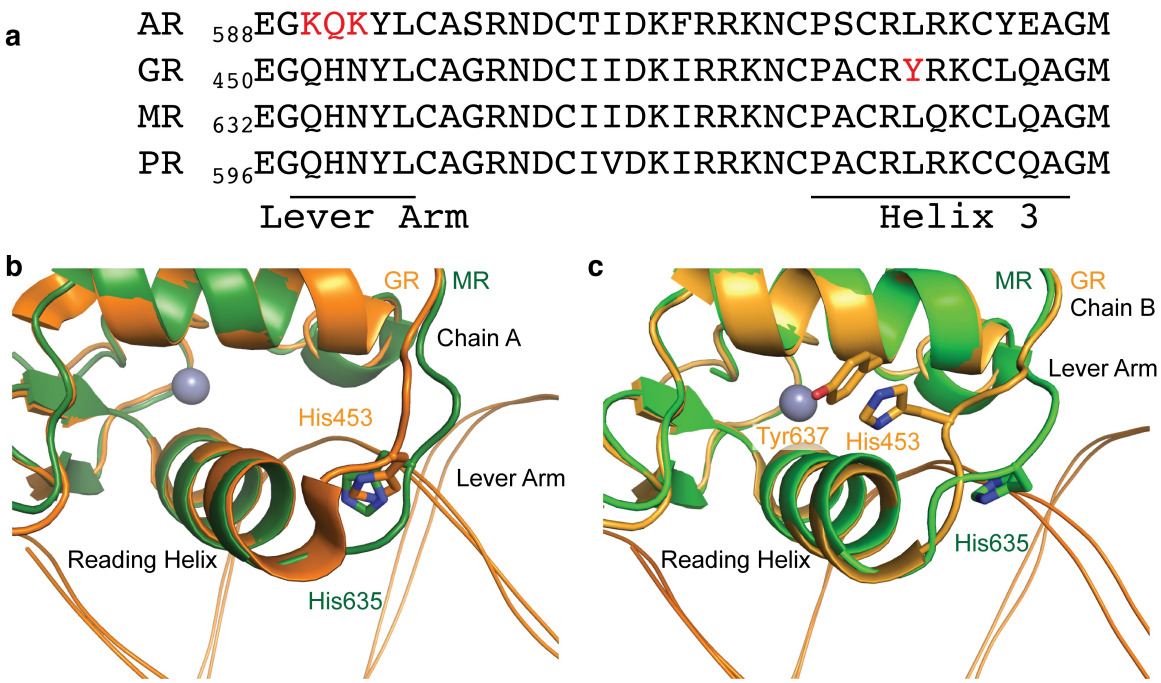


Figure 4.3: Lever arm conformation differs between MR and GR. (a) Sequence alignment of the lever arm through helix 3 of the NR3C receptors. AR and GR contain divergent sequence at the lever arm and helix 3, respectively (red). In a dimer of GR molecules on DNA (see Figure 4.1c), the side chain of His453 can assume two conformations. (b) Monomer A of the GR - FKBP5 (+)GRE complex contains His453 in a flipped conformation, where the histidine side chain sits between the DNA and DBD reading helix; a similar conformation is seen in monomer A of the MR DBD - (+)GRE complex. (c) However, His453 in the second GR DBD monomer assumes a packed position against Tyr478 in the core of the GR DBD fold. This conformation does not occur in the MR DBD - (+)GRE structure, likely due to the presence of a leucine rather than tyrosine at position 660 (GR position 478). In panels (b) and (c), DNA is shown as a ribbon helix below the protein.

4.3.4 Dissecting the protein-DNA interface

In addition to protein-based allostery, recent studies have shown that the shape of DNA itself regulates protein-DNA binding. Occupancy of one protein binding site on a DNA double helix affects the occupancy of additional binding sites in a periodic manner.[765] This phenomenon may be exploited by the glucocorticoid receptor to prevent cooperative dimerization at negative glucocorticoid response elements (nGREs; see Chapter 3). The MR - (+)GRE structure reveals that the MR dimer perturbs DNA in a very similar manner as a GR dimer (Figure 4.4a). Both receptors induced a similar widening of the DNA major groove to 18 Å (Figure 4.4b). This is noticeably distinct from interactions of GR with nGREs, where the major groove is constricted relative to (+)GREs bound to GR.[333]

We also analyzed the buried surface area of each MR monomer-DNA interface as well as the MR dimer interface. The MR dimer interface buries 574 Å^2, similar to the GR dimer interface when bound to the FKBP5 (+)GRE, which buries 555 Å^2 . However, each the MR - (+)GRE interface is comprised of a much smaller surface area than the corresponding GR - (+)GRE interface. The two MR DBD - DNA interfaces bury 373 Å^2, and 369 Å^2, compared to 554 Å^2 and 520 Å^2 for GR. This is consistent with GR's potential ability to bind to DNA as a monomer.[333]

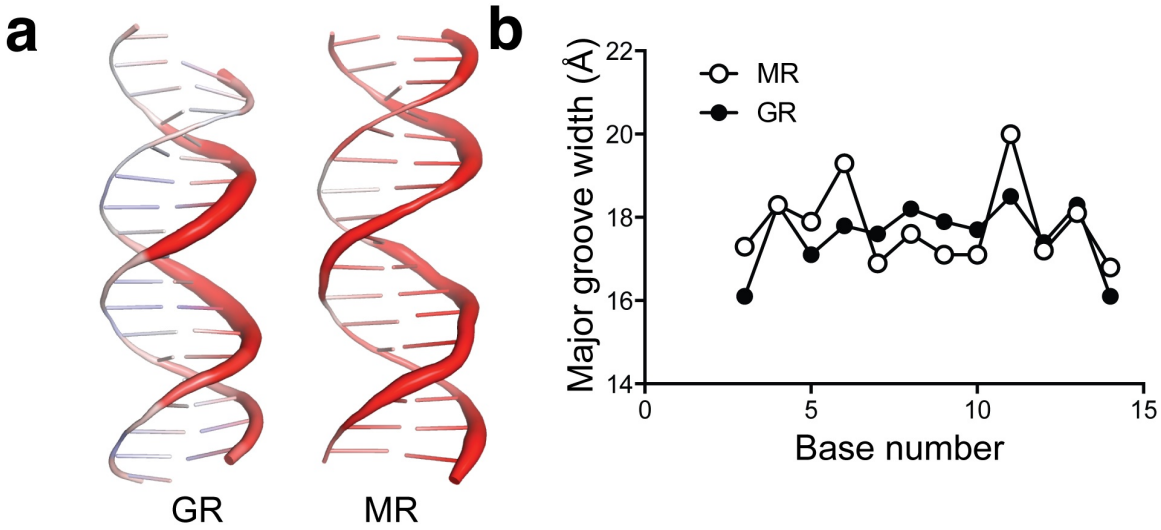


Figure 4.4: Analysis of the MR - DNA interface. (a) Thermal motion of (+)GREs when bound to GR (left) and MR (right). Thicker, red sections of DNA indicate higher B-factors and therefore higher thermal motion. (b) Major groove width at each position of the (+)GRE when bound to the MR and GR DBDs.

4.4 Discussion

Mutations to the mineralocorticoid receptor are the primary cause of type 1 pseudohypoaldosteronism, or PHA1.[766] Many of these mutations target the receptor DBD, including nonsense, missense, and frameshift mutations (Figure 4.5). The missense mutations in PHA1 include the mutation of the Zn^{2+}-coordinating Cys645 to serine, which would be devastating for folding of the zinc finger (Figure 4.6a). Additional PHA1 missense mutations include the mutation of lever arm Gly633 to arginine[767] and the mutation of Arg659 to serine at the DNA binding interface.[766] The Arg659 mutation removes a charge-charge interaction between the MR DBD and the DNA backbone, likely reducing DNA binding activity without altering sequence specificity (Figure 4.6b). The lever arm mutation of Gly633 to arginine does not alter DNA binding affinity,[766] and the structure of the MR DBD reveals this residue is solvent exposed (Figure 4.6c). However, this mutation reduces MR's transactivation ability by 40% compared to wild-type receptor, supporting the lever arm's predicted role in receptor activation.[669]

In addition to endocrine disorders, steroid receptors frequently assume malicious roles in cancer, with PR and the estrogen receptor (ER) often driving breast cancer growth and AR driving prostate cancer cell growth.[768–770] While MR and GR are less studied with respect to their action in cancer cells, there is accumulating evidence that these receptors also play key roles in neoplastic diseases.[771–774] A recent study demonstrated that a decrease in MR expression was associated with increased angiogenesis and poor patient survival in colorectal cancer.[773] MR is mutated in up to 6% of colorectal cancer samples on the cBioPortal database and is also frequently ($\geq 5\%$) altered in skin cutaneous melanoma, uterine, bladder, and stomach cancers.[775,776]

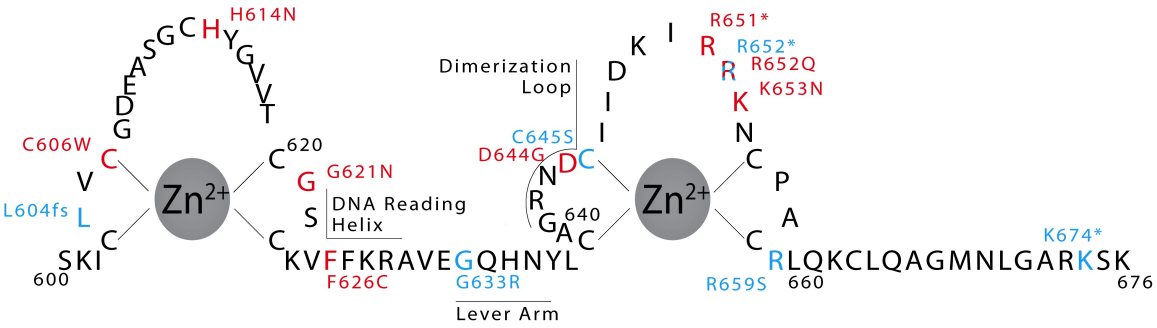


Figure 4.5: MR DBD mutations found in disease. Mutations found in type I pseudohypoaldosteronism are in blue and mutations found in cancer are in red. An asterisk indicates a nonsense mutation, and fs indicates a frameshift mutation.

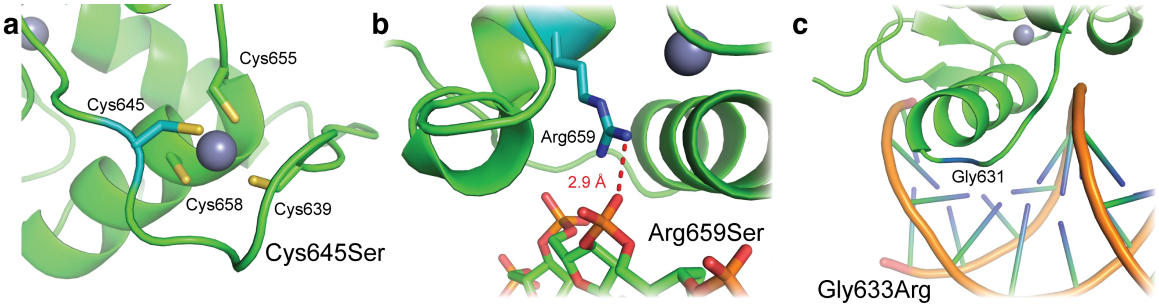


Figure 4.6: MR DBD mutations driving PHA1. Mutated residues are shown in blue. (a) Cys645 is one of four cysteines that coordinate a Zn^{+2} ion in MR's second zinc finger. Its mutation to serine would destroy the zinc finger fold of the DBD. (b) Arg659 makes non-specific interactions with the DNA backbone, and is mutated to serine in some cases of PHA1. (c) Gly633 is part of the DBD lever arm, which is important for receptor activation.[669] Mutation of this residue to arginine affects receptor activation without affecting its affinity for DNA.[767]

Several mutations found in cancer affect the MR DBD, including nonsense mutations that truncate part of or the entire domain (Figure 4.5). Four missense mutations affect the DNA binding interface of the MR DBD (Figure 4.7a-d). His614, which interacts with both the phosphate DNA backbone and a serine side chain, is mutated in a kidney renal papillary cell carcinoma sample to asparagine (Figure 4.7a). Arg652, which also interacts with the DNA backbone, is mutated to glutamine in a uterine corpus endometrioid carcinoma sample (Figure 4.7b). MR Lys653 is mutated to asparagine in multiple cancer types. This residue may make non-specific contacts with the minor groove, but does not have strong electron density to support its side chain position (Figure 4.7c). The most interesting mutation at the DNA interface is that of Gly621 to aspartic acid (Figure 4.7d). This glycine residue is strictly conserved in GR, AR, and PR, but ER contains a glutamic acid at the homologous position (Figure 4.7e).[127] Mutation of the homologous residue in GR, Gly439, to aspartic acid results in a DNA binding domain that poorly discriminates among (+)GREs and estrogen response elements.[777]

Two additional cancer mutations target the hydrophobic core of the MR DBD (Figure 4.7f,g). One mutation found in a glioblastoma multiforme patient targets Cys606, which is one of four cysteine residues that coordinate a Zn^{+2} ion in one of MR's two zinc fingers (Figure 4.7f). Like the mutation of Cys656 in PHA1, this mutation to tryptophan would be devastating for folding of the DBD. A second cancer mutation within the hydrophobic core, Phe626 to cysteine, may also affect the DBD's core fold (Figure 4.7g). Finally, one interesting mutation in colorectal cancer targets the dimerization interface of the MR DBD (Figure 4.7h). In this case, Asp644 is mutated to glycine; this aspartic acid participates in two salt bridge interactions that link the two MR DBD monomers. Such a mutation may affect cooperative binding of the receptor to DNA and subsequent gene activation.

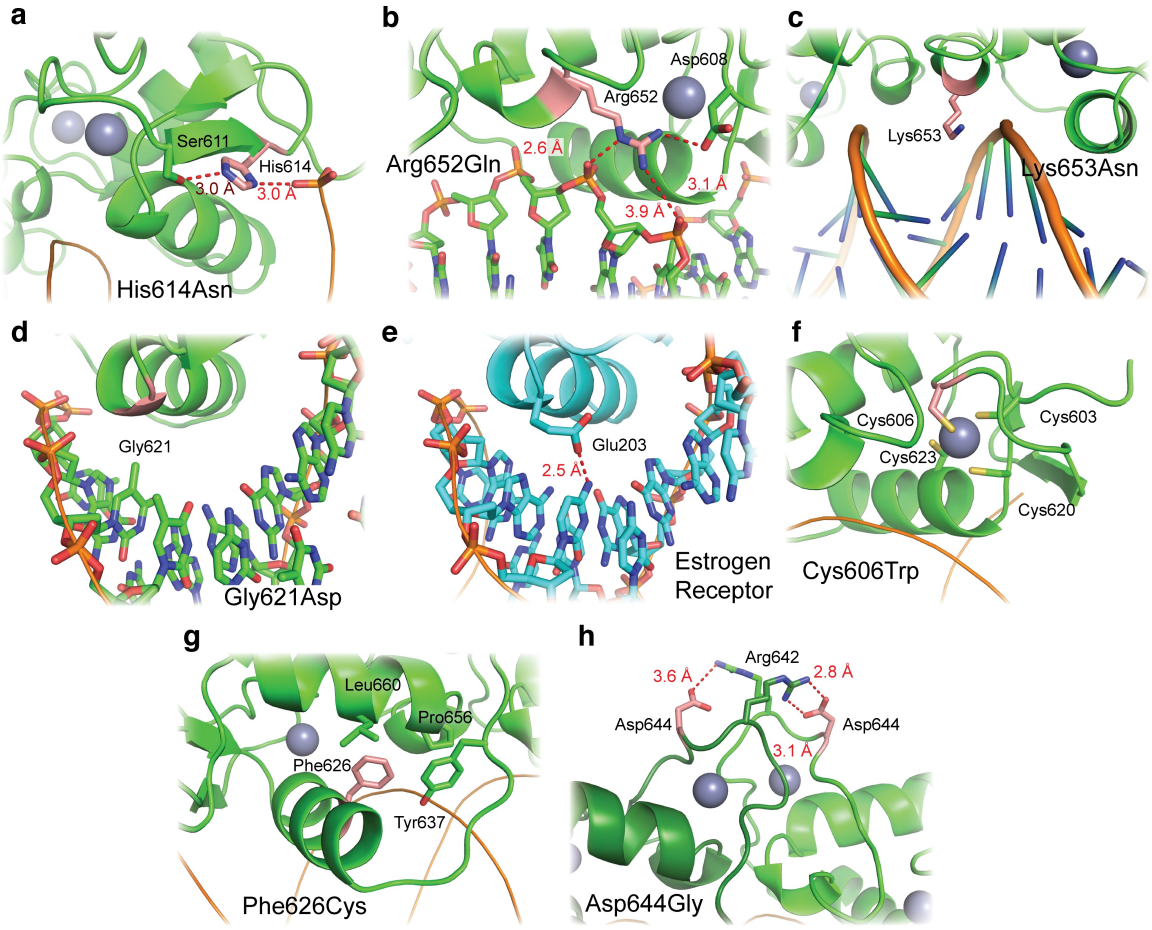


Figure 4.7: MR DBD mutations found in cancer. Mutated residues are shown in red. (a) His614 interacts with both the DNA backbone and Ser611. (b) Similarly, Arg652 also interacts with the DNA backbone and a neighboring amino acid, aspartic acid 608. (c) Lys653 is within an appropriate distance to make non-specific contacts with the minor groove of a (+)GRE. (d) Gly621 is mutated to aspartic acid in a stomach cancer sample. In the estrogen receptor, the homologous amino acid is the similar glutamic acid, which participates in base-specific DNA recognition (panel e). (f) Cys606 is one of four cysteine residues to coordinate a Zn^{+2} ion in one of MR's two zinc fingers. (g) Phe626 comprises part of the hydrophobic core of the DBD. (h) Asp644 is a key mediator of MR dimerization, forming a salt bridge with Arg642 of the second monomer.

In addition to PHA1 and cancer, MR mutation also occurs in hypertension;[778] mutations of genes in steroid metabolic pathways upstream of MR can lead to similar disorders.[779] Some MR DBD mutations in PHA1 are frameshift or nonsense mutations,[766] but many are missense mutations that affect the dimerization, DNA binding, or hydrophobic core structure of the domain (Figure 4.6). The MR mutations found in cancer are very similar, affecting the fold of the DBD, its DNA binding interface, and its dimerization loop. Since these types of mutation diminish MR's transcriptional activity in PHA1, the MR mutations found in cancer also likely abrogate receptor activity. This is consistent with the decreased MR expression found in some types of cancer.[780,781]

Such mutations found in PHA1 and cancer may also lead to structural changes of elements flanking the DBD, such as the nuclear localization sequences immediately to the C- and N-terminus of the MR DBD.[782] Several post-translational modifications also occur at the DBD flanks, including acetylation at Lys677[783] and phosphorylation at Ser601.[782] Amino acid changes in MR are not limited to cancer and PHA1: numerous human SNPs within MR's coding region have been identified, including the change of Val617 to alanine in the DBD (rs373194830).[784] This mutation likely may have minimal effect on MR activity, since the androgen receptor contains an alanine at the homologous position. However, it is possible that such SNPs lead to quantifiable physiological differences, as has been noted with MR polymorphisms in the N-terminal domain.[748] Finally, one MR mutation in cancer may change the DNA binding specificity of the receptor. The stomach cancer mutation of Gly621 to aspartic acid mirrors the mutation of Gly439 previously performed with the GR.[777] This mutation within GR led to a DBD that poorly discriminated among estrogen and glucocorticoid response elements.[777] Such diverse DNA recognition *in vivo* may allow for the receptor to activate a more diverse set of target genes than wild type MR.

MR and GR can heterodimerize,[785] implying that MR mutations that affect the

core structure and dimerization of the receptor may also affect responses to glucocorticoids. Other domains of MR are key for other protein-protein interactions, including the unstructured N-terminal domain.[786] Common polymorphisms in the N-terminal domain lead to phenotypic changes in stress response, including saliva production and heart rate.[748]

In addition to its relevance to human disease, the comparison between MR and GR is interesting due to the different responses activation of each receptor produces despite similar DNA and ligand binding properties. To provide a framework to the interpretation of differences between GR and MR, we report the structure of the MR DBD in complex with a DNA (+)GRE. The DBDs of GR and MR make identical contacts with DNA, but GR buries a larger surface area at the (+)GRE interface and makes very favorable monomer-DNA interactions at nGREs.[333] Although MR binding to consensus nGREs[574] has not been tested, this difference in DNA binding may allow GR to bind to a greater diversity of DNA sequences. ChIP-seq analyses have found a large number of motifs at GR binding sites on genomic DNA, including not only the (+)GRE but also AP-1, AML1, UNKN, NF-κB, HNF3, TAL1, and NF1 response elements.[677] In ChIP-seq studies of the MR, only palindromic motifs were explored; it is unclear whether MR binds a similarly wide array of genomic binding sites.[754] Future work is required to determine how such similar receptors can effect disparate function *in vivo* and whether this disparate function is based on differences in DNA binding preference.

4.5 Materials and Methods

4.5.1 Protein expression and purification

The DNA binding domain (DBD) of the human MR (amino acids 593-671, UniProt P08235.1) was cloned into the pMCSG7 vector, which contains a 6X-histidine tag.

Both the MR and GR DBDs were expressed and purified as described previously (see Chapter 2). Briefly, BL-21 (DE3) pLysS *E. coli* transformed with the expression construct were grown in TB media. At an OD_{600} of 0.8, cultures were induced with 300 μM IPTG for four hours at 30 °C. Cells were pelleted via centrifugation at 4,000 *g* for 20 minutes and frozen at -80 °C until purification.

For purification, cells were thawed and resuspended in a buffer containing 1 M NaCl, 20 mM Tris-HCl pH 7.4, 5% glycerol, and 25 mM imidazole. Cells were lysed on ice via sonication and centrifuged for 1 hour at 4 °C and 35,000 *g*. DBD was purified from the supernatant using a 5 ml HisTrap affinity column followed by gel filtration with a HiPrep 26/60 S300 Sephacryl column (GE Healthcare) into a buffer containing 100 mM NaCl, 20 mM Tris-HCl pH 7.4, and 5% glycerol. Protein was concentrated to 3 mg/ml, flash frozen, and stored at -80 °C.

4.5.2 Crystallization, data collection, and structure determination

Crystals of the MR DBD - (+)GRE complex were grown by hanging-drop vapor diffusion in 0.2 M sodium malonate and 12% PEG 3350 at a protein concentration of 3.0 mg/ml and a 2:1 molar ratio of DNA to protein. Crystals were cryoprotected in 0.2 M sodium malonate, 12% PEG 3350 and 20% glycerol and flash cooled in liquid N_2.

Data were collected remotely on the 22-ID beamline at the Southeast Regional Collaborative Access Team (SER-CAT) at the Advanced Photon Source (Argonne, IL, USA). Data were processed using HKL-2000 software and phased using previously solved structures of the GR DBD bound to (+)GREs.[669] Phasing, refinement, and omit map generation were performed in the PHENIX software suite (version 1.9_1692).[691] Model building was performed in COOT (version 0.6.1).[690] As the data are highly anisotropic, completeness is only 67% in the highest resolution shell (Table

1), despite 49.2% of reflections in the shell from 2.39 - 2.30 Å having an I / σI > 5. To balance completeness and using the available data, 2.39 Å was chosen as the resolution cutoff for refinement. The PyMOL software suite (Schrödinger, LLC) was used to visualize the structure and generate figures. Amino acids are numbered according to the human MR sequence (UniProt P08235.1). 3DNA was used to analyze nucleic acid structure,[605] and the PISA server was used to calculate buried surface areas of each interface.[722] The coordinates and structure factors for the MR DBD - (+)GRE complex were deposited in the Protein Data Bank under accession code 4TNT.

4.5.3 Nucleic acid binding assay

A synthesized 6-carboxyfluorescein (6-FAM) labeled (+)GRE (Integrated DNA Technologies) was annealed in 10 mM NaCl and 20 mM Tris-HCl pH 8.0 by heating to 90 °C in a 1 L water bath and slow cooling to room temperature. The (+)GRE sequences used for binding were 5'-[FAM]CCAGAACAGAGTGTTCTGA-3' and 5'-TCAGAACACTCTGTTCTGG-3', where [FAM] indicates the position of 6-FAM. Indicated amounts of DBD were added to wells containing 10 nM of 6-FAM-labeled (+)GRE, and formation of DBD - GRE complexes was monitored by fluorescence polarization with a Biotek Synergy plate reader at an excitation wavelength of 485 nm and emission wavelength of 528 nm. Reactions were performed in buffer containing 100 mM NaCl, 20 mM Tris-HCl (pH 7.4), and 5% glycerol. Prism version 6.0d (Graphpad Software, Inc.) was used for data analysis and graph generation.

4.5.4 Cancer mutations

Cancer mutations were accessed via the cBioPortal for Cancer Genomics.[776]

4.5.5 Sequences

Sequence numbering for the steroid receptors are for the human proteins and derived from the following sequences: androgen receptor, UniProtKB P10275.2; progesterone receptor, UniProtKB P06401.4; mineralocorticoid receptor, UniProt P08235.1; glucocorticoid receptor, GenBank ADP91252.1.

Chapter 5

The evolution of DNA-dependent transrepression by steroid receptors

5.1 Abstract

Gene duplication is a key factor in the acquisition of novel protein function over evolution. Most species, including humans, encode numerous families of paralogs - genes with divergent function that evolved from a common ancestral gene after a duplication event. Promiscuity of ancestral genes is a major factor in evolution, and many models have been proposed to explain how new functions arise and are subdivided among paralogous proteins. Understanding these processes is critical for understanding evolutionary processes as well as developing therapies to target specific proteins contained within larger gene families. In this work, we examine the mechanism by which two ancestral, dissimilar, and sequence-specific DNA binding functions were subdivided and enhanced throughout the evolution of a remarkably small protein domain, the DNA binding domain of steroid hormone receptors.

This chapter is adapted from a manuscript in preparation for publication:
Hudson WH, Kossmann BK, de Vera IMS, Chuo S-W, Weikum ER, Ivanov I, Kojetin DJ, Ortlund EA. Epistasis drives the evolution of a multifunctional protein domain.

5.2 The evolution of DNA-dependent transrepression by steroid receptors

Steroid hormones are cholesterol-derived molecules that mediate long-range, long-lasting physiologic effects by signaling through steroid hormone receptors (SRs) in target tissue.[787,788] SRs are members of the nuclear receptor superfamily, consisting of 48 proteins (in humans) that are important drug targets due to their powerful transcriptional effects in response to lipophilic ligands.[789,790] The glucocorticoid receptor (GR), a SR, is the target of more approved drugs than any other protein, due to its powerful anti-inflammatory transcriptional effects.[790] GR is capable of both activating[791,792] and repressing[574,576,793] transcription in response to a DNA binding event. To elicit these opposite responses, the GR's DNA binding domain (DBD) is capable of recognizing both activating glucocorticoid response elements ((+)GREs) as well as repressive negative glucocorticoid response elements (nGREs) in a sequence-specific manner (Figure 5.1a). These two modes of binding require alternate kinetics: GR - (+)GRE binding exhibits positive cooperativity, while GR - nGRE binding shows negative cooperativity (Chapter 3). Likewise, 2D NMR spectra of ^{15}N-labeled GR DBD bound to a (+)GRE and nGRE are strikingly divergent: the DBD's dimerization loop undergoes significant perturbation upon (+)GRE - but not nGRE - binding, demonstrating that the GR DBD adopts distinct DNA-bound conformers to elicit opposing transcriptional responses (Figure 5.1b, Figure 5.2).

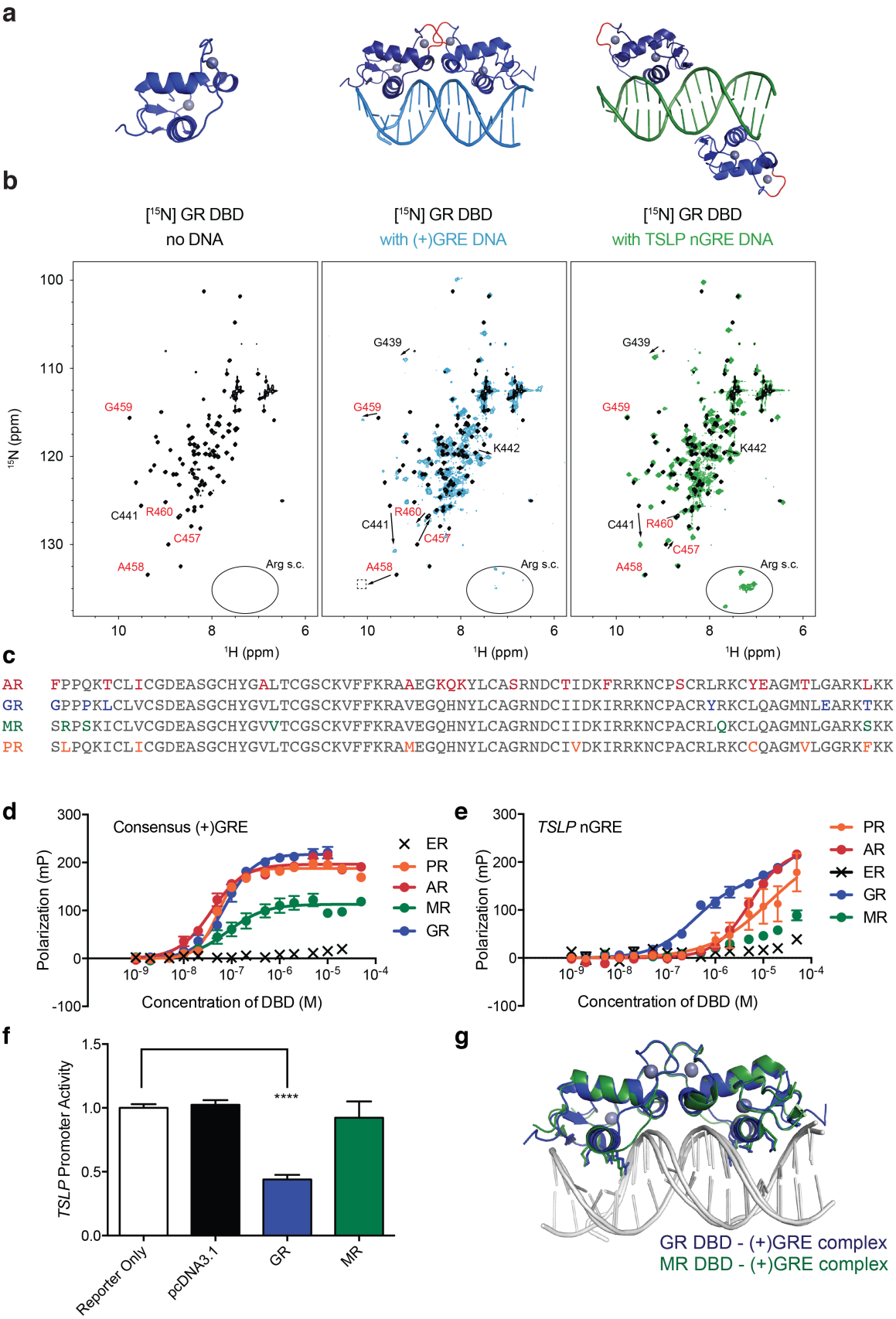

a
b
[15N] GR DBD
no DNA
[15N] GR DBD
with (+)GRE DNA
[15N] GR DBD
with TSLP nGRE DNA
15N (ppm)
1H (ppm)
G459
C441
R460
C457
A458
Arg s.c.
G439
K442
c
AR FPPQKTCLICGDEASGCHYGALTCGSCKVFFKRAAEGKQKYLCASRNDCTIDKFRRKNCPSCRLRKCYEAGMTLGARKLKK
GR GPPPKLCLVCSDEASGCHYGVLTCGSCKVFFKRAVEGQHNYLCAGRNDCIIDKIRRKNCPACRYRKCLQAGMNLEARKTKK
MR SRPSKICLVCGDEASGCHYGVVTCGSCKVFFKRAVEGQHNYLCAGRNDCIIDKIRRKNCPACRLQKCLQAGMNLGARKSKK
PR SLPQKICLICGDEASGCHYGVLTCGSCKVFFKRAMEGQHNYLCAGRNDCIVDKIRRKNCPACRLRKCCQAGMVLGGRKFKK
d
Consensus (+)GRE
Polarization (mP)
Concentration of DBD (M)
ER
PR
AR
MR
GR
e
TSLP nGRE
f
TSLP Promoter Activity
****
Reporter Only
pcDNA3.1
GR
MR
g
GR DBD - (+)GRE complex
MR DBD - (+)GRE complex

Figure 5.1: The glucocorticoid receptor (GR) DNA binding domain (DBD) uniquely adopts distinct conformations to activate and repress transcription. (a) At left, the NMR structure of free GR DBD in solution[794] is shown. In the center, the GR DBD can dimerize on an inverted repeat (+)GRE element[669] to activate transcription, or can bind as two monomers on an nGRE to repress transcription (left, see also Chapter 3). The dimerization loop of the DBD is shown in red. (b) NMR spectra of ^{15}N-labeled GR DBD in the absence of DNA (left), with (+)GRE DNA (center), and nGRE DNA (right). While DNA contacting residues (such as Gly439 or Lys442) are shifted upon nGRE or (+)GRE binding, residues of the dimerization loop (red) are significantly shifted upon (+)GRE binding but not nGRE binding. Additionally, significant differences between (+)GRE- and nGRE-bound GR DBD occur at other residues, including the appearance of arginine side chain density at lower right, which indicates altered rates of exchange of DBD - (+)GRE and DBD - nGRE complexes. Together, these results indicate that GR DBD adopts two distinct conformations in order to activate or repress transcription when bound to DNA. See also Fig 5.2. (c) Sequence alignment of the human 3-keto steroid receptor DBDs; consensus sequence is shown in gray, and difference from the consensus sequence is shown in color. (d) All 3-keto steroid receptor DBDs bind to a consensus (+)GRE with nanomolar affinity, but only human GR binds to nGREs with nanomolar affinities (e). (f) While full-length human GR is capable of repressing a constitutively active nGRE-containing reporter, MR is not, despite significant sequence (g) and structural similarity within the DBDs (ref. 669 and Chapter 4). In (f), **** indicates $P < 0.0001$ by two-way ANOVA followed by Tukey's post-hoc test. In all panels, points or bars indicate mean, and error bars represent SEM.

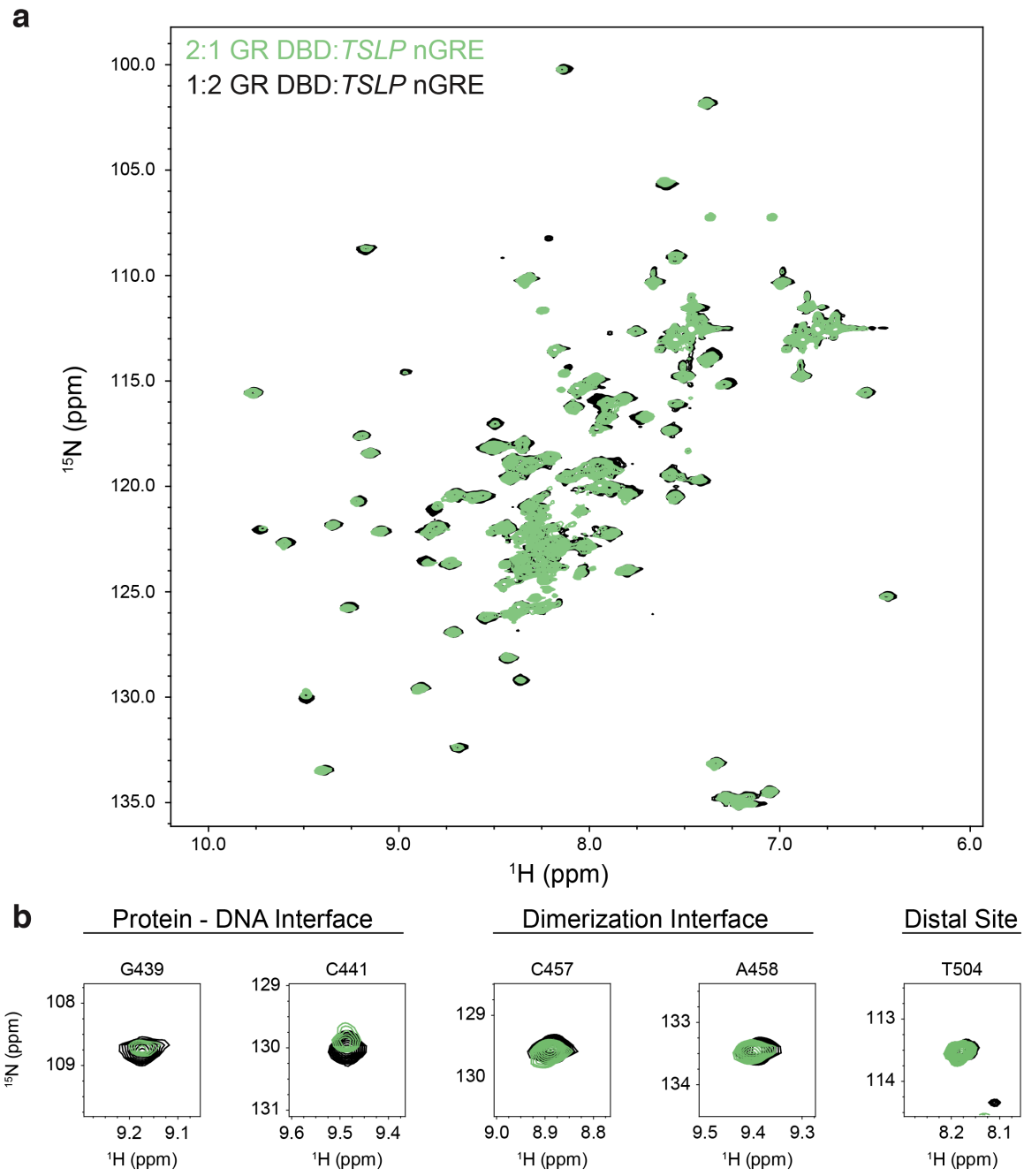


Figure 5.2: The interaction of GR DBD with nGREs is characterized by two non-identical, monomeric binding events. (a) NMR spectra of ^{15}N-labeled GR DBD in the presence of 0.47x or 2.1x *TSLP* nGRE DNA (approximately 2:1 and 1:2, respectively). In the 0.47x sample (black), two proteins are likely bound to one nGRE - one to a high affinity site and one to a lower affinity site - and in the 2.1x (+)GRE (green), it is likely that a single protein is bound to the higher affinity site. (b) Peaks of individual residues at the DNA binding interface (Gly439, Cys441) are affected by the lower affinity dimer interaction, supporting the previously-observed low micromolar affinity for a second DNA binding event. Residues in the dimerization interface (Cys457, Ala458) are not affected, indicating that the GR DBD does not dimerize on nGREs as it does on (+)GREs. Finally, a distant residue to the DNA binding interface (Thr504), is not affected, indicating that the chemical shifts observed are specific to the DNA binding interface.

GR, along with the mineralocorticoid, progesterone, and androgen receptors (MR, PR, and AR, respectively), comprise the 3-keto steroid binding (NR3) subfamily of nuclear receptors. All of these receptors are capable of binding (+)GREs through a highly conserved DBD (Figure 5.1c-d). However, only the GR DBD can bind to the *TSLP* nGRE with nanomolar affinity (Figure 5.1e), indicating that the conformational diversity of GR's DNA binding ability may be unique among the steroid receptors. Furthermore, full-length MR - the most closely related paralog of GR - is incapable of repressing a constitutively driven nGRE-containing reporter in cells (Figure 5.1f). This divergence in function is quite remarkable given the close sequence (Figure 5.1c) and structural (Figure 5.1g) similarity of GR and MR, as well as the small size of their DBDs (75 amino acids).

To trace the evolutionary history of divergent response element specificity among the steroid receptors from their well-established phylogeny,[795] we determined sequences of ancestral DBDs from key nodes within the 3-keto steroid receptor evolutionary lineage (Tables 5.1, 5.2, 5.3, 5.4, 5.5); all reconstructed DBDs had very high posterior probabilities (Figure 5.3). To determine the ability of the DBDs within the GR/MR lineage to repress transcription at nGREs, we used overlap-extension PCR to insert these DBDs into the full-length human GR (Figure 5.4a) and tested their ability to both activate a simple (+)GRE reporter and repress the *TSLP* nGRE under a constitutively-active promoter. As expected, all extant and ancestral proteins within both the GR and MR lineage activated a (+)GRE reporter (Figure 5.4b-c). Surprisingly, the AncSR2 and AncCR DBDs - ancestors common to the GR and MR lineage - were able to repress the *TSLP* nGRE from baseline, although these activities were not statistically significant when corrected for multiple comparisons (Figure 5.4d). This slight repressive ability was lost in the ancestral MR (AncMR; Figure 5.4e) but enhanced in the GR lineage (Figure 5.4d), supporting the hypothesis that subfunctionalization of ancestral moonlighting functions is an important mechanism

of divergent function among paralogs.[796] In particular, the transition from AncGR to AncGR2 - which represent the ancestor of cartilaginous fishes and bony vertebrates and the ancestor of ray- and lobe-finned fishes, respectively[797] - caused a dramatic increase in the ability of GR to repress at nGREs (Figure 5.4d).

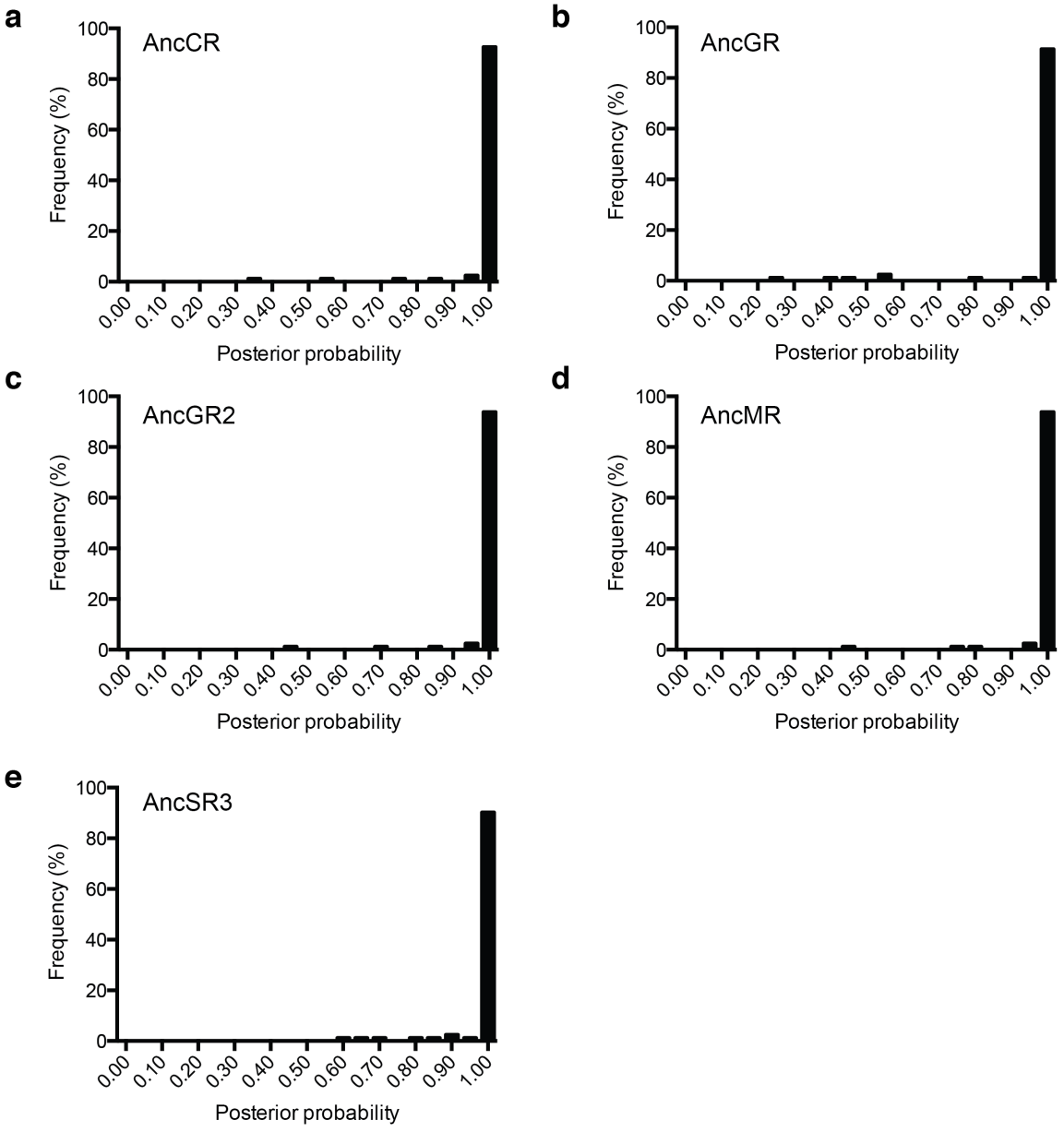


Figure 5.3: Distribution of the posterior probabilities of individual amino acid residues among the ancestral DBDs reconstructed in this study. Mean posterior probabilities are 0.98 for AncCR (a), 0.96 for AncGR1 (b), 0.99 for AncGR2 (c), 0.97 for AncMR (d), and 0.98 for AncSR3 (e).

Position	ML state	PP	Alt. State 1	PP	Alt. State 2	PP	Alt. State 3	PP
415	G	0.42	S	0.407	A	0.098	T	0.034
416	P	0.964	S	0.022	T	0.005	A	0.004
417	P	0.977	T	0.009	S	0.007	A	0.004
418	P	0.273	Q	0.258	H	0.239	S	0.199
419	K	0.998	R	0.001				
420	I	0.549	V	0.355	L	0.068	M	0.021
421	C	1						
422	L	0.997	M	0.001				
423	V	0.993	I	0.007				
424	C	1						
425	G	0.552	S	0.415	A	0.017	N	0.004
426	D	1						
427	E	0.997	D	0.002				
428	A	1						
429	S	1						
430	G	1						
431	C	1						
432	H	1						
433	Y	1						
434	G	1						
435	V	1						
436	L	0.996	I	0.001	M	0.001	V	0.001
437	T	1						
438	C	1						
439	G	1						
440	S	1						
441	C	1						
442	K	1						
443	V	1						
444	F	1						
445	F	1						
446	K	1						
447	R	1						
448	A	1						
449	V	0.999	I	0.001				
450	E	1						
451	G	1						
452	Q	0.999						
453	H	0.999						
454	N	0.998	D	0.001	S	0.001		
455	Y	1						
456	L	0.999						
457	C	1						
458	A	1						
459	G	1						
460	R	1						
461	N	1						
462	D	1						
463	C	1						
464	I	1						
465	I	1						
466	D	1						
467	K	1						
468	I	1						
469	R	1						
470	R	1						
471	K	1						
472	N	1						
473	C	1						
474	P	1						
475	A	1						
476	C	1						
477	R	1						
478	F	0.986	L	0.012	Y	0.002		
479	R	1						
480	K	1						
481	C	1						
482	L	0.999	F	0.001				
483	Q	1						
484	A	1						
485	G	1						
486	M	1						
487	N	1						
488	L	1						
489	E	0.468	D	0.419	G	0.112	A	0.001
490	A	1						
491	R	1						
492	K	1						
493	T	0.82	S	0.171	N	0.005	A	0.003
494	K	1						
495	K	1						

Table 5.1: Maximum likelihood states (ML states) and posterior probabilities (PPs) for the reconstructed AncCR DBD. Amino acid positions are numbered to correspond to the human glucocorticoid receptor α (GenBank ADP91252.1).

Position	ML state	PP	Alt. State 1	PP	Alt. State 2	PP	Alt. State 3	PP
415	S	0.907	F	0.052	A	0.016	T	0.008
416	P	0.643	S	0.276	L	0.026	Q	0.02
417	P	1						
418	Q	1						
419	K	1						
420	I	0.887	V	0.063	L	0.024	T	0.02
421	C	1						
422	L	1						
423	I	1						
424	C	1						
425	G	1						
426	D	1						
427	E	1						
428	A	1						
429	S	1						
430	G	1						
431	C	1						
432	H	1						
433	Y	1						
434	G	1						
435	V	0.998	A	0.002				
436	L	0.999	I	0.001				
437	T	1						
438	C	1						
439	G	1						
440	S	1						
441	C	1						
442	K	1						
443	V	1						
444	F	1						
445	F	1						
446	K	1						
447	R	1						
448	A	1						
449	V	0.718	I	0.174	A	0.093	S	0.013
450	E	1						
451	G	1						
452	Q	0.99	H	0.005	K	0.005		
453	H	1						
454	N	0.996	K	0.004				
455	Y	1						
456	L	1						
457	C	1						
458	A	1						
459	G	0.999	S	0.001				
460	R	1						
461	N	1						
462	D	1						
463	C	1						
464	I	0.995	T	0.005				
465	I	0.998	V	0.001				
466	D	1						
467	K	1						
468	I	1						
469	R	1						
470	R	1						
471	K	1						
472	N	1						
473	C	1						
474	P	1						
475	A	0.996	S	0.004				
476	C	1						
477	R	1						
478	L	1						
479	R	1						
480	K	1						
481	C	1						
482	C	0.845	F	0.129	Y	0.021	L	0.005
483	Q	0.996	A	0.002	E	0.001	H	0.001
484	A	1						
485	G	1						
486	M	1						
487	T	0.605	V	0.391	I	0.002	A	0.001
488	L	1						
489	G	1						
490	A	0.996	G	0.004				
491	R	0.998	H	0.002				
492	K	0.98	R	0.02				
493	L	0.966	S	0.024	F	0.008		
494	K	0.999	R	0.001				
495	K	1						

Table 5.2: Maximum likelihood states (ML states) and posterior probabilities (PPs) for the reconstructed AncSR3 DBD. Amino acid positions are numbered to correspond to the human glucocorticoid receptor α (GenBank ADP91252.1).

Position	ML state	PP	Alt. State 1	PP	Alt. State 2	PP	Alt. State 3	PP
415	G	0.42	S	0.407	A	0.098	T	0.034
416	P	0.964	S	0.022	T	0.005	A	0.004
417	P	0.977	T	0.009	S	0.007	A	0.004
418	P	0.273	Q	0.258	H	0.239	S	0.199
419	K	0.998	R	0.001				
420	I	0.549	V	0.355	L	0.068	M	0.021
421	C	1						
422	L	0.997	M	0.001				
423	V	0.993	I	0.007				
424	C	1						
425	G	0.552	S	0.415	A	0.017	N	0.004
426	D	1						
427	E	0.997	D	0.002				
428	A	1						
429	S	1						
430	G	1						
431	C	1						
432	H	1						
433	Y	1						
434	G	1						
435	V	1						
436	L	0.996	I	0.001	M	0.001	V	0.001
437	T	1						
438	C	1						
439	G	1						
440	S	1						
441	C	1						
442	K	1						
443	V	1						
444	F	1						
445	F	1						
446	K	1						
447	R	1						
448	A	1						
449	V	0.999	I	0.001				
450	E	1						
451	G	1						
452	Q	0.999						
453	H	0.999						
454	N	0.998	D	0.001	S	0.001		
455	Y	1						
456	L	0.999						
457	C	1						
458	A	1						
459	G	1						
460	R	1						
461	N	1						
462	D	1						
463	C	1						
464	I	1						
465	I	1						
466	D	1						
467	K	1						
468	I	1						
469	R	1						
470	R	1						
471	K	1						
472	N	1						
473	C	1						
474	P	1						
475	A	1						
476	C	1						
477	R	1						
478	F	0.986	L	0.012	Y	0.002		
479	R	1						
480	K	1						
481	C	1						
482	L	0.999	F	0.001				
483	Q	1						
484	A	1						
485	G	1						
486	M	1						
487	N	1						
488	L	1						
489	E	0.468	D	0.419	G	0.112	A	0.001
490	A	1						
491	R	1						
492	K	1						
493	T	0.82	S	0.171	N	0.005	A	0.003
494	K	1						
495	K	1						

Table 5.3: Maximum likelihood states (ML states) and posterior probabilities (PPs) for the reconstructed AncGR DBD. Amino acid positions are numbered to correspond to the human glucocorticoid receptor α (GenBank ADP91252.1).

Position	ML state	PP	Alt. State 1	PP	Alt. State 2	PP	Alt. State 3	PP
415	G	0.929	S	0.066	A	0.004	T	0.001
416	P	0.999	S	0.001				
417	P	0.96	T	0.026	S	0.009	A	0.004
418	P	0.425	H	0.395	Q	0.117	S	0.054
419	K	1						
420	I	0.714	L	0.141	V	0.124	M	0.02
421	C	1						
422	L	1						
423	V	1						
424	C	1						
425	S	0.992	G	0.007	A	0.001		
426	D	1						
427	E	1						
428	A	1						
429	S	1						
430	G	1						
431	C	1						
432	H	1						
433	Y	1						
434	G	1						
435	V	1						
436	L	1						
437	T	1						
438	C	1						
439	G	1						
440	S	1						
441	C	1						
442	K	1						
443	V	1						
444	F	1						
445	F	1						
446	K	1						
447	R	1						
448	A	1						
449	V	1						
450	E	1						
451	G	1						
452	Q	1						
453	H	1						
454	N	1						
455	Y	1						
456	L	1						
457	C	1						
458	A	1						
459	G	1						
460	R	1						
461	N	1						
462	D	1						
463	C	1						
464	I	1						
465	I	1						
466	D	1						
467	K	1						
468	I	1						
469	R	1						
470	R	1						
471	K	1						
472	N	1						
473	C	1						
474	P	1						
475	A	1						
476	C	1						
477	R	1						
478	F	0.982	Y	0.018				
479	R	1						
480	K	1						
481	C	1						
482	L	1						
483	Q	1						
484	A	1						
485	G	1						
486	M	1						
487	N	1						
488	L	1						
489	E	0.987	D	0.013	G	0.001		
490	A	1						
491	R	1						
492	K	1						
493	T	0.848	S	0.131	N	0.019	A	0.001
494	K	1						
495	K	1						

Table 5.4: Maximum likelihood states (ML states) and posterior probabilities (PPs) for the reconstructed AncGR2 DBD. Amino acid positions are numbered to correspond to the human glucocorticoid receptor α (GenBank ADP91252.1).

Position	ML state	PP	Alt. State 1	PP	Alt. State 2	PP	Alt. State 3	PP
415	S	0.787	T	0.127	A	0.075	G	0.005
416	P	0.453	S	0.284	T	0.165	R	0.061
417	P	1						
418	S	0.97	P	0.011	A	0.008	Q	0.006
419	K	1						
420	V	0.771	I	0.226	M	0.001	L	0.001
421	C	1						
422	L	1						
423	V	1						
424	C	1						
425	G	0.974	A	0.016	S	0.01		
426	D	1						
427	E	1						
428	A	1						
429	S	1						
430	G	1						
431	C	1						
432	H	1						
433	Y	1						
434	G	1						
435	V	1						
436	L	0.979	V	0.018	I	0.001	M	0.001
437	T	1						
438	C	1						
439	G	1						
440	S	1						
441	C	1						
442	K	1						
443	V	1						
444	F	1						
445	F	1						
446	K	1						
447	R	1						
448	A	1						
449	V	1						
450	E	1						
451	G	1						
452	Q	1						
453	H	0.997	Q	0.003				
454	N	1						
455	Y	1						
456	L	1						
457	C	1						
458	A	1						
459	G	1						
460	R	1						
461	N	1						
462	D	1						
463	C	1						
464	I	1						
465	I	1						
466	D	1						
467	K	1						
468	I	1						
469	R	1						
470	R	1						
471	K	1						
472	N	1						
473	C	1						
474	P	1						
475	A	1						
476	C	1						
477	R	1						
478	L	0.999						
479	R	1						
480	K	1						
481	C	1						
482	L	0.999	F	0.001				
483	Q	0.997	H	0.001	N	0.001		
484	A	1						
485	G	1						
486	M	1						
487	N	1						
488	L	1						
489	G	1						
490	A	1						
491	R	1						
492	K	1						
493	S	0.985	T	0.013	A	0.001		
494	K	1						
495	K	1						

Table 5.5: Maximum likelihood states (ML states) and posterior probabilities (PPs) for the reconstructed AncMR DBD. Amino acid positions are numbered to correspond to the human glucocorticoid receptor α (GenBank ADP91252.1).

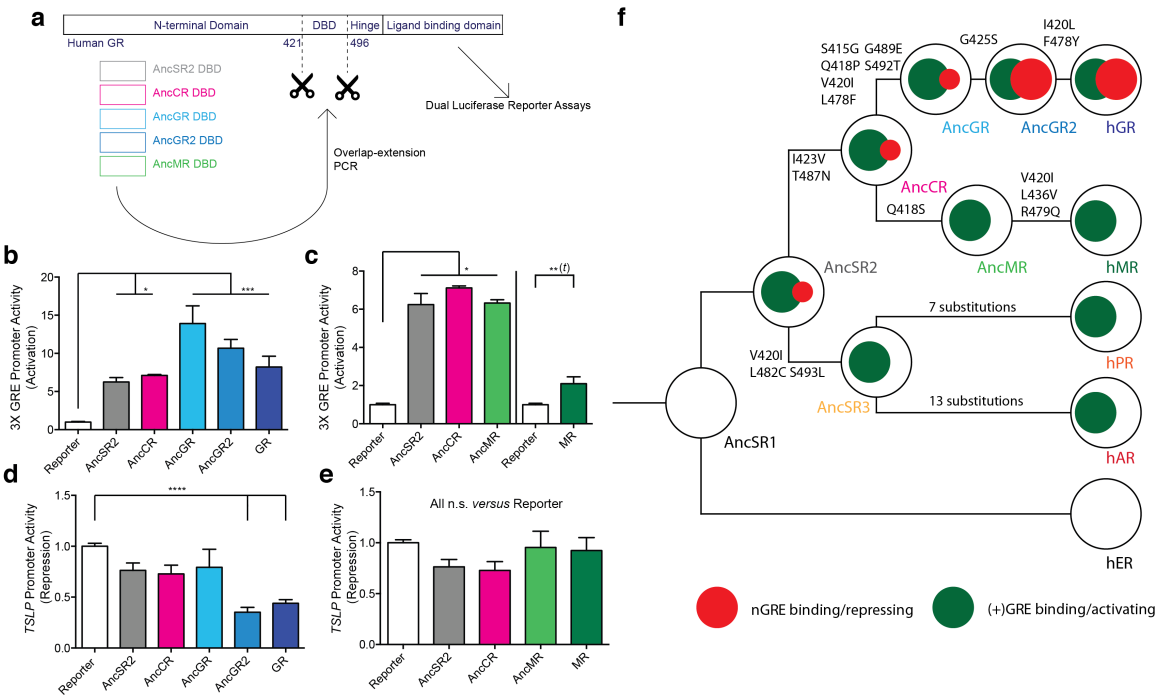


Figure 5.4: The GR lineage improved upon an ancestral cellular repressive function that was lost in MR. (a) Reconstructed ancestral DBDs were inserted into full-length human GR using overlap-extension. These constructs were transfected into HeLa cells with the indicated reporter and treated with 1 μM dexamethasone (or aldosterone, for human MR only). (b) All DBDs in the GR lineage, including AncSR2, are capable of activating a (+)GRE reporter; similar results are found in the MR lineage (panel c). (d) The AncSR2 DBD is capable of repressing transcription from an nGRE-containing promoter, and this ability was retained in the GR lineage and further enhanced at the AncGR2 node. (e) However, the ability repress at nGREs was lost in the MR lineage at AncMR, consistent with *in vitro* binding data (see also Figure 5.5). Above each branch are the amino acid substitutions between each node, using human GR numbering. *, **, ***, and **** indicate $P < 0.05$, 0.01, 0.001, or 0.0001, respectively. One-way ANOVA with Tukey's post hoc test was used in all panels except for hMR vs reporter (Student's t test), because the full-length MR protein was used. Bars represent mean $\pm$ SEM. Experiments were performed at least in duplicate with three internal replicates. (f) To summarize, (+)GRE and nGRE-binding and repressive ability of SR DBDs are mapped onto their phylogeny. (+)GRE binding was derived at the AncSR2 node,[798] along with a moonlighting nGRE binding and transrepressive function. While (+)GRE binding was preserved throughout the clade, nGRE binding was lost at AncCR and AncSR3 and preserved (and enhanced) in the GR lineage. Green circles represent the ability of a given DBD to bind to and activate from (+)GREs. Red circles indicate the ability of a DBD to repress the *TSLP* nGRE in cells more or less than 50% (large and small circles, respectively). Above each branch are the amino acid substitutions between each node, using human GR numbering.

That so few mutations led to divergence of function within a small protein domain allows a unique opportunity to pursue the detailed mechanisms by which functions are altered and distributed among paralogous proteins. To obtain the biochemical and structural mechanisms by which repression at nGREs was selectively retained and enhanced in the GR lineage, we recombinantly expressed and tested all ancestral DBDs for *in vitro* binding to both (+)GRE and nGRE DNA (Figure 5.4f, Table 5.6). All ancestral DBDs retained binding to a (+)GRE, as expected given the ability of all extant 3-keto steroid receptor DBDs to activate from these sequences (Figure 5.1d and reference 799). The ancestral 3-keto steroid receptor, AncSR2, was capable of binding nGREs with nanomolar affinity, implying that nGRE-binding originated at the ancestor of all 3-keto steroid receptors and has been selectively retained in GR. Unlike human GR, the AncSR2 DBD does not bind to nGREs with negative cooperativity; negative cooperativity of DBD - nGRE binding increased progressively from AncSR2 to human GR (hGR), where two distinct, negatively-cooperative binding sites on nGRE DNA are present (Table 5.6, Figure 5.2). nGRE binding ability was lost in AncSR3, the common ancestor of AR and PR, through the accumulation of three amino acid substitutions that spare (+)GRE binding (Figure 5.4f). Individually, none of these changes affect the affinity of AncSR2 for nGREs or (+)GREs, yet negative epistatic interaction among these substitutions results in an ablation of nGRE binding (Figure 5.5).

Protein-DNA complex	K_d (μM)	SEM (μM)	r^2	Notes
S425A GR DBD - (+)GRE	.026	.002	0.99	
S425A - *TSLP* nGRE	0.132, 9.8	.018, 1.4	1.00	2 site binding event
WT GR - (+)GRE	0.073	.003	1.00	
WT GR - *TSLP* nGRE	0.363, 63.2	0.60, 89.3	1.00	2 site binding event
Ser425Gly GR DBD - (+)GRE	0.085	0.005	1.00	
Ser425Gly - *TSLP* nGRE	48.4	59.6	0.99	
WT MR DBD - (+)GRE	0.055	0.006	0.99	
WT MR DBD - *TSLP* nGRE	21.6	19.8	0.98	
Gly607Ser MR DBD - (+)GRE	0.166	0.025	0.99	
Gly607Ser MR DBD - *TSLP* (+)GRE	38.8	19.6	0.99	
AncSR2 DBD - (+)GRE	0.125	0.007	1.00	
AncSR2 DBD- *TSLP* nGRE	0.369	0.042	0.99	h = 1.45
AncGR DBD - (+)GRE	0.110	0.13	0.91	
AncGR DBD - *TSLP* nGRE	7.4	1.3	0.96	h = 0.94
AncGR2 DBD - (+)GRE	0.119	0.008	1.00	
AncGR2 DBD - *TSLP* nGRE	15.9	8.8	0.98	h = 0.75
AncMR DBD - (+)GRE	0.106	0.005	1.00	
AncMR DBD - *TSLP* nGRE	5.7	1.2	0.99	
WT MR DBD - (+)GRE	0.055	0.006	0.99	
WT MR DBD - *TSLP* nGRE	21.6	19.8	0.98	
Gly607Ser MR DBD - (+)GRE	0.166	0.022	0.99	
Gly607Ser MR DBD - *TSLP* nGRE	44.1	7.6	0.99	
Gln418Ser AncSR2 DBD - (+)GRE	0.087	0.007	0.99	
Gln418Ser AncSR2 DBD - nGRE	1.1	0.159	0.99	
Ile423Val AncSR2 DBD - (+)GRE	0.135	0.010	0.99	
Ile423Val AncSR2 DBD - *TSLP* nGRE	0.569	0.029	1.00	
Thr487Asn AncSR2 DBD - (+)GRE	0.164	0.006	0.99	
Thr487Asn AncSR2 DBD - *TSLP* nGRE	2.2	0.238	0.99	
Gln418Ser Ile423Val AncSR2 DBD - (+)GRE	0.071	0.005	1.00	
Gln418Ser Ile423Val AncSR2 DBD - *TSLP* nGRE	4.5	1.1	0.99	
Gln418Ser Thr487Asn AncSR2 DBD - (+)GRE	0.081	0.011	0.98	
Gln418Ser Thr487Asn AncSR2 DBD - *TSLP* nGRE	6.1	1.9	0.99	
Ile423Val Thr487Asn AncSR2 (AncCR) DBD - (+)GRE	0.187	0.005	1.00	
Ile423Val Thr487Asn AncSR2 (AncCR) DBD - *TSLP* nGRE	1.1	0.090	1.00	h = 0.89
Val420Ile AncSR2 DBD - (+)GRE	0.041	0.004	0.99	
Val420Ile AncSR2 DBD - *TSLP* nGRE	3.3	0.450	1.00	
Leu482Cys AncSR2 DBD - (+)GRE	0.129	0.018	0.98	
Leu482Cys AncSR2 DBD - *TSLP* nGRE	2.4	0.276	1.00	
Ser493Leu AncSR2 DBD - (+)GRE	0.046	0.004	0.99	

Ser493Leu AncSR2 DBD - *TSLP* nGRE	1.7	0.109	1.00	
Val420Ile Leu482Cys AncSR2 DBD - (+)GRE	0.100	0.011	0.99	
Val420Ile Leu482Cys AncSR2 DBD - *TSLP* nGRE	1.5	0.138	0.99	
Val420Ile Ser493Leu AncSR2 DBD - (+)GRE	0.239	0.060	0.98	
Val420Ile Ser493Leu AncSR2 DBD - *TSLP* nGRE	16.1	3.7	1.00	
Leu482Cys Ser493Leu AncSR2 DBD - (+)GRE	0.126	0.022	0.97	
Leu482Cys Ser493Leu AncSR2 DBD - *TSLP* nGRE	5.2	3.1	0.93	
AncSR3 DBD - (+)GRE	0.161	0.015	0.99	
AncSR3 DBD - *TSLP* nGRE	24.8	19.4	0.96	
PR DBD - nGRE	13.5	18.3	0.61	
AR DBD - nGRE	5.2	0.536	0.97	
ER DBD - nGRE	n/a			Fit does not converge
PR DBD - (+)GRE	0.098	0.010	0.91	
AR DBD - (+)GRE	0.060	0.004	0.97	
MR DBD - (+)GRE	0.055	0.003	0.99	
ER DBD - (+)GRE	n/a			Fit does not converge

Table 5.6: Interaction of DBDs with (+)GRE and nGREs, as measured by fluorescence polarization.

While the DBD of the ancestral corticosteroid receptor (AncCR), the ancestor of MR and GR, retained nGRE-binding ability, one additional mutation in the MR lineage resulted in a DBD deficient in nGRE binding, consistent with cellular repression data (Figure 5.4e, Figure 5.5). Again, this mutation acted in concert with the two amino acid substitutions from AncSR2 to AncCR to selectively abolish nGRE binding. Both computational and directed evolution studies have implicated epistasis as a primary factor in molecular evolution,[800,801] but whether these mechanisms extend to natural and historical protein divergence has remained an open question.[802] Here, we demonstrate that negative epistatic interactions among accumulating neutral amino acid substitutions enables selective loss-of-function events in promiscuous proteins, or even small protein domains.

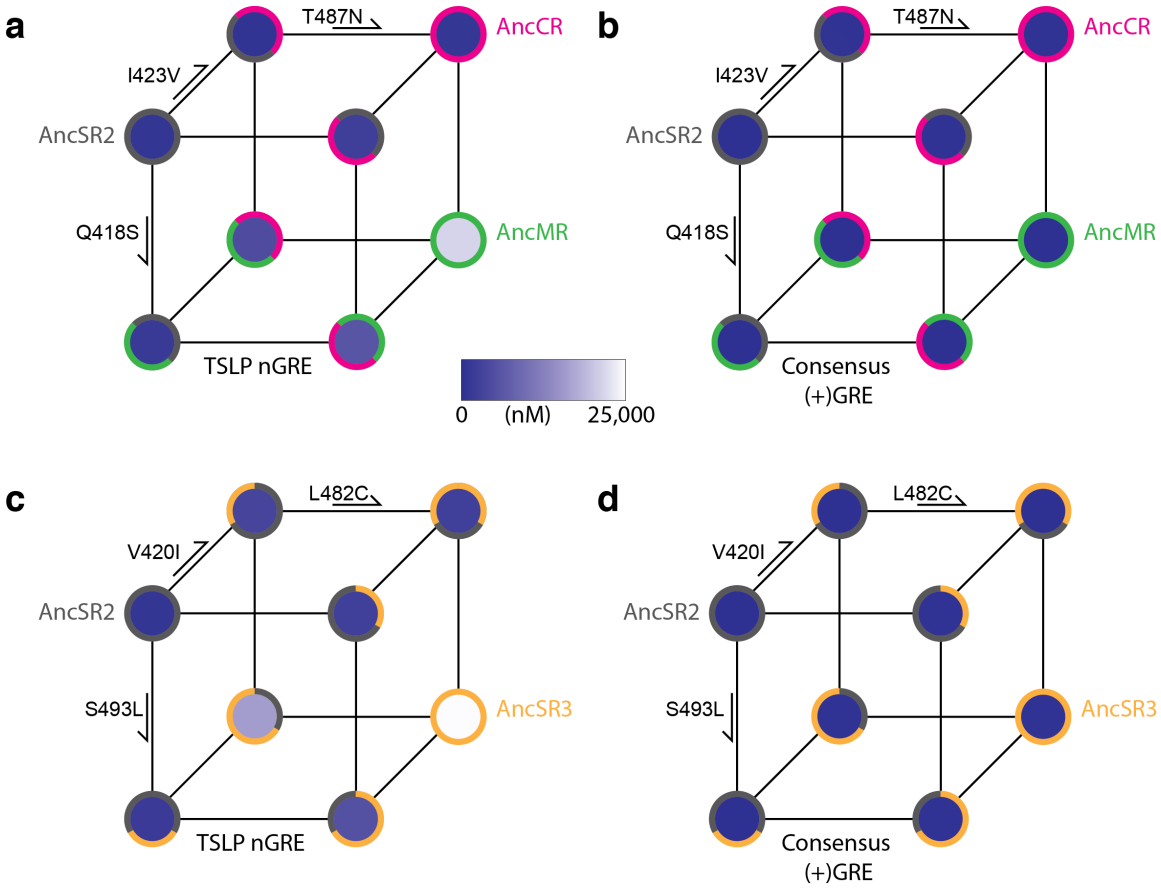


Figure 5.5: Negative epistasis drives the loss of nGRE binding among the 3-keto steroid receptors. *In vitro* binding affinity for indicated DNA response elements is represented on a blue-white scale (center) at each node or amino acid substitution (circles). (a) Three substitutions occurred at the AncSR2 to AncMR transition, none of which affected nGRE binding individually. However, the single substitution from the AncCR to AncMR node (Gln418Ser) interacted in negative epistatic concert with two previous neutral mutations to ablate nGRE binding. (b) None of these mutations (or any combination) affect (+)GRE binding. (c) Three substitutions also define the AncSR2 to AncSR3 transition; no single substitution affects nGRE binding, but the combination is devastating to nGRE binding ability. Again, none of the substitutions or any potential combination affects (+)GRE binding (panel d).

After establishing that GR alone retained an ancestral ability to bind to nGREs, we sought to obtain the structural mechanisms by which this expanded DNA specificity was subfunctionalized to and enhanced in the GR. The crystal structure of the AncSR2 DBD - *TSLP* nGRE complex reveals that the GR - nGRE binding orientation and sequence specificity originated at the ancestor of all 3-keto steroid receptors, before the emergence of vertebrates[803] (Figure 5.6a-d, Table 5.7), despite the superior ability of GR DBD to repress nGRE-mediated transcription in cells (Figure 5.4d). NMR data from (+)GRE- and nGRE-bound GR (Figures 5.1, 5.2) indicated that DBD conformation differs when bound to distinct response elements, and such allosteric changes have been shown to affect transcriptional output.[669,672] Molecular dynamics (MD) trajectories followed by community and sub-optimal path analysis of the steroid receptor - DNA complexes show that AncSR2's daughter proteins in the GR and MR lineage contain more complex community organizations, resulting in diverging allosteric communication upon DNA binding among AncSR2-derived paralogs (Figure 5.7): at nGREs, the community connecting DBD monomers is much larger in hGR versus AncSR2, indicating that DNA-mediated inter-monomer communication and resulting negative cooperativity at nGREs (Chapter 3) is enhanced in hGR as opposed to AncSR2 (Figure 5.6e-h). At (+)GREs, the community structure of AncSR2 bound to a (+)GRE is relatively simple compared to human MR (hMR) and hGR, with each monomer consisting of one large community (α and β) in direct allosteric communication, with some communication diverted through residues in the protein dimerization loops and through the DNA response element itself (Figure 5.7).

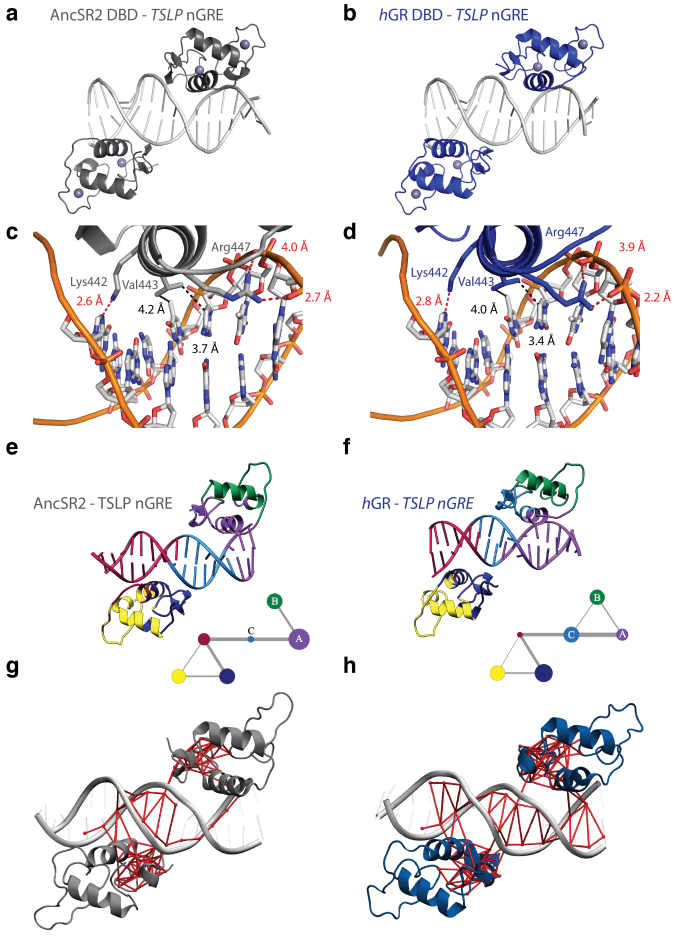


Figure 5.6: Although nGRE binding orientation, stoichiometry, and sequence specificity originated at the common ancestor of all 3-keto SRs, GR is capable of enhanced DNA-mediated allosteric communication at nGREs. (a) Overall structure of the AncSR2 DBD - *TSLP* nGRE complex. (b) Overall structure of the human GR DBD - *TSLP* nGRE complex (see Chapter 3). Both crystal structures revealed two DBD monomers bound as an everted repeat on the opposite sides of DNA. hGR retains ancient amino acid-DNA contacts to ensure sequence specificity when binding nGREs, as shown by the AncSR2 - nGRE (panel c) and human GR - nGRE crystal structures (panel d). (e,f) Community analysis of nGRE-bound AncSR2 and human GR DBDs. At nGREs, each DBD monomer communicates through a central community (community C). Community C largely consists of DNA in the AncSR2 DBD - DNA complex, but this community is expanded to include part of the GR protein in the human GR DBD - DNA complex. The larger community in hGR, relative to AncSR2, enhances communication between DBD monomers at nGREs, as also shown by suboptimal path analysis between the AncSR2 monomers (panel g; 4,158 pathways) and human GR monomers (panel h; 26,165 pathways) when nGRE-bound. This enhanced communication corresponds to an increase at negative cooperativity at nGREs observed throughout the GR lineage.

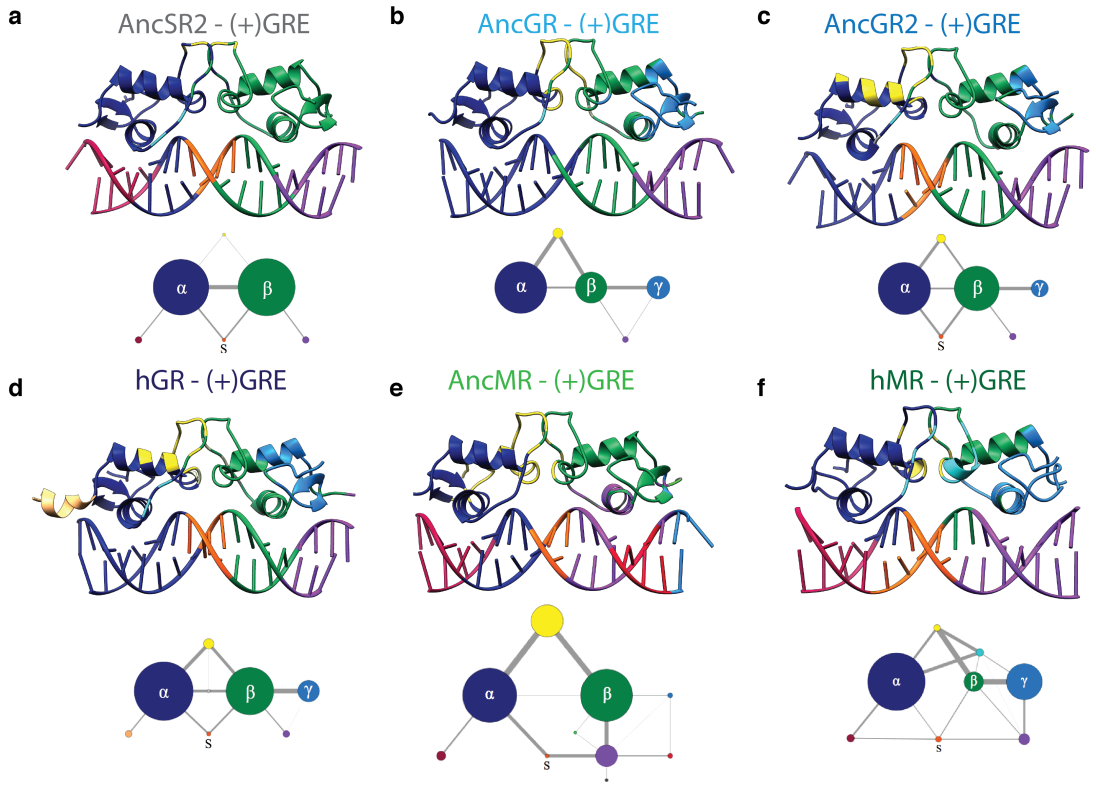


Figure 5.7: Allosteric communication at (+)GREs diverges among paralogous evolutionary pathways. Shown are community analyses of 200 ns MD trajectories performed starting from the crystal structures of (+)GRE-bound AncSR2[798] (panel a), AncGR (panel b), AncGR2 (panel c), human GR[669] (panel d), AncMR (panel e), and human MR (Chapter 4; panel f). The most ancient 3-keto steroid receptor ancestor, AncSR2 (a) possesses a simple community structure, with each monomer consisting of one large community (α and β) with direct communication between α and β, with some diversion of communication through the dimerization loops (yellow) as well as the DNA spacer (community S). Evolution from AncSR2 to AncGR (b) results in a bifurcation of community β - into β and γ, and evolution from AncGR to AncGR2 (c) enhances protein-protein communication outside the dimerization loop (including through S), which is retained in human GR (d). The MR branch of the phylogenetic tree (e,f) has a qualitatively different graph structure, with a greater fraction of the total communication between monomers occuring directly through the dimerization interface, especially in human MR.

Data collection	AncGR - (+)GRE	AncGR2 - (+)GRE	AncMR - (+)GRE	AncSR2 - *TSLP* nGRE	hGR Ser425Gly - (+)GRE
Space group	$P2_1$	C2	$P2_1$	$P2_12_12_1$	C2
Cell dimensions					
a, b, c (Å)	47.8, 81.6, 116.5	131.6, 38.8, 98.3	47.5, 81.2, 116.4	50.3, 72.5, 104.9	176.7, 39.1, 132.1
α,β,γ (°)	90, 97.1, 90	90, 119.7, 90	90, 96.8, 90	90, 90, 90	90, 101.1, 90
Resolution (Å)	40 - 2.00 (2.07-2.00)	50 - 2.00 (2.07-2.00)	40 - 2.20 (2.28-2.20)	45 - 2.40 (2.49-2.40)	50 - 2.30 (2.38-2.30)
R_{merge}	12.5 (51.8)	6.9 (35.5)	9.8 (57.3)	10.4 (69.0)	10.6 (41.8)
I / σI	14.5 (2.1)	20.9 (2.8)	14.7 (1.8)	34.5 (2.1)	13.7 (2.4)
Completeness (%)	95.4 (84.8)	91.3 (61.3)	98.8 (92.2)	99.8 (98.5)	94.2 (79.8)
Redundancy	3.0 (2.6)	4.8 (2.9)	2.8 (2.2)	8.3 (4.0)	3.3 (2.5)
Refinement					
Resolution (Å)	2.0	2.0	2.2	2.4	2.3
No. reflections	57246	27116	44748	15335	36884
R_{work} / R_{free}	21.1 / 22.9	18.0 / 21.6	20.3 / 23.4	21.4 / 23.7	24.5 / 27.9
No. atoms					
Protein/DNA	3731	1882	3697	1745	3736
Water	159	138	120	2	20
B-factors					
Protein / DNA	49.2	37.8	49.2	82.4	70.3
Water	46.3	34.2	41.9	60.7	66.9
R.m.s. deviations					
Bond lengths (Å)	0.005	0.009	0.010	0.012	0.010
Bond angles (°)	0.82	1.17	1.25	1.42	1.41
Ramachandran favored (%)	98	98	97	91	96
Ramachandran outliers (%)	0	0	0	0	1
PDB accession code	5CBX	5CBY	5CBZ	5CC0	5CC1

Table 5.7: X-ray data collection and refinement statistics.

In contrast to both AncSR2 and hGR, such extensive DNA-mediated communication between DBD monomers is not a feature of the MR branch of the SR phylogeny. MD trajectories and resulting community analysis of the AncMR - (+)GRE (Table 5.7) and MR - (+)GRE crystal structures (Chapter 4) reveal a qualitatively different community organization and rotes of communication, compared to their GR paralogs (Figure 5.7). A greater fraction of the total communication between AncMR monomers occurs directly through the dimerization interface - as opposed to through the DNA - relative to the GR lineage, a trend that is more pronounced in hMR. Given the link between transcriptional output and allosteric communication,[669,672] the fracturing of the community organization and weakening of allosteric cohesiveness leads to a network in AncMR and hMR with decreased intra-DNA allosteric communication. This correlates with weaker binding to nGRE sequences, where protein dimerization does not occur. It is noteworthy that all amino acid substitutions driving these allosteric changes occurred far from the DNA binding surface, which is likely necessary to maintain (+)GRE binding (Figure 5.8). Therefore, mutations driving divergent function among structurally similar proteins are likely distant from functional interfaces and act through subtle, allosteric mechanisms.

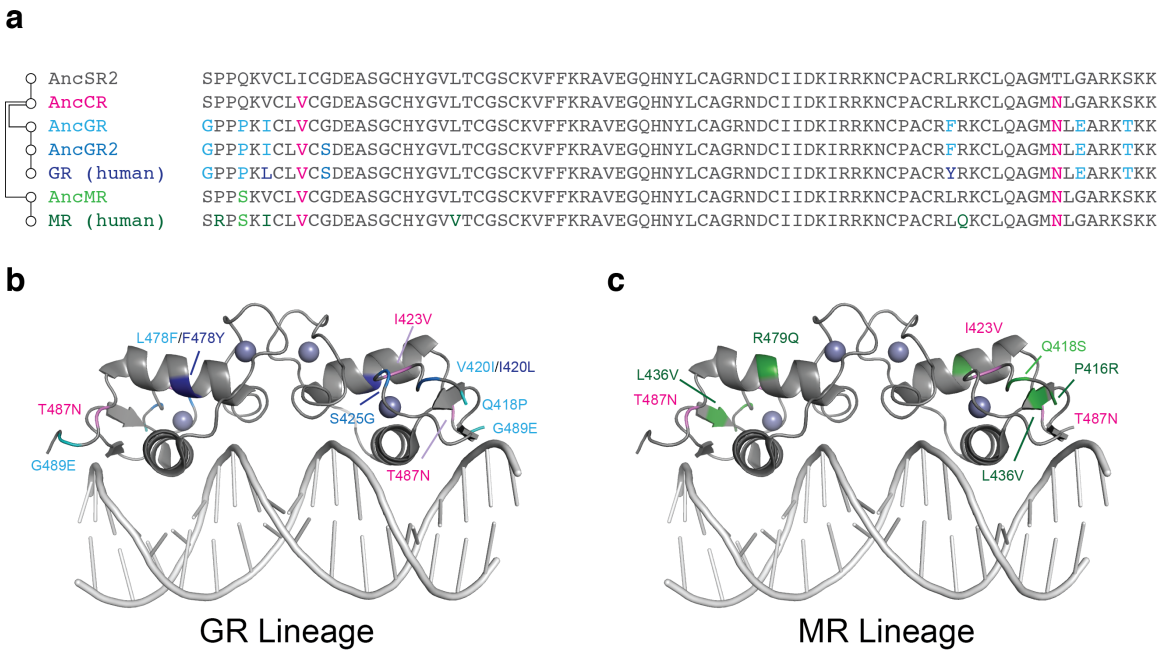


Figure 5.8: Amino acid substitutions that drive changes in DNA binding specificity among paralogous receptors are distant from the nucleic acid-binding interface. (a) Sequence alignment showing AncSR2 and its daughter genes in the GR/MR lineage. Substitutions are colored by the order of their appearance. These substitutions, when mapped onto the three-dimensional structure of the AncSR2 DBD, cluster far from the DNA binding interface in both the paralogous lineages of GR (b) and MR (c).

Such a mutation occurred in the GR lineage: a single substitution - Gly425Ser - greatly enhanced glucocorticoid-mediated repression at nGREs (Figure 5.4). In the AncGR2 - DNA crystal structure (Table 5.7), the sidechain of Ser425 is solvent exposed and makes no contacts with DNA or the remainder of the GR DBD (Figure 5.9a-b). Due to the conformational freedom granted by glycine residues (Figure 5.9c), we hypothesized that subtle changes in backbone conformation may underlie the large effect caused by the Gly425Ser substitution. All non-GR 3-keto SRs - extant and ancestral - contain a glycine at position 425, and these proteins all occupy glycine-only backbone conformations when bound to (+)GREs, as visualized by a Ramachandran plot (Figure 5.9d-f). Intriguingly, the AncSR2 - (+)GRE crystal structure[798] contains two dimers: Gly425 of one dimer occupies glycine-only Ramachandran space and Gly425 of the second dimer occupies general Ramachandran space (Figure 5.9d), indicating that AncSR2 may be more dynamic and able to occupy a wider range of conformational ensembles. The Gly425Ser substitution at the AncGR2 node locked the GR lineage into a restricted subset of Ramachandran space, as compared to the MR/AR/PR lineages (Figure 5.9e-f). MD trajectories show that position 425 of AncGR2 occupies a separate subset of Ramachandran space from AncGR, eliminating any artifacts from crystal packing (Figure 5.9g). As a result of these conformational changes, community analysis reveals that the Gly425Ser substitution decreases direct communication of the two DBD monomers via the dimerization interface and instead increases inter-protein communication via DNA on activating elements, a trend that is preserved in hGR (Figure 5.7). These results are extremely consistent with previous hypotheses that 'evolvable' proteins - such as AncSR2 - may exist in closely-related but functionally distinct conformers, whose distribution may be easily perturbed by mutation, such as the Gly425Ser substitution observed here.[804]

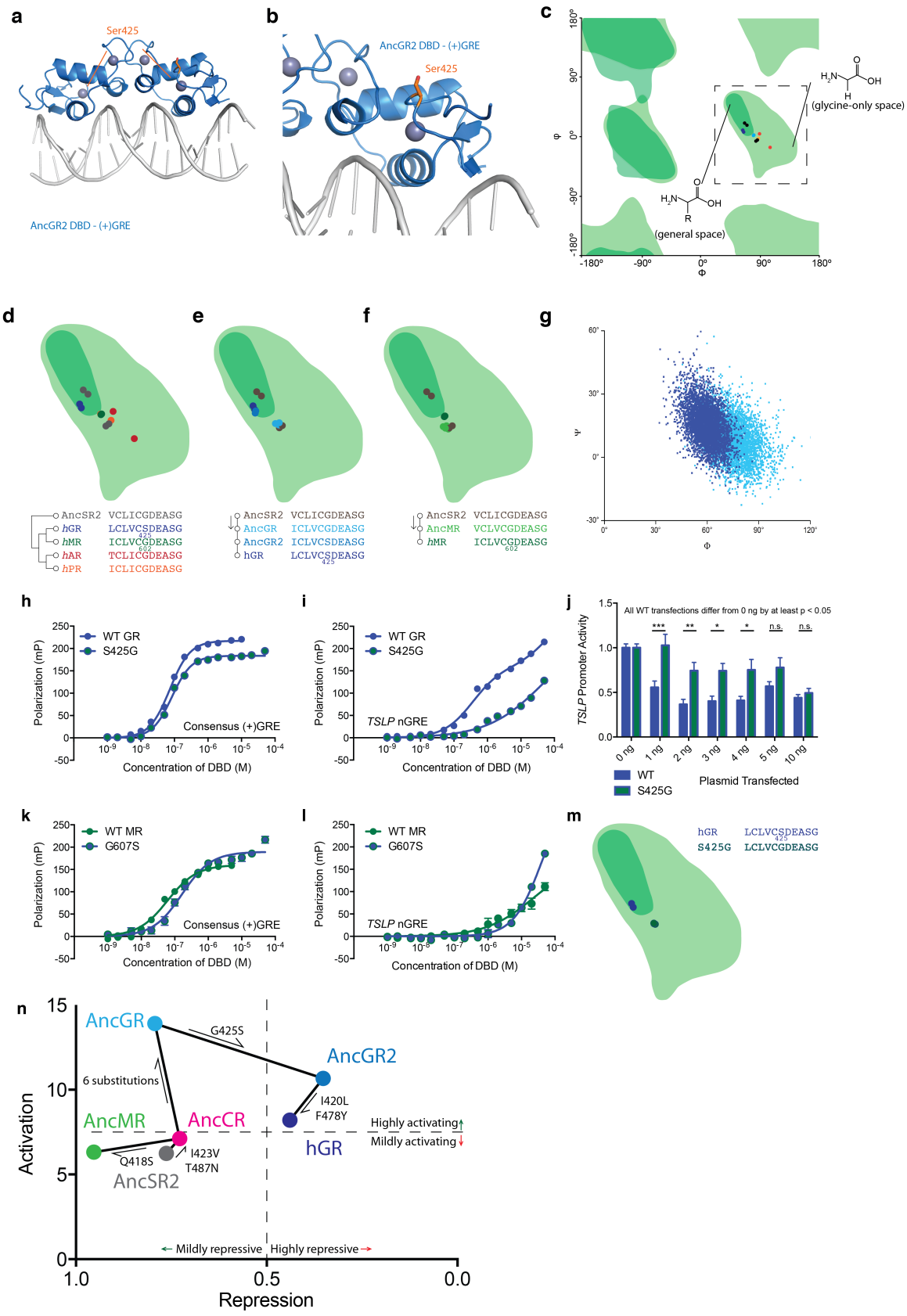

a
Ser425
AncGR2 DBD - (+)GRE
b
AncGR2 DBD - (+)GRE
Ser425
c
(glycine-only space)
(general space)
d
e
f
g
AncSR2 VCLICGDEASG
hGR LCLVCSDEASG
hMR ICLVCGDEASG
hAR TCLICGDEASG
hPR ICLICGDEASG
AncGR ICLVCGDEASG
AncGR2 ICLVCSDEASG
AncMR VCLVCGDEASG
h
i
WT GR
S425G
Consensus (+)GRE
TSLP nGRE
Polarization (mP)
Concentration of DBD (M)
j
All WT transfections differ from 0 ng by at least p < 0.05
TSLP Promoter Activity
WT
S425G
Plasmid Transfected
k
l
WT MR
G607S
m
n
AncGR
AncGR2
AncMR
AncCR
AncSR2
hGR
G425S
6 substitutions
I420L
F478Y
Q418S
I423V
T487N
Highly activating
Mildly activating
Mildly repressive
Highly repressive
Activation
Repression

Figure 5.9: A single amino acid substitution far from the DNA binding interface - Ser425Gly - led to an improvement in nGRE binding through subtle effects in SR backbone conformation. (a) Location of serine 425 within the overall crystal structure of the AncGR2 DBD - (+)GRE complex. (b). Serine 425 is solvent exposed and does not contact any other side chains in the AncGR2 DBD. (c) Ramachandran plot of the AncSR2 and extant human DBDs. (d) Non-glycine (general) amino acids in a protein polymer are restricted in their backbone conformations (phi/psi angles) as compared to glycine-containing residues (dark green vs. light green, respectively). (d) Inset of panel (c); AncSR2 and hGR DBDs occupy general Ramachandran space at position 425, as opposed to the other three extant human DBDs. (e) hGR's conformational space was decided at the AncGR2 node, where the G425S substitution occurred. (f) MR retains a glycine at the homologous position, and assumes a backbone conformation only accessible to glycine residues. (g) Backbone conformation of AncGR - and AncGR2 - (+)GRE complexes differs over the course of 200 ns MD trajectories. (h,i) Comparison of binding between WT and Ser425Gly GR DBD at (+)GREs and nGREs. (j) Full-length WT or Ser425Gly GR was transfected at indicated amounts in to HeLa cells co-transfected with a luciferase target gene controlled by the *TSLP* nGRE and treated with dexamethasone. (k,l) Comparison of binding between WT and Gly607Ser MR DBD at (+)GREs and nGREs. While the Gly607Ser mutation does not decrease binding to (+)GREs (k), the mutation also has no effect on nGRE binding (l). (m) The Ser425Gly mutation reverts GR to an MR-like backbone conformation. *, **, and *** indicate $P < 0.05$, 0.01, or 0.001, respectively. Two-way ANOVA with Sidak's multiple comparisons test was used in all panel c. All error bars and points represent mean ± SEM. (n) Functional landscape of activation and repression among the SR DBDs, showing that the Gly425Ser switch led to a large shift in functional ability in conunction with the conformational changes demonstrated above.

Reversal of this substitution (Ser425Gly) in human GR discriminates between the activating and repressive elements *in vitro*: Ser425Gly retains high-affinity binding to a (+)GRE but much lower affinity for nGREs (Figure 5.9h-i); in cells, full-length GR with the Ser425Gly mutation is a much less potent repressor of the *TSLP* nGRE (Figure 5.9j). Together, these data indicate the two amino acid changes between AncGR2 and hGR, Ile420Leu and Phe478Tyr, locked hGR into dependence on the Gly425Ser substitution for even low-affinity binding to nGREs, a hypothesis supported by sequence alignments (Figure 5.10)[648] and similar to the evolutionary 'ratchet' observed in SR ligand binding domains.[805] The crystal structure of the GR DBD Ser425Gly mutant bound to DNA (Figure 5.11) reveals that position 425 reverts to a glycine-only conformation (Figure 5.9m), confirming that backbone conformation is the likely mechanism for the large-effect seen by the G425S substitution. Even when not bound to DNA, 2D NMR reveals reversal of the Gly425Ser substitution in the hGR DBD results in large conformational changes in residues comprising the DNA-binding interface, consistent with the historical substitution's effects on DNA-mediated inter-protein communication (Figure 5.12). Mutation of residue 425 to a non-glycine amino acid, such as alanine, does not cause a loss of nGRE binding (Figure 5.13). As an aside, the GR Ser425Gly mutation has deleterious effects on the repression of T-bet, AP-1, and NF-κB by GR28-31,[806–809] suggesting DNA-mediated effects play a larger role in pro-inflammatory transcriptional repression by the GR than has been previously assumed.

a

Glucocorticoid Receptor	425
ENSP00000231509_Hsap/1-778	LCLVCSDEASG
ENSACAP00000013512_Acar/1-773	LCLVCSDEASG
ENSAMEP00000010671_Amel/1-784	LCLVCSDEASG
ENSAMXP00000011973_Amex/1-769	ICLVCSDEASG
ENSAMXP00000020637_Amex/1-760	ICLVCSDEASG
ENSAPLP00000006970_Apla/1-773	LCLVCSDEASG
ENSBTAP00000025941_Btau/1-781	LCLVCSDEASG
ENSCAFP00000009468_Cfam/1-780	LCLVCSDEASG
ENSCJAP00000033976_Cjac/1-777	LCLVCSDEASG
ENSCPOP00000014391_Cpor/1-771	LCLVCSDEASG
ENSCSAP00000008542_Csab/1-776	LCLVCSDEASG
ENSDARP00000054263_Drer/1-746	ICLVCSDEASG
ENSDNOP00000008277_Dnov/1-773	LCLVCSDEASG
ENSDORP00000009401_Dord/1-778	LCLVCSDEASG
ENSECAP00000007606_Ecab/1-775	LCLVCSDEASG
ENSEEUP00000008183_Eeur/1-627	LCLVCSDEASG
ENSETEP00000011589_Etel/1-730	LCLVCSDEASG
ENSFALP00000009344_Falb/1-681	LCLVCSDEASG
ENSFCAP00000019629_Fcat/1-453	LCLVCSDEASG
ENSGACP00000024074_Gacu/1-779	ICLVCSDEASG
ENSGACP00000027400_Gacu/1-764	ICLVCSDEASG
ENSGALP00000011948_Ggal/1-772	LCLVCSDEASG
ENSGGOP00000004255_Ggor/1-778	LCLVCSDEASG
ENSGMOP00000006025_Gmor/1-766	ICLVCSDEASG
ENSGMOP00000019143_Gmor/1-698	TCLVCSDEASG
ENSLACP00000017638_Lcha/1-770	ICLVCSDEASG
ENSLAFP00000004197_Lafr/1-776	LCLVCSDEASG
ENSLOCP00000012883_Locu/1-786	ICLVCSDEASG
ENSMEUP00000009009_Meug/1-767	LCLVCSDEASG
ENSMGAP00000009653_Mgal/1-771	LCLVCSDEASG
ENSMICP00000008614_Mmur/1-771	LCLVCSDEASG
ENSMLUP00000010283_Mluc/1-776	LCLVCSDEASG
ENSMMUP00000033941_Mmul/1-777	LCLVCSDEASG
ENSMODP00000012934_Mdom/1-777	LCLVCSDEASG
ENSMPUP00000007823_Mpfu/1-782	LCLVCSDEASG
ENSMUSP00000111229_Mmus/1-792	LCLVCSDEASG
ENSNLEP00000011915_Nleu/1-778	LCLVCSDEASG
ENSOANP00000009152_Oana/1-779	LCLVCSDEASG
ENSOARP00000002152_Oari/1-781	LCLVCSDEASG
ENSOCUP00000010039_Ocun/1-772	LCLVCSDEASG
ENSOGAP00000002995_Ogar/1-775	LCLVCSDEASG
ENSONIP00000010662_Onil/1-803	ICLVCSDEASG
ENSONIP00000022569_Onil/1-633	ICLVCSDEASG
ENSOPRP00000014186_Opri/1-568	LCLVCSDEASG
ENSORLP00000001939_Olat/1-790	ICLVCSDEASG
ENSORLP00000007570_Olat/1-778	ICLVCSDEASG
ENSPANP00000000661_Panu/1-773	LCLVCSDEASG
ENSPFOP00000005862_Pfor/1-765	ICLVCSDEASG
ENSPFOP00000010412_Pfor/1-767	ICLVCSDEASG
ENSPPYP00000017784_Pabe/1-777	LCLVCSDEASG
ENSPSIP00000017259_Psin/1-767	LCLVCSDEASG
ENSPTRP00000029679_Ptro/1-778	LCLVCSDEASG
ENSPVAP00000013434_Pvam/1-775	LCLVCSDEASG
ENSPVAP00000013434_Pvam/1-775	LCLVCSDEASG
ENSRNOP00000044335_Rnor/1-382	LCLVCSDEASG
ENSSARP00000000561_Sara/1-438	LCLVCSDEASG
ENSSHAP00000001942_Shar/1-768	LCLVCSDEASG
ENSSSCP00000015324_Sscr/1-782	LCLVCSDEASG
ENSSTOP00000014793_Itri/1-777	LCLVCSDEASG
ENSTBEP00000014153_Tbel/1-775	LCLVCSDEASG
ENSTGUP00000000147_Tgut/1-773	LCLVCSDEASG
ENSTNIP00000011800_Tnig/1-779	ICVVCSDEASG
ENSTNIP00000021375_Tnig/1-759	ICLVCSDEASG
ENSTRUP00000015645_Trub/1-765	ICLVCSDEASG
ENSTSYP00000012562_Tsyr/1-777	LCLVCSDEASG
ENSTTRP00000003057_Ttru/1-782	LCLVCSDEASG
ENSXETP00000003968_Xtro/1-780	LCLVCSDEASG
ENSXMAP00000001513_Xmac/1-481	ICLVCSDEASG
ENSXMAP00000009196_Xmac/1-767	ICLVCSDEASG

b

Mineralocorticoid Receptor	607
ENSP00000350815_Hsap/1-984	ICLVCGDEASG
ENSACAP00000003608_Acar/1-988	VCLVCGDEASG
ENSAMEP00000002171_Amel/1-978	ICLVCGDEASG
ENSAPLP00000015040_Apla/1-981	VCLVCGDEASG
ENSBTAP00000003291_Btau/1-984	ICLVCGDEASG
ENSCAFP00000011569_Cfam/1-973	ICLVCGDEASG
ENSCHOP00000003343_Chof/1-836	ICLVCGDEASG
ENSCJAP00000004277_Cjac/1-987	ICLVCGDEASG
ENSCPOP00000006998_Cpor/1-980	ICLVCGDEASG
ENSDARP00000130941_Drer/1-970	VCLVCGDEASG
ENSDNOP00000018070_Dnov/1-983	ICLVCGDEASG
ENSDORP00000015711_Dord/1-984	ICLVCGDEASG
ENSECAP00000014151_Ecab/1-984	ICLVCGDEASG
ENSEEUP00000005140_Eeur/1-677	ICLVCGDEASG
ENSETEP00000000543_Etel/1-913	ICLVC-DE-SG
ENSFALP00000005645_Falb/1-987	VCLVCGDEASG
ENSFCAP00000017618_Fcat/1-677	ICLVCGDEASG
ENSGACP00000022713_Gacu/1-1002	VCLVCGDEASG
ENSGALP00000016283_Ggal/1-981	VCLVCGDEASG
ENSGGOP00000001086_Ggor/1-988	ICLVCGDEASG
ENSGMOP00000002015_Gmor/1-1001	VCLVCGDEASG
ENSLAFP00000017164_Lafr/1-983	ICLVCGDEASG
ENSLOCP00000010583_Locu/1-987	VCLVCGDEASG
ENSMEUP00000005763_Meug/1-987	ICLVCGDEASG
ENSMGAP00000003822_Mgal/1-986	VCLVCGDEASG
ENSMLUP00000007404_Mluc/1-982	ICLVCGDEASG
ENSMMUP00000019623_Mmul/1-706	ICLVCGDEASG
ENSMODP00000000987_Mdom/1-993	ICLVCGDEASG
ENSMPUP00000006969_Mpfu/1-982	ICLVCGDEASG
ENSMUSP00000105539_Mmus/1-980	ICLVCGDEASG
ENSNLEP00000006887_Nleu/1-988	ICLVCGDEASG
ENSOANP00000008376_Oana/1-990	VCLVCGDEASG
ENSOCUP00000002724_Ocun/1-982	ICLVCGDEASG
ENSOCUP00000002724_Ocun/1-982	ICLVCGDEASG
ENSOGAP00000012544_Ogar/1-983	ICLVCGDEASG
ENSONIP00000012611_Onil/1-994	VCLVCGDEASG
ENSOPRP00000012148_Opri/1-985	ICLVCGDEASG
ENSORLP00000009438_Olat/1-994	VCLVCGDEASG
ENSPFOP00000009763_Pfor/1-993	VCLVCGDEASG
ENSPPYP00000016883_Pabe/1-981	ICLVCGDEASG
ENSPSIP00000007055_Psin/1-991	VCLVCGDEASG
ENSPTRP00000057800_Ptro/1-984	ICLVCGDEASG
ENSPVAP00000010281_Pvam/1-980	ICLVCGDEASG
ENSRNOP00000045942_Rnor/1-877	ICLVCGDEASG
ENSSARP00000001043_Sara/1-992	ICLVCGDEASG
ENSSTOP00000003652_Itri/1-989	ICLVCGDEASG
ENSTBEP00000011021_Tbel/1-986	ICLVCGDEASG
ENSTGUP00000002663_Tgut/1-972	VCLVCGDEASG
ENSTRUP00000037994_Trub/1-996	VCLVCGDEASG
ENSTSYP00000004880_Tsyr/1-965	ICLVCGDEASG
ENSTTRP00000013694_Ttru/1-983	ICLVCGDEASG
ENSVPAP00000006448_Vpac/1-986	ICLVCGDEASG
ENSXETP00000040897_Xtro/1-980	VCLVCGDEASG
ENSXMAP00000013853_Xmac/1-993	VCLVCGDEASG

Figure 5.10: Alignments of position 425 in GR (a) and its homologous position in MR (b; residue 607). All unique orthologs available from Ensembl are shown.[648]

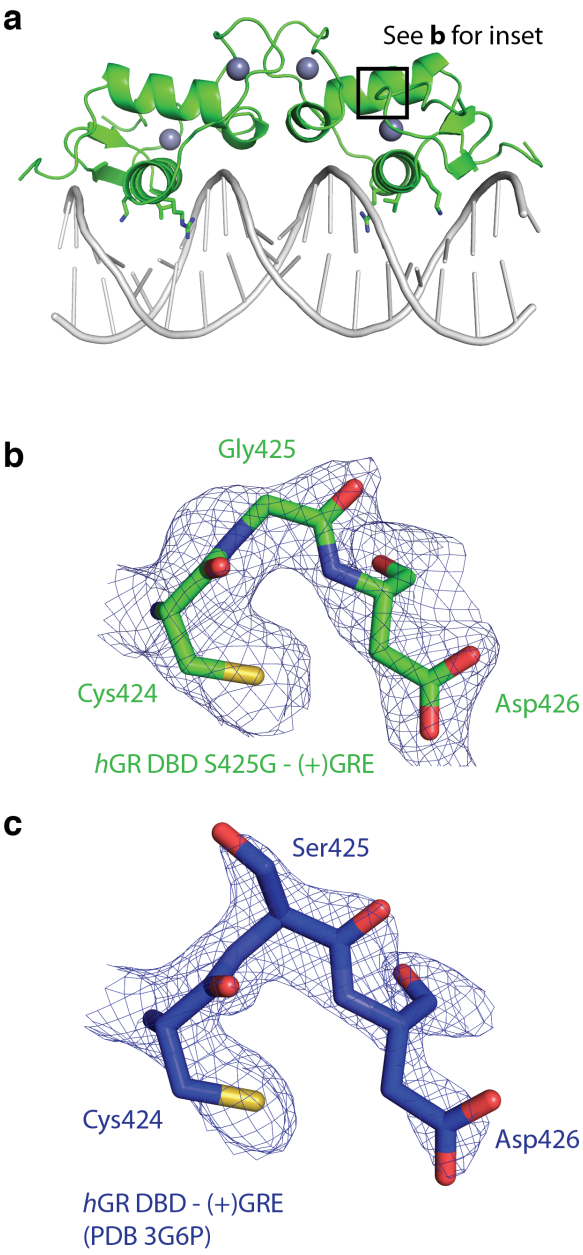


Figure 5.11: Crystal structure of the human GR DBD Ser425Gly mutant. (a) Overall structure of the GR DBD Ser425Gly mutant bound to a (+)GRE. (b) $2F_o$-$F_c$density, contoured to 1 σ, of position 425, confirming the presence of the Ser425Gly mutation. For comparison, density of the WT GR DBD - (+)GRE complex[669] from PDB 3G6P.

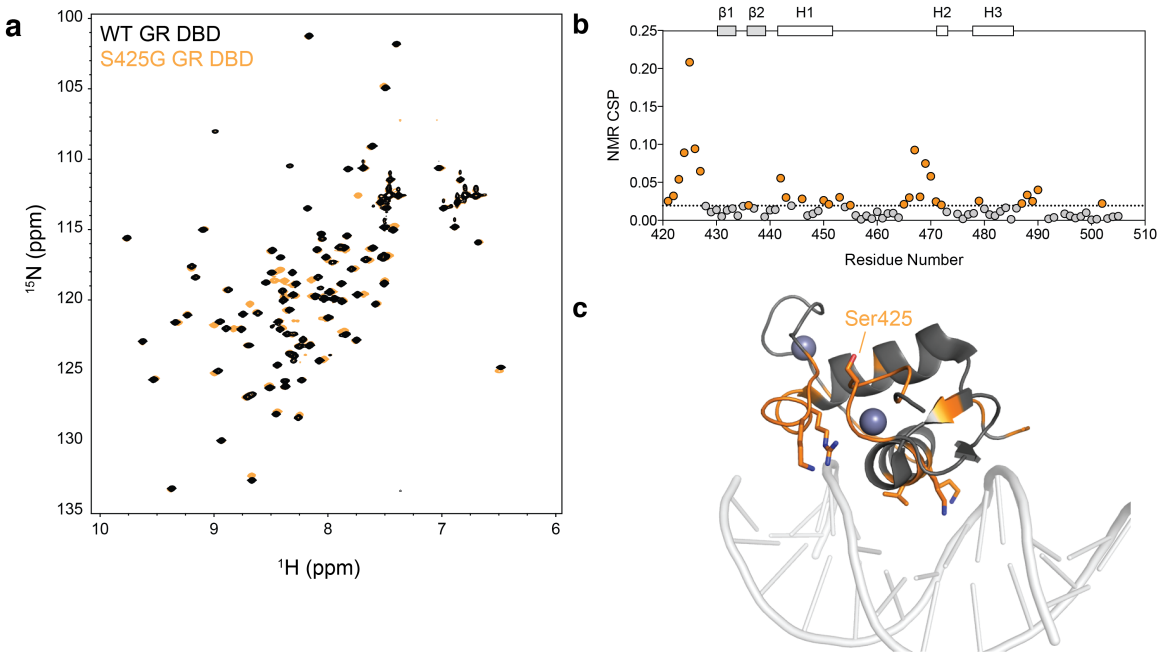


Figure 5.12: Reversal of the Gly425Ser substitution alters the conformation of the GR DBD's DNA binding interface, even in the absence of DNA. (a) 2D HSQC spectra of ^{15}N labeled GR DBD in the absence of DNA; WT protein is in black and Ser425Gly protein in orange. (b) Chemical shift perturbations greater than 2 standard deviations from the mean are plotted as a function of protein residue. (c) Perturbed residues are mapped onto the GR DBD structure in orange; many affected residues are distal from position 425 and DNA-contacting.

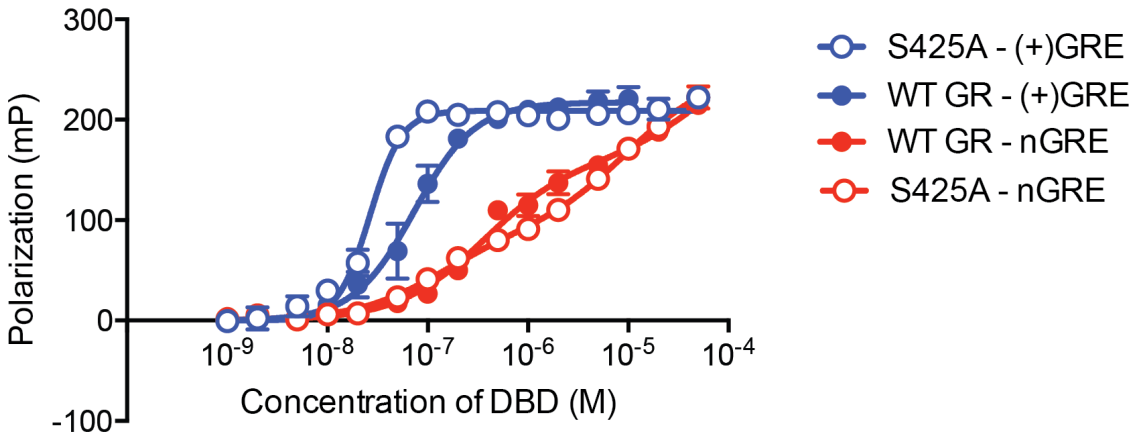


Figure 5.13: Mutation of position 425 to alanine does not affect GR DBD - nGRE binding. Shown are fluorescence polarization values for increasing amounts of WT or S425A proteins into fluoresecently-labeled nGRE or (+)GRE DNA. Calculated high-site affinity (K_d) for WT and S425A on nGREs is 363 and 132 nM, respectively. Error bars represent SEM. Binding studies were conducted in duplicate with three internal replicates.

Finally, the analogous 'forward' substitution in the hMR DBD, Gly607Ser, retains high-affinity binding to (+)GREs (Figure 5.9k) but does not improve affinity for the *TSLP* nGRE (Figure 5.9l), indicating that the accumulation of amino acid substitutions neutral toward function at (+)GREs restricted paths toward the recognition of alternate DNA response elements such as nGREs. A limited difference in sequence background - such as that between the hMR and hGR DBDs - is sufficient to restrict and permit potential future structural conformations, limiting the evolution of new function. However, such functional limitations permit substrate specificity among a related group of proteins, in this case providing an evolutionary path whereby similar transcription factors can control diverse transcriptional programs and effect specific, distinctive responses to signaling input.

5.3 Acknowledgements

We thank Joseph Thornton and Geeta Eick for the ancestral DBD sequences and Christine Youn for assistance with molecular biology. W.H.H. is supported by both an Emory-National Institute of Health Pharmacological Sciences graduate training grant (5T32GM008602) and American Heart Association (AHA) predoctoral fellowship (13PRE16920012). E.A.O. is supported by NIH grant R01DK095750 from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), by an AHA grant (14GRNT20460124) and a W. M. Keck Foundation Medical Research Grant. D.J.K. is supported by NIH grant R01DK101871 from the NIDDK. I.I. is supported by NSF CAREER Award MCB-1149521 from the National Science Foundation. B.K. is supported by a Molecular Basis of Disease fellowship at Georgia State University. Computational resources were provided in part by a National Science Foundation XSEDE allocation CHE110042 and through an allocation at National Energy Research Scientific Computing Center (NERSC) supported by the U.S. De-

partment of Energy Office of Science contract DE-AC02-05CH11231. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38.

5.4 Methods

5.4.1 Protein expression and purification

Proteins were expressed as described previously (Chapters 3,4). Proteins were cloned into pMCSG7 vector, which contains an N-terminal 6X His tag. The expression vector was transformed into BL-21 (DE3) pLySs *E. coli*, which were grown in TB media and induced with 300 μM IPTG at an OD of approximately 0.8 for 4 hours at 30 °C. Cells were lysed via sonication in 20 mM Tris 7.4, 1 M NaCl, 25 mM imidazole, and 5% glycerol. SR DBDs were purified from the supernatant using a nickel affinity chromatography column (HisTrap) with fast protein liquid chromatography (FPLC). Gel filtration via FPLC was used to further purify the DBDs into a buffer of 20 mM Tris 7.4, 100 mM NaCl, and 5% glycerol. Protein was concentrated to approximately 4 mg/ml and flash frozen in liquid N_2 until further use.

5.4.2 Protein - DNA Binding Assays

All DNA oligos for this study were purchased from Integrated DNA Technologies. For DNA binding assays, increasing amounts of indicated protein were added to 10 nM of 5' carboxyfluorescein-labeled DNA oligos and fluorescence polarization values were measured using a Biotek Synergy plate reader. Sequences used for binding assays were: (+)GRE: 5'-(FAM)CCAGAACAGAGTGTTCTGA-3' and 5'-TCAGAACACTCTGTTCTGG-3'; *TSLP* nGRE: 5'-(FAM)CCGCCTCCGGGAGAGCTG and 5' - CAGCTCTCCCGGAGGCGG - 3', where (FAM) indicates the position of the carboxyfluorescein dye. hGR - TSLP nGRE and hMR - (+)GRE binding data

were also reported in Chapters 3 and 4.

5.4.3 Crystallization

All structures were crystallized via hanging drop vapor diffusion. The AncSR2 - *TSLP* nGRE complex was crystallized in 0.1 M HEPES (pH 7.5), 10% PEG 20000, 5% glycerol, and 5% ethanol. The AncGR - (+)GRE complex was crystallized in 0.1 M HEPES (pH 7.5), 12% PEG 20000, and 5% glycerol. The AncGR2 - (+)GRE complex was crystallized in 0.1 M HEPES (pH 7.5) and 15% PEG 8000. The AncMR - (+)GRE complex was crystallized in 0.1 M MES (pH 6.5), 20% PEG 6000, and 5% glycerol. The GR Ser425Gly - (+)GRE complex was crystallized in 0.1 M HEPES (pH 7.5), 20% PEG 8000, and 4% ethylene glycol. Crystals were flash-cooled in liquid N_2 after soaking in cryoprotectant consisting of the crystallization condition plus additional PEG and glycerol. DNA constructs used for crystallization were 5' - CGCCTCCGGGAGAGCT - 3'and 5' - AGCTCTCCCGGAGGCG - 3' for the AncSR2 - *TSLP* nGRE complex, and 5'- TCAGAACACTCTGTTCTG -3' and 5'- CCAGAACAGAGTGTTCTG -3' for the AncGR-, AncGR2- AncMR- and hGR Ser425Gly-(+)GRE complexes.

Data were collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-ID (or 22-BM) beamline at the Advanced Photon Source (APS), Argonne National Laboratory, Argonne, IL, USA. Supporting institutions may be found at www.ser-cat.org/members.html. Data were processed using HKL-2000 and phased with molecular replacement using PHASER in the PHENIX suite and refined using phenix.REFINE.[691,810] COOT was used for model building,[690] and PyMOL (Schrödinger, LLC) and Chimera[811] (Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco) were used for structure visualization and figure generation. The PDBREDO server was used for structure optimization and validation.[812] Sample electron density for all structures is located in Figure 5.14.

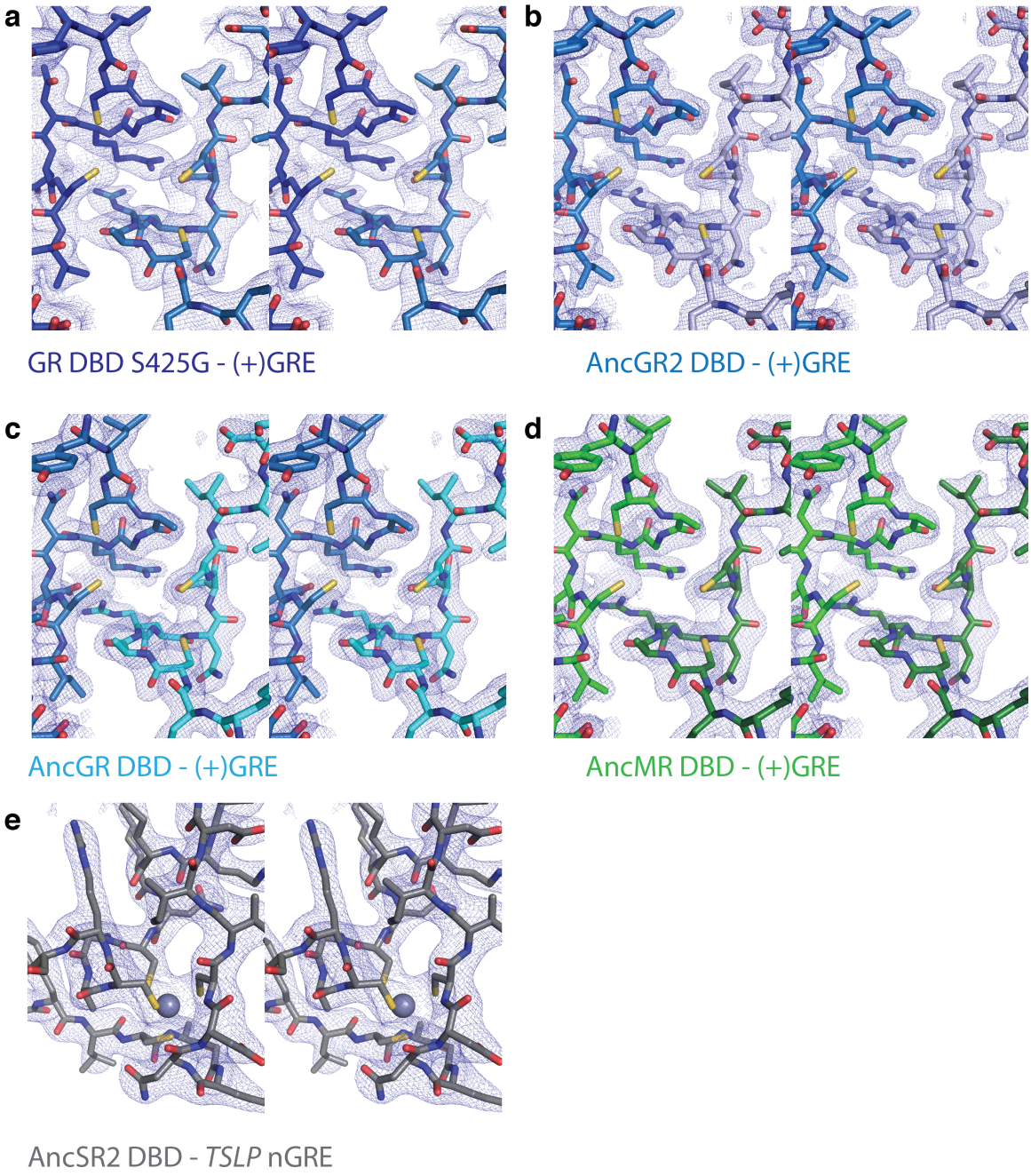


Figure 5.14: Stereo electron density for the crystal structures reported in this chapter. (a-d) Electron density of the dimerization interface of the GR DBD Ser425Gly mutant (a), AncGR2 DBD (b), AncGR DBD (c), AncMR (d) in complex with (+)GRE DNA. Separate protein chains are shown in different shades. In panel (e), sample electron density of the AncsSR2 - *TSLP* nGRE complex is shown.

5.4.4 Molecular Dynamics Simulations

Eight systems were prepared for molecular dynamics (MD) simulation: (+)GRE-AncSR2 (PDB 4OOR), AncGR (PDB 5CBX), AncGR2 (PDB 5CBY), hGR (PDB 3G6R), AncMR2 (PDB 5CBZ), hMR (PDB 4TNT); nGRE- AncSR2 (PDB 5CC0), and hGR (PDB 4HN5). Each protein monomer was capped with acetyl and N-methyl groups at the N- and C- termini, respectively. The complexes were placed in a rectilinear simulation box extending no less than 12 Å from the protein-DNA along each dimension. The systems were solvated with TIP3P[813] water molecules. Na^+ and Cl^- ions were added to maintain neutral charge with excess 0.15 M NaCl concentration to mimic physiological conditions. Systems were set up with the xLeap module of the AmberTools11[814] package with the parm99-bsc0 forcefield.[815,816] Cys-Zn^{2+} tetrahedral geometries in the zinc fingers were maintained using dummy atoms.

All minimization and MD was performed with the NAMD2.9 simulation package[817] using a 2 fs timestep, the r-RESPA method[818] for force integration, and the SHAKE[819] algorithm to restrain bonds all bonds involving hydrogen atoms. Non-bonded interactions were calculated within a 12 Å cutoff, with a switching function applied between 10 Å and 12 Å. Long-range electrostatics were treated with the smooth particle mesh Ewald (PME) scheme,[820] with full electrostatics calculated every 2 steps and 1-4 interaction scaling set at 0.83. Each system was subjected to 10,000 steps of conjugate gradient minimization, followed by a 200 ps MD simulation in the NVT ensemble, with 5 kcal/mol $\times$ $Å^2$ harmonic restraints on all protein and DNA heavy atoms, while smoothly heating from 0 - 300 K. Seven stages of restraint release were then carried out in the NPT ensemble in 1 ns segments, incrementally releasing first protein sidechains, followed by protein backbone, DNA nucleosides, and finally DNA backbone heavy atoms. An unrestrained, 20 ns NPT simulation was performed to fully equilibrate the system. Each system was then sampled for 200 ns in the NPT ensemble. 10,000 evenly spaced frames were taken from each trajectory

for network analysis in the NetworkView[821] plugin of the VMD[822] visualization and analysis program.

We have employed network theory to highlight evolutionary alterations in the allosteric network of 3-keto SR complexes. Our networks are constructed by defining all protein alpha-carbon and DNA C1' atoms in a system as nodes and using Cartesian covariance as a measure of communication within the network. Any pair of nodes that reside within a 4.5 Å cutoff for more than 75% of the MD trajectory are connected via an edge, with the weight of the edge being proportional to the covariance between the nodes. From this raw data, we have resolved the networks into communities, groups of nodes with correlated motions, using the Girvan-Newman algorithm.[823] While constructing the community graphs, we generate the minimum number of communities possible while maintaining at least 92.3% of the maximum modularity[823] to prevent excessive community subdivision. We have quantified the magnitude of communication flow between communities by the total betweenness[824] of all edges that transition between communities. In the case of the nGRE complexes, we have also identified suboptimal paths between monomers and through the DNA using the Floyd-Warshall algorithm.[825]

Covariance plots for each of the MD simulations are found in Figure 5.15.

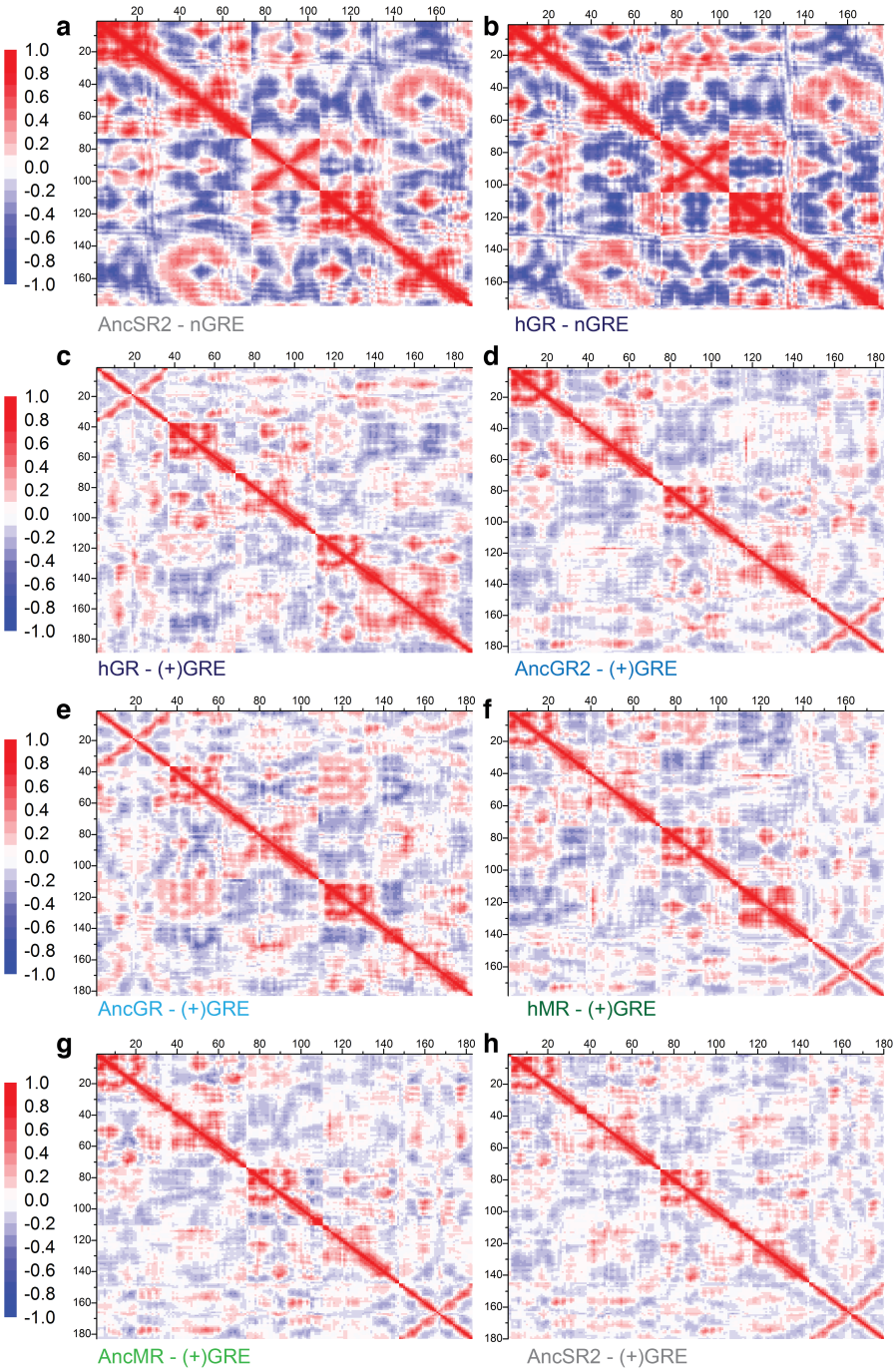


Figure 5.15: Covariance plots for the MD simulations performed in this chapter.

5.4.5 Cellular Activation and Repression Assays

HeLa cells were transfected with 10 ng of the indicated receptor, 50 ng of the indicated firefly luciferase reporter, and 10 ng of Renilla luciferase under the control of the pRLTK (constitutively active) promoter. For transfection, OptiMEM media (without antibiotics) was used with FuGene HD, according to the manufacturer's protocols. Otherwise, HeLa cells were passaged in MEMα (Life Technologies) supplemented with 10% stripped FBS (Atlanta Biologicals) and penicillin/streptomycin (Life Technologies). 24 hours after transfection, cells were treated with 1 μM dexamethasone, except for hMR conditions, in which aldosterone was used. 24 hours after ligand treatment, firefly and Renilla luciferase activities were measured using the DualGlo kit (Promega) according to the manufacturer's protocol on a Biotek Synergy plate reader.

5.4.6 NMR

^{15}N-labeled GR DBD and S425G were expressed in *E. coli* BL21 (DE3) pLysS cells as TEV-cleavable 6X-His tagged fusion proteins using M9 media, with $^{15}NH_4Cl$ as the sole nitrogen source. Proteins were purified through a Ni-NTA column against a 500 mM imidazole gradient, using wash buffer containing 50 mM Tris (pH 8.0), 100 mM NaCl, 15 mM imidazole, 1 mM TCEP. The His tag was subsequently cleaved with TEV protease overnight at 4 °C. The protein solutions were passed through the Ni NTA column anew, and the flow through containing purified protein was collected. Proteins were verified to be >99% pure by SDS-PAGE.

NMR data were collected on a Bruker 700 MHz Bruker NMR instrument equipped with a QCI cryoprobe. For DNA experiments, the 19 nt (+)GRE and TSLP nGRE DNA duplexes were reconstituted in 20 mM phosphate (pH 6.7), 100 mM NaCl, 1 mM TCEP, 10% D_2O buffer to 1.7-2.0 mM, subsequently annealed by denaturing at 95 °C for 3 minutes and equilibrated to room temperature (20-23 °C) overnight. For

protein experiments, 2D [^{1}H,^{15}N]-HSQC spectra were collected at 25 °C for 250 uM of free ^{15}N-labelled GR DBD protein and complexed with 1.5:1 of GRE or 0.44:1/2.1:1 of TSLP nGRE DNA duplex in the same NMR buffer. Chemical shift perturbations were assigned using previously published GR DBD NMR chemical shifts[672] and calculated using the minimum chemical shift perturbation procedure. Data were processed using Bruker Topsin and analysed with NMRViewJ (OneMoon Scientific, Inc).

Chapter 6

Beyond nGREs: GR-mediated, DNA-dependent repression of NF-κB

6.1 Abstract

The glucocorticoid receptor (GR) is a ubiquitously expressed nuclear receptor that controls immunity, metabolism, and responses to stress. GR agonists, glucocorticoids (GCs), are potent repressors of NF-κB activity, making them a preferred choice for treatment of inflammation-driven conditions. GCs repress 25% of LPS-activated genes, yet current models are inadequate to explain GR's role within this important signaling pathway. In contrast to the current tethering hypothesis, which posits that GR interacts with NF-κB in a DNA-independent manner, we support a model whereby DNA binding is required for GR-driven repression of NF-κB. In this chapter, we report the crystal structure of the GR DNA binding domain (DBD) in complex with five NF-κB response elements (κBREs) and demonstrate that GR recognizes cryptic sequences between the binding sites of each NF-κB subunit in order to bind to and repress the promoters of pro-inflammatory genes. These cryptic sequences exhibit high sequence and functional conservation, despite a lack of NF-κB sequence-specific contacts, suggesting that GR binding to κBREs is an evolutionarily important mechanism of controlling immunity.

This chapter is adapted from a manuscript in preparation for publication:
Hudson WH, Weikum ER, Nwachukwu JC, de Vera IMS, Herbst AG, Kojetin DJ, Nettles KW, Ortlund EA. Cryptic glucocorticoid receptor binding sites pervade genomic NF-κB response elements.

6.2 Introduction

Glucocorticoids (GCs) are a class of steroid hormones that are widely prescribed for inflammation-driven conditions such as asthma and arthritis.[826] GCs exert their effects by binding to the glucocorticoid receptor (GR), a ubiquitously expressed nuclear receptor that drives both the activation and repression of its target genes. Ligand-bound GR is able to antagonize the activity of immunogenic transcription factors such as NF-κB, AP-1, and T-bet,[575,576,793,827] resulting in a potent attenuation of inflammation. Unfortunately, GR's anti-inflammatory actions are concomitant with induction of a host of undesirable side effects that include skin atrophy, glaucoma, osteoporosis and hypertension.[709] These opposing actions by GR have led to an intense - and largely unsuccessful - search for dissociated glucocorticoids that would separate the receptor's anti-inflammatory properties from its more malicious effects.[566]

Upon binding of GCs to GR's ligand binding domain (LBD), the receptor translocates to the nucleus, where its agonist-bound conformation enables the recruitment of transcriptional co-activators such as TIF2 and SRC-1.[828] The receptor's DNA binding domain (DBD) can bind activating glucocorticoid response elements, or (+)GREs, which are inverted repeat sequences containing two AGAACA (or similar) half sites separated by 3 base pairs.[334] Thus, transcriptional activation mediated by GR appears to be largely dependent on DNA binding by the receptor, with additional allosteric modulation potentially occurring through slight variation of the (+)GRE sequence.[669,672] In contrast, GR's repressive effects on pro-inflammatory transcription factors are generally thought to be DNA-independent. While GR can upregulate the transcription of IκB - an inhibitor of NF-κB - through a DNA-dependent mechanism,[829] this activity is not sufficient for repression of NF-κB.[806] Instead, GR is believed to tether, or interact directly, with DNA-bound NF-κB in a DNA-independent manner.[830]

Recently, some GR-mediated transcriptional repression has been attributed to

direct interactions of the receptor with DNA. In 2011, the discovery of a negative glucocorticoid receptor element (nGRE) DNA sequence was found to mediate glucocorticoid-induced repression of hundreds of genes.[574] Subsequent crystallographic analyses demonstrated that the GR DBD binds nGREs in a distinct orientation from (+)GREs (see Chapter 3). Simultaneously, ChIP-seq studies examining GR's role in attenuating the immune response have shown GR is recruited to some pro-inflammatory genes such as *CCL2* in the absence of pro-inflammatory signaling, indicating that the tethering model is insufficient to explain GR's recruitment to and repression of NF-κB target genes.[831] Here, we propose that GR in fact binds directly to NF-κB response elements in order to repress pro-inflammatory genes.

6.3 Results

6.3.1 Classic GR mutations dissociate (+)GRE and nGRE binding

The glucocorticoid receptor is one of five paralogous steroid receptors in humans. The other four steroid receptors include the mineralocorticoid, androgen, progesterone, and estrogen receptors. Given the critical physiological roles played by these receptors and their similar protein structure,[832] we sought to determine whether these proteins could mediate repression from nGREs (see also Chapter 5). Interestingly, while the DBD of the mineralocorticoid receptor (MR) - GR's closest paralog - can bind with high affinity to (+)GREs, the protein was unable to bind to an nGRE from the *TSLP* promoter (Figure 6.1a). Additionally, the full-length MR protein was unable to repress a constitutively active nGRE-containing promoter in cells, unlike GR (Figure 6.1b). Given the high sequence identity between the DBDs of MR and GR - 69 of 75 identical residues (Figure 5.1) - we hypothesized that reverting one or more of the divergent residues in the GR DBD could dissociate repressive and activating DNA

binding by GR. Indeed, a single amino acid substitution in GR to the homologous residue in MR (Ser425Gly) was able to decrease GR's affinity for nGREs by two orders of magnitude, while preserving dimeric (+)GRE binding (Figure 6.1c,d). Critically, this mutant also diminished the ability of full-length GR to repress from nGRE-containing elements in cells (Figure 6.1e; see also Figure 5.9).

Consistent with our data, previous studies have shown that the Ser425Gly GR mutant is fully capable of activating transcription; however, at least five previous studies have shown that the Ser425Gly mutant is incapable of repressing the pro-inflammatory transcription factors AP-1 and NF-κB.[575,576,793,807,808] Given the relatively recent identification of the nGRE motif,[574] the inability of the Ser425Gly mutant to repress transcription led to the hypothesis that GR's DNA binding function was dispensable for repression of NF-κB. Interestingly, two other GR mutants[759,833] that repress NF-κB but inhibit (+)GRE-mediated repression appear to selectively impair (+)GRE binding more than nGRE binding *in vitro* (Figure 6.2). These observations led us to hypothesize that GR might bind directly to DNA in order to repress NF-κB activity.

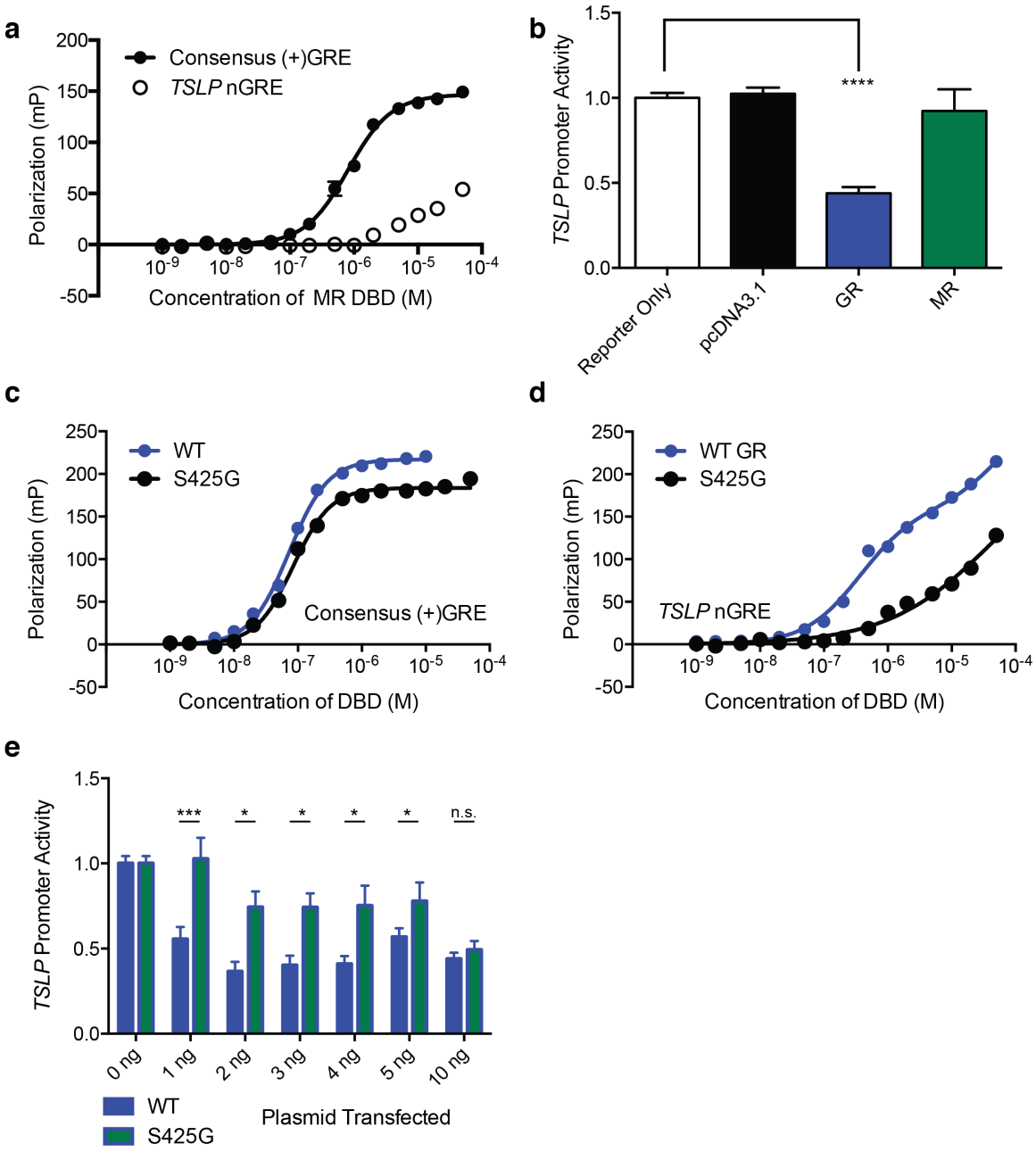


Figure 6.1: A single mutation differentiates the repressive function of two homologous receptors. (a) Recombinantly-expressed human mineralocorticoid receptor (MR) DNA binding domain (DBD) binds a consensus (+)GRE, but not repressive nGRE elements. (b) Full-length GR (10 ng) is able to repress a constitutively active *TSLP* reporter in HeLa cells, in contrast to full length MR. This difference can be traced to a single amino acid mutation, Ser425Gly, which does not affect GR DBD binding to a (+)GRE (c), but diminishes binding to the *TSLP* nGRE (d). (e) The full-length Ser425Gly protein is a less potent repressor of a constitutively-active nGRE-driven promoter in HeLa cells than WT GR. In panel (c), **** indicates $P < 0.0001$ after one-way ANOVA followed by Tukey's multiple comparison test. In panel (e), all WT GR transfections differ from 0 ng by at least $P < 0.05$. ***, *, and n.s. indicate $P < 0.001$, < 0.05, and not significant between the groups shown by two-way ANOVA followed by Sidak's multiple comparison test. See also Figure 5.9.

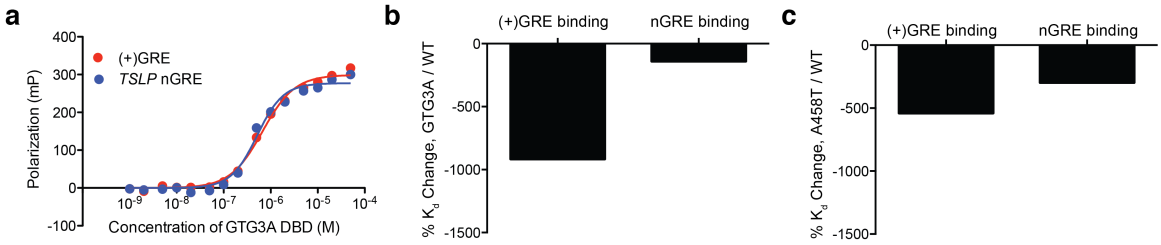


Figure 6.2: Two GR DBD mutants selectively impair (+)GRE binding vs. nGRE binding. (a) Binding of the GTG3A mutant[759] to the *TSLP* nGRE and a consensus (+)GRE reveals similar binding affinities for the two elements. However, this represents a large decrease in (+)GRE binding vs. nGRE binding, due to WT GR's lower affinity for nGREs (panel b). (c) The GRdim (Ala458Thr) mutation also affects (+)GRE binding more than nGRE binding.

6.3.2 GR binds directly to NF-κB response elements

In order to test this hypothesis, we investigated the ability of a GR double mutant, Lys442Ala Arg447Ala, to repress constitutively active reporters containing NF-κB response elements. This mutant lacks two key side chains critical for sequence-specific DNA recognition by GR at multiple response elements (reference 669 and Chapter 3). While WT GR was able to repress several of these reporters, including *IL8*, *CCL2*, *RELB*, *PLAU*, and *ICAM1*, the Lys442Ala Arg447Ala mutant was generally unable to repress more than transfection with an empty vector (Figure 6.3). These results indicated that GR's DNA binding ability is critical for its ability to repress transcription, and we sought to determine the DNA sequences through which GR mediated these effects.

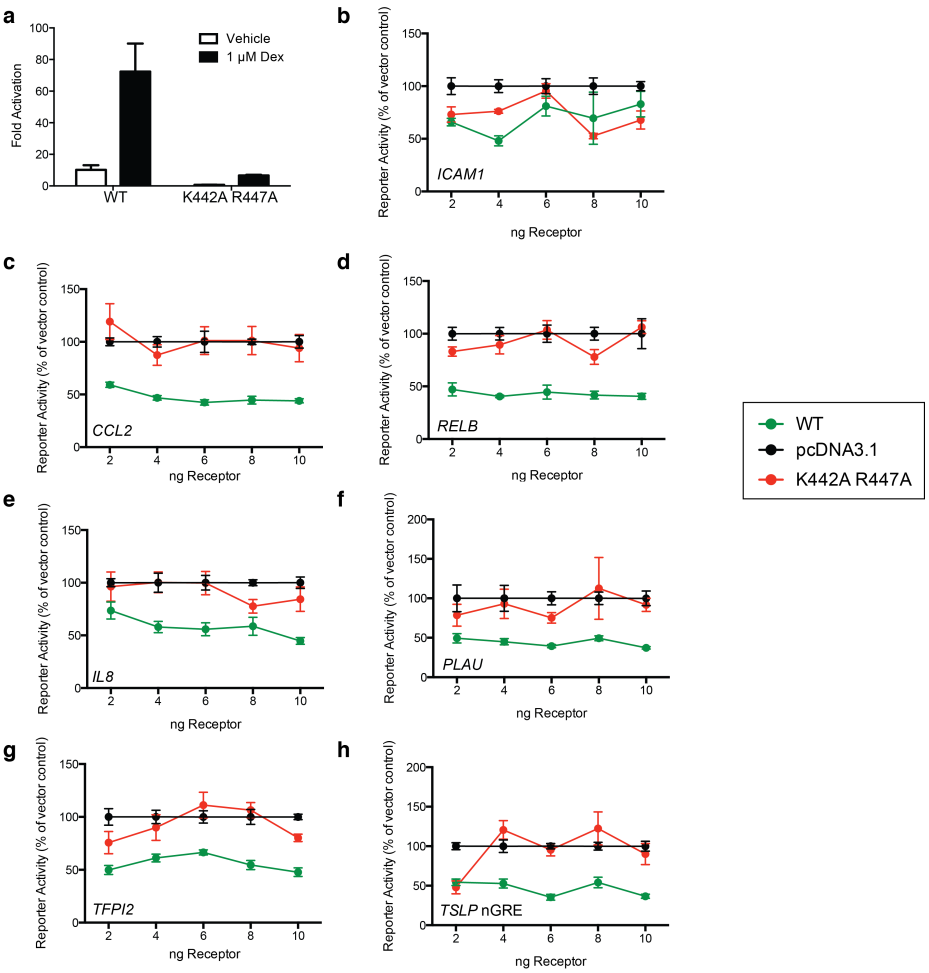


Figure 6.3: The Lys442Ala Arg447Ala mutant is unable to repress constitutively active reporters containing NF-κB response elements. (a) The Lys442Ala Arg447Ala mutant is greatly deficient in activation of the (+)GRE-containing *SGK* promoter in U2OS cells. (b-g) Increasing amounts of plasmid containing WT full-length GR, Lys442Ala Arg447Ala full-length GR, or empty pcDNA3.1 were transfected into U2OS cells containing the indicated NF-κB response element constitutively activated by the SV40 promoter and enhancer. (h) The Lys442Ala Arg447Ala mutant was also tested with the *TSLP* nGRE, which requires GR-DNA interaction for repression,[574] as a control.

Several ChIP-seq studies have been performed to determine motifs enriched at GR binding sites. In many of these studies, NF-κB response elements are highly enriched at GR ChIP-seq peaks.[677,831,834] Most of these studies were performed in the absence of NF-κB activation, potentially implying that GR was not recruited to these elements via tethering. We examined ENCODE GR ChIP-seq data performed in A549 cells with ethanol or dexamethasone treatment.[565] While unliganded GR is not present in at the *CCL2*, *PLAU*, *RELB*, or *IL8* NF-κB response elements in A549 cells, dexamethasone treatment alone is sufficient to recruit GR to these elements (Figure 6.4a-d); similar results have been observed on the *IL6* promoter in macrophages.[831] This suggests that GR could bind directly to NF-κB response elements without the need for tethering to another transcription factor. In MCF-7 cells, we observed that GR is present at the *IL6* promoter, which contains an NF-κB response element, in the absence of ligand (Figure 6.4e). In order to identify whether potential tethering factors are present basally, we performed ChIP for p65 as well as AP-1 family members, and compared the association relative to IgG control. While we detected increased association of GR, along with Fra-1 and Fra-2 relative to IgG control immunoprecipitation, there was less recruitment than IgG for ERα, p65, c-Jun, JunB, c-Fos, and FosB, suggesting that these were not potential tethering factors (Figure 6.4e). After TNF-α treatment, these proteins were present at the *IL6* promoter, indicating that the antibodies were effective for these factors. (Figure 6.4f) Together, these results suggested that GR might bind directly to NF-κB response elements to repress pro-inflammatory genes.

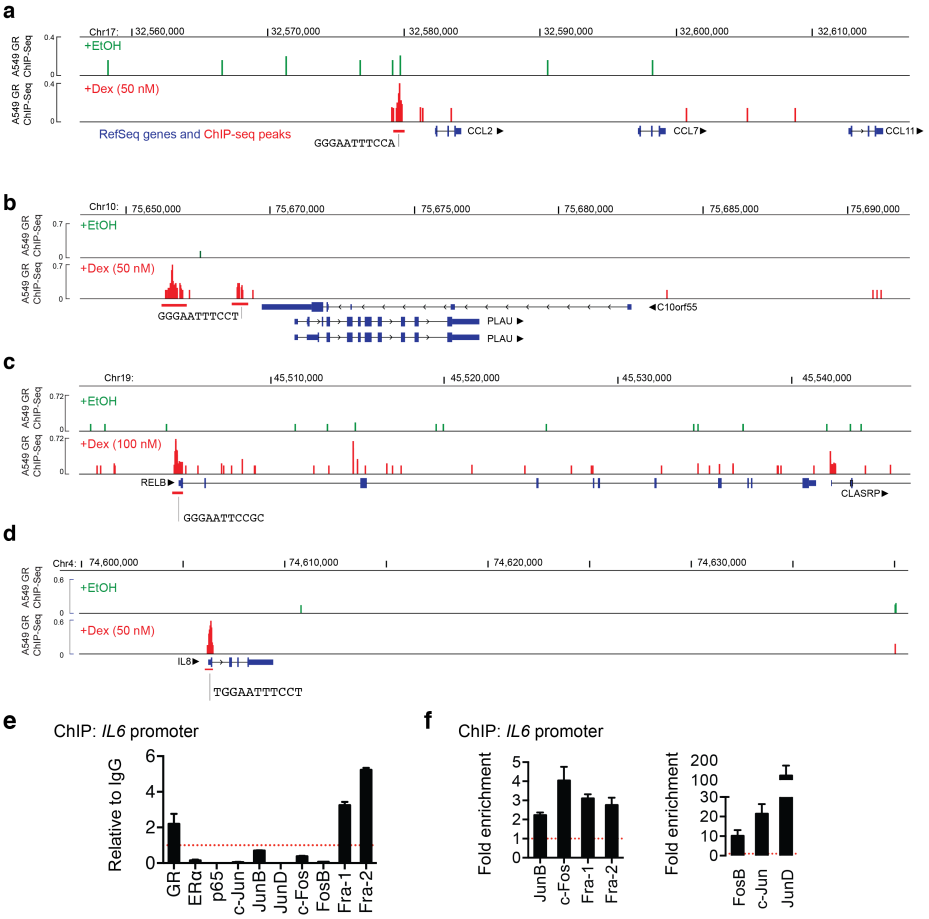
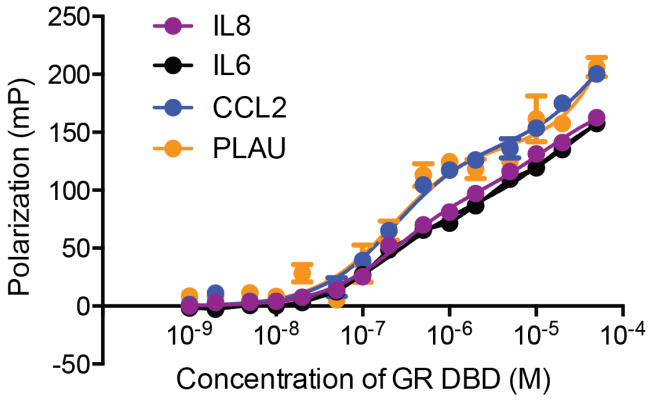


Figure 6.4: GR is recruited directly to NF-κB response elements. (a-d) ENCODE ChIP-seq data reveals that GR is recruited directly to the *CCL2*, *PLAU*, *RELB*, and *IL8* NF-κB response elements in the presence of dexamethasone alone. (e) ChIP at the *IL6* promoter reveals that GR is basally present, even without the presence of potential tethering factors such as p65 and AP-1 subunits. (f) ChIP at the *IL6* promoter with TNF-α stimulation reveals the presence of AP-1 subunits, validating the antibodies used in (e).

To test this hypothesis, we used fluorescence polarization to test the ability of the GR DBD to bind to the *IL6*, *IL8*, *CCL2*, and *PLAU* NF-κB response elements *in vitro*. GR DBD was able to bind to all four of these response elements, with each binding event fitting well to a two-site model (Figure 6.5). Interestingly, we previously showed that binding of GR to nGREs occurs in a similar fashion; affinity of the GR DBD for these NF-κB response elements closely mirrored the affinity of the DBD for nGREs (see Figure 3.1 and Table 3.1).

To uncover the mechanism by which GR binds to NF-κB response elements, we solved five crystal structures of GR DBD bound to NF-κB response elements from the *CCL2*, *ICAM1*, *IL8*, *PLAU*, and *RELB* promoters, at resolutions ranging from 1.85 to 2.30 Å (Table 6.1, Figure 6.6a-e). In all of these crystal structures, the GR DBD formed a dimer within the asymmetric unit (Figure 6.6a-e). However, in all structures, one DBD monomer was consistently located above the end-stacking junction of the pseudo-continuous DNA helix formed by crystal packing. Therefore, it is likely that monomeric GR DBD is bound to these NF-κB response elements at all but the highest concentrations.

a



b

Response element	High site K_d	Low site K_d	*p* for 2-site binding
IL8	239 nM	17.0 μM	> 0.0001
IL6	221 nM	68.5 μM	> 0.0001
CCL2	233 nM	67.9 μM	> 0.0001
PLAU	215 nM	>50 μM	0.0051

Figure 6.5: GR DBD binds to NF-κB response elements *in vitro*. (a) Fluorescence polarization was used to monitor binding of GR DBD to four NF-κB response elements. (b) These binding events showed two-site binding curves, similar to that of GR DBD when binding to an nGRE (Figure 3.1 and Table 3.1). An extra sum-of-squares F-test was used to compare a two-site specific binding event to a one-site specific binding event using GraphPad Prism; the resulting P values are shown in the right column of the panel.

	GR DBD - *IL8* κBRE	GR DBD - *ICAM1* κBRE	GR DBD - *PLAU* κBRE	GR DBD - *RELB* κBRE	GR DBD - *CCL2* κBRE
Data collection					
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions					
a, b, c (Å)	39.2, 97.2, 103.3	39.1,97.1, 104.0	39.9, 95.4, 103.5	39.2, 97.0, 103.3	39.1, 97.4, 103.9
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	34.27 - 1.85 (1.92-1.85)	50.00 - 2.17 (2.25-2.17)	50.00 - 2.15 (2.23-2.15)	50.00 - 2.25 (2.33-2.25)	48.70 - 2.20 (2.28-2.20)
R_{merge}	10.9	6.9	19.1	9.8	9.5
I / σI	14.0 (2.8)	20.0 (1.8)	16.2 (1.7)	19.8 (1.9)	20.9 (2.1)
Completeness (%)	88.4 (74.7)	96.6 (91.6)	100.0 (100.0)	98.1 (84.9)	98.3 (87.8)
Redundancy	4.4 (2.3)	5.1 (4.3)	7.1 (6.7)	6.3 (2.9)	6.7 (5.1)
Refinement					
Resolution (Å)	1.85	2.40	2.20	2.25	2.20
No. reflections	30412	15785	20243	19106	20451
R_{work} / R_{free}	22.0 / 24.2	25.7/28.4	23.6/26.6	20.4/22.8	23.3/27.1
No. atoms					
Protein/DNA	1774	1767	1774	1777	1808
Water	96	19	51	75	29
B-factors					
Protein/DNA	51.5	97.2	50.7	51.3	58.0
Water	50.0	61.1	48.6	46.2	46.2
RMS deviations					
Bond lengths (Å)	0.005	0.008	0.005	0.004	0.011
Bond angles (°)	0.5	1.17	0.85	0.72	1.18
Ramachandran					
Favored (%)	98	92	96	96	94
Outliers (%)	0	1	0	0	0

Table 6.1: Data collection and refinement statistics (molecular replacement) for x-ray crystal structures reported in this chapter.

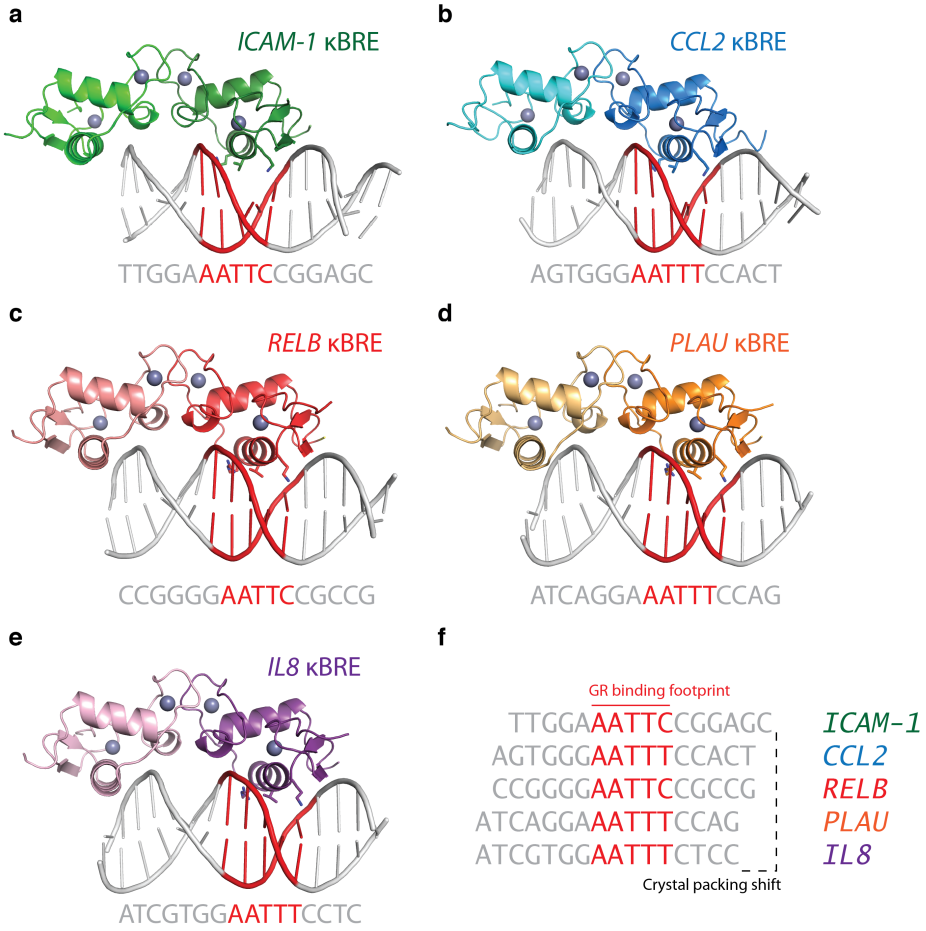


Figure 6.6: Five crystal structures of the GR DBD bound to NF-κB response elements reveal its binding footprint. (a-e) Overall representations of five GR DBD - NF-κB response element crystal structures. In each case, the GR DBD crystallized as a dimer in the asymmetric unit, with one monomer positioned over a DNA end stacking junction. The DNA sequence is shown below each structure, with the GR binding footprint in red. (f) The DNA binding footprints of GR are aligned, showing that crystal packing shifts in order to accommodate binding of GR to a specific AATTY sequence (Y, pyrimidine).

6.3.3 GR recognizes NF-κB response elements in a sequence specific manner

An interesting feature of the five crystal structures in complex with NF-κB response elements is the common binding footprint for the central GR DBD monomer (Figure 6.6f). In each structure, GR DBD recognizes an AATTY sequence, where Y represents a pyrimidine base. In fact, while each of the five structures were solved with a 16 bp oligonucleotide, the DNA packs in three distinct conformations to ensure that GR specifically recognizes the AATTY sequence (Figure 6.6f). PISA analysis[722] indicates that the free-energy gain upon formation of the specifically-bound GR DBD monomer and the *IL8* NF-κB response element is a very favorable -8.7 kcal/mol, similar to the change seen upon nGRE binding (see Chapter 3.3.3).

A close examination of the DBD - DNA interface shows that GR recognizes the AATTY sequence in a specific manner (Figure 6.7). Through the side chains of Lys442, Val443, and Arg447, the GR DBD makes contacts with four of the five bases within the AATTY motif. Arg447 recognizes the first two bases of this sequence: its guanidino group makes van der Waals contacts with the first adenine and a terminal amine forms a hydrogen bond with the second adenine in the motif (Figure 6.7a-e). Val443 participates in a van der Waals interaction with the C7 of the thymine in the central A:T base pair; this distance is constant across the five crystal structures (3.8-4.0 Å; Figure 6.7a-e). Finally, Lys442 makes a moderately strong, electrostatic hydrogen bond with the purine residue opposite the final base in the AATTY motif (Fig 4a-e). These interactions are similar to GR's contacts at nGREs, with the exception of the positioning of Arg447 (Figure 6.7f).

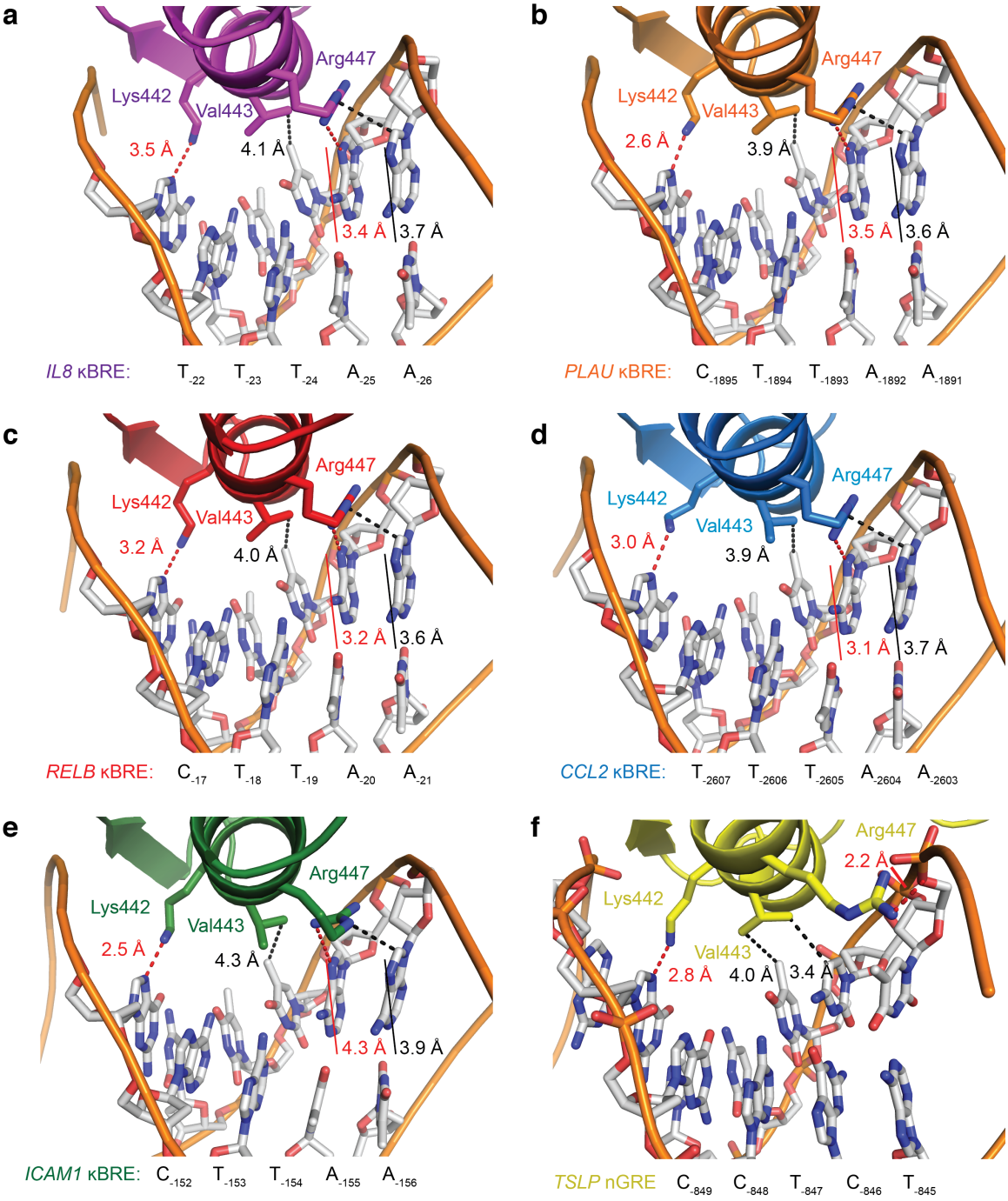


Figure 6.7: The GR DBD makes sequence-specific contacts with NF-κB response elements. (a-e) Sequence specific contacts between the GR DBD and the *CCL2*, *IL8*, *PLAU*, *RELB*, and *ICAM1* NF-κB response elements. For comparison, sequence specific contacts between the GR DBD and the *TSLP* nGRE are shown in panel (f).

To confirm these observations, we performed NMR footprinting analysis to map the interaction between GR DBD and the *IL8* NF-κB DNA response element. Two dimensional (2D) homonuclear [^{1}H,^{1}H]-NOESY NMR analysis reveals the GR DBD binding footprint on the *IL8* NF-κB DNA observed is consistent with the crystal structure (Figure 6.8a-c). Furthermore, the NMR data suggest the nucleotides near the AATTY sequence are most perturbed (Figure 6.8c). The largest chemical shift perturbation occurred at guanine$_{-21}$, which is directly adjacent to the first adenine of the AATTY motif. This is in strong agreement with the crystal structure of the GR DBD - *IL8* NF-κB response element structure, as the GR DBD makes two close interactions (2.8 Å) with the DNA backbone at this position (Figure 6.8f). The next two largest chemical shift perturbations upon GR DBD binding occurred in the AATTY motif, including adenine$_{-24}$, which is directly contacted by Lys442 of the GR DBD. Crucially, thymine$_{-24}$, which is contacted directly by Val443 of the GR DBD, also had a significant chemical shift perturbation upon GR binding.

On the protein, 2D [^{1}H,^{15}N]-HSQC NMR analysis reveals that binding of *IL8* NF-κB DNA to ^{15}N-GR DBD causes large chemical shift perturbations for residues that contact DNA, such as Cys441 and Val488 (Figure 6.8d). When GR binds to (+)GRE DNA as a homodimer, this causes large chemical shift perturbations in the dimerization loop (or D-loop) residues, such as Ala458 and Gly459 (Figure 6.8e).[672] However, when bound to the *IL8* NF-κB response element, these D-loop residues were not affected (Figure 6.8d), confirming that when GR binds to this DNA sequence it does not bind as a D-loop engaged dimer, but instead likely binds as a monomer. This is consistent with recent reports that a monomeric full-length GR protein carrying mutations at both the DBD and LBD dimerization interfaces remains capable of repressing NF-κB activity.[682]

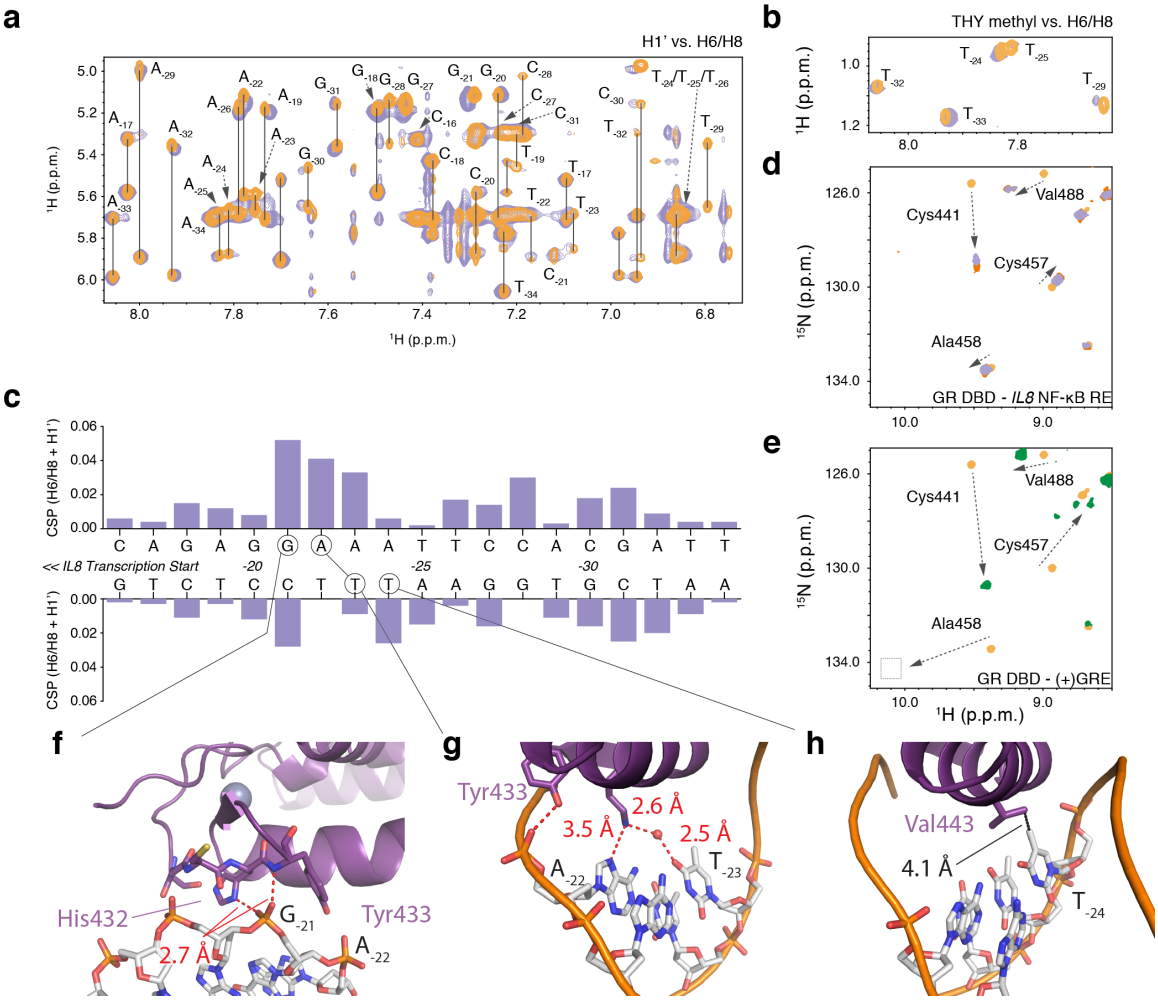


Figure 6.8: NMR reveals an interaction footprint of the GR DBD - *IL8* NF-κB response element complex that is consistent with crystal structures. (a,b) 2D homonuclear [^{1}H,^{1}H]-NOESY NMR data for *IL8* NF-κB response element alone (orange) or bound to GR DBD (2.3:1 molar ratio, purple). (c) NMR chemical shift perturbation analysis for data shown in (a). Panels (f-h) show insets of highly perturbed bases. (d) 2D [^{1}H,^{15}N]-HSQC NMR analysis of ^{15}N-labeled GR DBD (orange), and the same bound to *IL8* NF-κB response element (0.44:1 molar ratio, dark orange; 2.3:1 molar ratio, purple). (e) 2D [^{1}H,^{15}N]-HSQC NMR analysis of ^{15}N GR DBD (orange), and the same bound to (+)GRE consensus DNA response element. (f) Base G_{-21}, which makes two close contacts with the GR DBD, is highly perturbed upon formation of the complex. The same is true of bases A_{-22} (panel g) and T_{-24} (panel h).

6.3.4 Cryptic GR binding sites within NF-κB response elements are highly conserved

NF-κB is a homo- or heterodimer of Rel homology domain-containing proteins. NF-κB binds to its response elements by specifically recognizing two binding footprints surrounding a central sequence (Figure 6.9a). In a p50/p65 heterodimer, the central base pairs of the NF-κB response elements are not specifically bound NF-κB itself (Figure 6.9b).[835] Despite this lack of sequence discrimination, an AATTY motif is overrepresented in the NF-κB response element spacer region.[836] No satisfactory explanation for this overrepresentation has been explained by structural analyses of NF-κB binding alone,[836] and a recent SELEX study to determine the optimal NF-κB binding motif revealed little sequence preference by the protein at this spacer sequence.[837] We propose that the AATTY motif is widely prevalent not due to any structural requirement for NF-κB binding, but to ensure that these elements can be bound and repressed by the glucocorticoid receptor. Indeed, this AATTY motif is present in many NF-κB-responsive genes that are regulated by dexamethasone (Table 6.2).

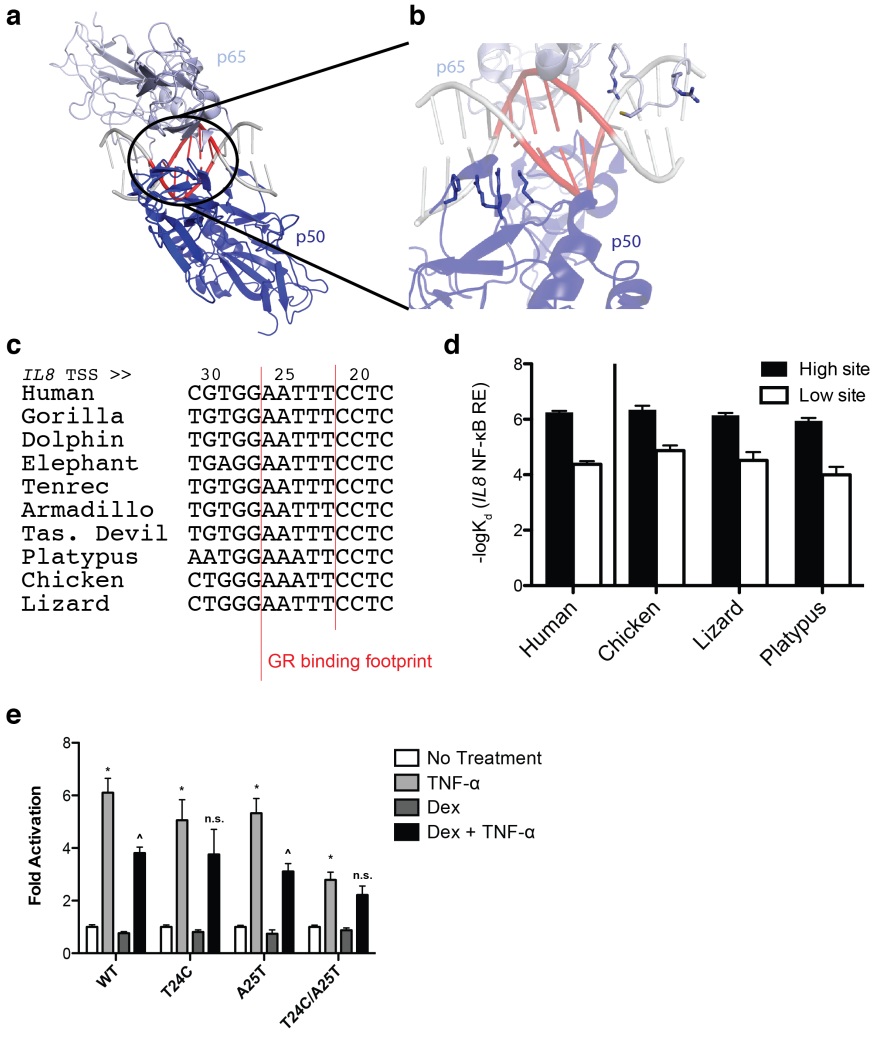


Figure 6.9: The GR DBD binds to a cryptic sequence in the spacer NF-κB response elements. (a) Crystal structure of the p50/p65 NF-κB heterodimer bound to its cognate response element in the IFNβ promoter.[838] The AATTY sequence at this response element is shown in red (panel b). At this element (and many of its other response elements), NF-κB largely makes sequence-specific contacts with the regions flanking the AATTY motif. Despite the lack of sequence-specific contacts by NF-κB at this region, the AATTY motif is highly conserved; the conservation of the AATTY motif within the *IL8* NF-κB response element is shown in panel (c). (d) Additionally, the GR DBD can bind to the *IL8* NF-κB response element from divergent species with nearly-identical affinities. (e) Finally, mutation of central, conserved bases within the AATTY motif of the *IL8* promoter affects the ability of dexamethasone to repress transcription from this response element in HeLa cells. *, $p < 0.001$ of TNF-α vs. no treatment; $\wedge$, $p < 0.001$ of TNF-α vs. TNF-α + Dex; n.s., no significance between TNF-α vs. TNF-α + Dex. Statistics are two-way ANOVA with Tukey's post-hoc test.

Gene	NF-κB response element sequence	Regulation by glucocorticoids	References
CARD15	GGG**AATTT**CC	Up	839, 840
CCL2	GGG**AATTT**CC	Down	834, 841
CCL5	AGA**AATTT**TTCC	Down	842, 843
CD40	GGG**AATTT**CC	Down	844, 845
CD40L	GG**AATTT**TCC	Up	846, 847
CR2	GGG**AATTC**TCT	Up	848, 849
CXCL5	GGG**AATTT**CCC	Down	850, 851
ICAM1	GGA**AATTC**C	Down	852
IER3	CGG**AATTT**CC	Down	834, 853
IFNB1	GGGA**AATTC**C	Down	854, 855
IL8	GTGG**AATTT**CC	Down	856
NFKB2	GGG**AATTC**CC	Down	834, 857
NFKBIA	GG**AAATT**CCCC	Down	834, 858
PLAU	GGG**AATTT**CC	Down	840, 859
RELB	GGGG**AATTC**C	Down	834, 860
TFPI2	GGGG**AATTC**C	Down	858, 861
TNC	GGG**AATTC**CT	Down	862, 863

Table 6.2: Many glucocorticoid-regulated pro-inflammatory genes contain NF-κB response elements containing cryptic GR binding sites. AATTY sequence is shown in bold.

If GR indeed recognizes this motif to repress the transcription of pro-inflammatory genes, we would expect this motif to be highly conserved despite its lack of sequence-specific recognition by NF-κB binding. Indeed, we find that the AATTY motif is extremely well conserved - often more so than the bases that contact NF-κB itself (Figure 6.9c). In a remarkable example, the AATTT motif at the *IL8* NF-κB response element is perfectly conserved from mammals to reptiles (Figure 6.9c). This element also exhibits strong functional conservation, as human GR DBD retains the ability to bind the *IL8* NF-κB response element from multiple species (Figure 6.9d). Finally, mutation of one or more of these conserved bases reduces or ablates the ability of dexamethasone to repress the human *IL8* promoter (Figure 6.9e), supporting both a sequence-specific recognition of the element by GR as well as a functional role for these highly conserved spacer bases at the *IL8* NF-κB response element.

6.3.5 A distinct set of coregulators is required for repression of NF-κB by GR

A small-scale screen of approximately 25 siRNA targets for coregulators required for suppression of the *IL6* gene by the estrogen receptor-α (ERα) ligands, estradiol (E2) or resveratrol, was recently reported.[864] This work revealed a number of distinct coregulator complexes required for repression. Here, we also screened for the effects of the siRNAs on dexamethasone. These included the steroid receptor coactivators 1-3, which provide one of the primary scaffolds for recruitment of other coregulators, such as the acetyl transferases, CBP, p300, and PCAF. We also examined nuclear receptor corepressors, HDACs, and components of the CoRest1/LSD1 and sirtiun (SRT) repressor complexes. MCF-7 cells were transfected with the control or target siRNAs for 24 hours, and then stimulated for 2 hr with TNF-α ± 10 nM dexamethasone. To our surprise, dexamethasone showed the same functional requirements as E2 with respect to the steroid receptor coactivators and acetyl transferases. While SRC2 has

been previously reported to be required for dexamethasone's suppressive effects, we found both SRC2 and SRC3 were required (Figure 6.10a). However, SRC1 and SRC3 had distinct roles, as the siRNAs increased both the vehicle and TNF-α induced IL-6 expression, demonstrating a general role in suppressing this gene. As with estradiol, CBP was required for the suppressive effects of dexamethasone, but P300 and PCAF were not (Figure 6.10b). In fact, knockdown of p300 and PCAF completely blocked TNF-α induced induction of *IL6*. In our previous work, we showed that estradiol mediated recruitment of CBP to the *IL6* promoter and displacement of p300, which was associated with a block in acetylation of the p65 subunit of NF-κB at Lys310, an acetylation event required for transcriptional activity. Here we demonstrate that dexamethasone utilizes the same caste of coregulators for repression of *IL6*, suggesting a similar mechanism of action.

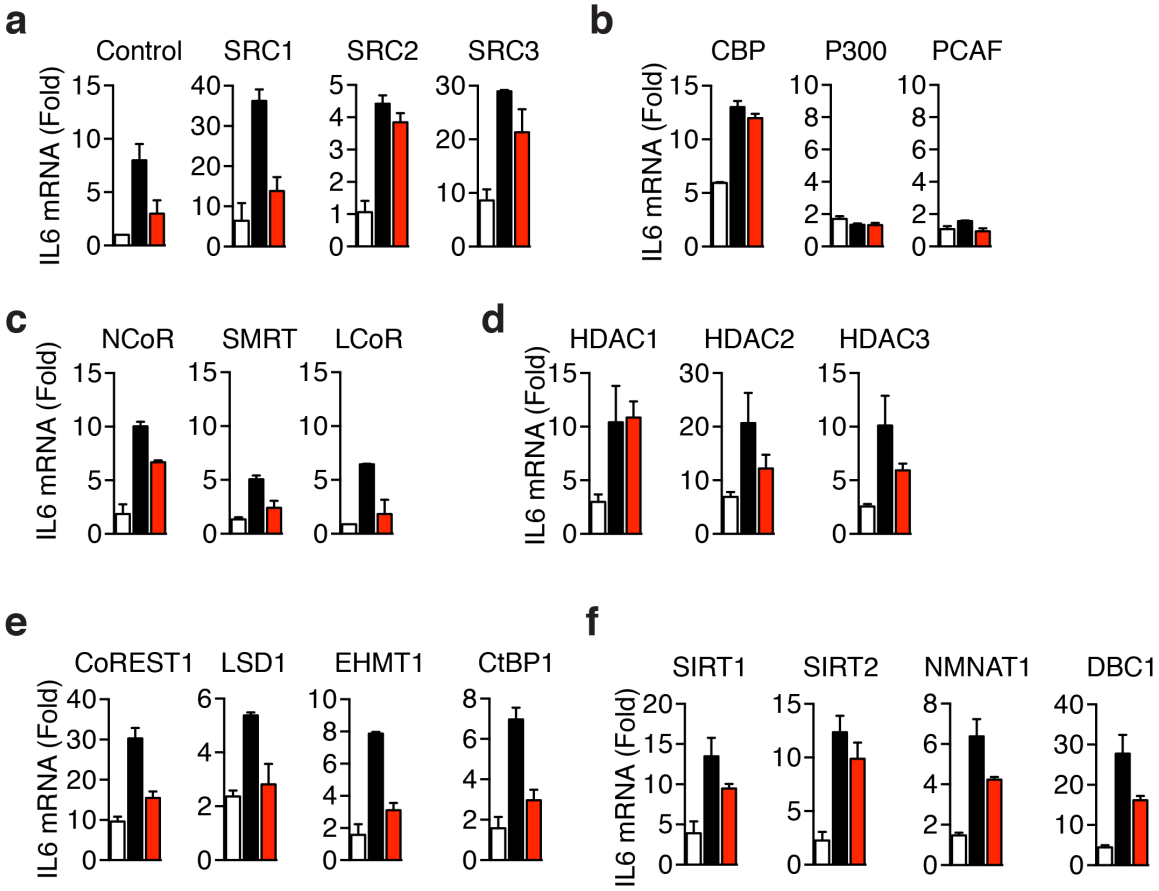


Figure 6.10: A distinct set of co-regulators is required for repression of TNF-α-driven transcription of the *IL6* gene. In all panels, red represents *IL6* levels (measured by RT-PCR) after 2 hours of TNF-α treatment, and black indicates *IL6* levels after TNF-α and dexamethasone co-treatment. siRNA used in each experiment is indicated above each graph. (a) The co-activator proteins SRC2 and SRC3 are both required for dexamethasone-mediated repression of *IL6*. (b) p300 and PCAF are required for TNF-α-mediated induction of *IL6*. The co-repressor protein NCoR is also required for dexamethasone-mediated repression, as is the histone-modifying enzyme HDAC1 (but not HDAC2 or HDAC3; panel d). (e,f) Effects of knockdown of CoREST complex and SIRT1 and SIRT2 and their associated proteins on dexamethasone-mediated repression of *IL6*.

Unlike suppression mediated by E2, dexamethasone also required the nuclear receptor corepressor, NCoR (Figure 6.10c), and the deacetylase HDAC1 (Figure 6.10d), which forms a complex with NCoR. Despite effective knockdown of SMRT and its homology to NCoR, SMRT was not required, nor was another corepressor, LCoR (Figure 6.10c). While not involved in the ligand dependent effect, HDAC2 had a similar role to SRC1 and SRC3 in maintaining global repression of the *IL6* gene (Figure 6.10d). We also examined elements of the CoRest1/LSD1 repressor complex, which interacts with ERα and is required for estradiol's suppressive effects on IL-6. As we previously reported, CoREST1 maintained global repression of the IL-6 gene, while the enzymatic components, LSD1, and EHMT1, were not involved. Lastly, we examined effects of knockdown of SIRT1 and SIRT2 and the SIRT1 associated proteins, NMNAT1 and deleted in breast cancer-1 (DBC1). SIRT1 and SIRT2 are a family of deacetylases that require NAD+ as an obligate cofactor, and have both been implicated in suppressing NF-κB signaling. Here again, the effects were different between dexamethasone and E2, where both SIRT1 and to a greater extent SIRT2 were required for dexamethasone mediated suppression of IL-6, but not for inhibition by E2. NMNAT1 produces NAD+ and interacts directly with SIRT1, enabling it produce local concentration gradients and 'feed' NAD+ to SIRT1. DBC1 also interacts with SIRT1, but acts to inhibit enzymatic activity. Here, knockdown of NMNAT1 and DBC1 showed moderate effects on dexamethasone suppression, in contrast to the absolute requirement for DBC1 for E2 inhibition. Thus E2 and dexamethasone share some common mechanisms of action, but also a number of different requirements for suppression of *IL6*.

6.4 Discussion

Here, we propose that direct binding of GR to NF-κB response elements can mediate GC-mediated repression of pro-inflammatory genes and demonstrate that GR is capable of binding to NF-κB response elements in a sequence-specific manner. Such a mechanism suggests that GR and NF-κB compete for the same binding site in order to regulate genes such as *IL6*, *IL8*, *RELB*, *PLAU*, and *CCL2*. Indeed, we find that TNF-α reduces GR occupancy of the *IL6* and *IL8* promoters, and this response can be overcome by co-administration of dexamethasone (Figure 6.11a). However, both GR and p65 occupancies at the *IL6*, *IL8*, and *ICAM1* promoters seem to cycle immediately following dexamethasone and/or TNF-α treatment (Figure 6.11a-c), potentially indicating a competition for the element, a fluctuating number of DNA binding sites (perhaps mediated by changing chromatin state),[677] an increase in non-coding RNA levels,[2] or other factors.

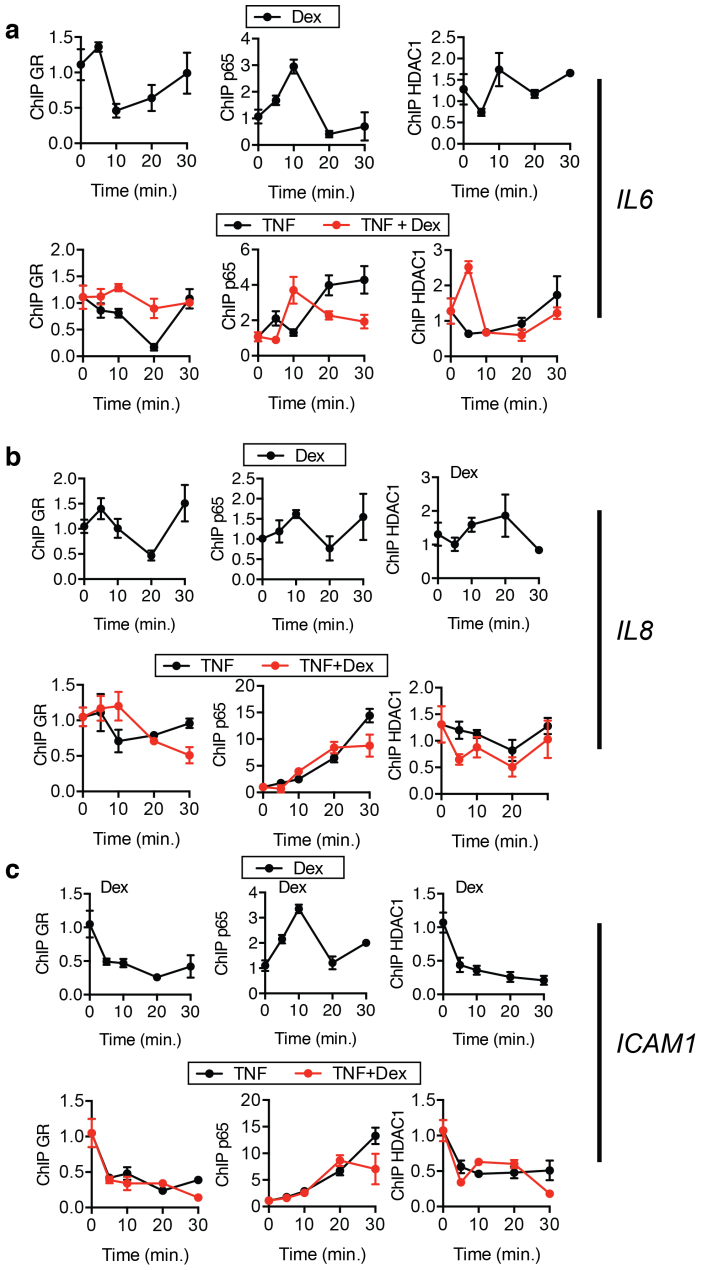


Figure 6.11: Kinetic ChIP at the *IL6*, *ICAM1*, and *IL8* promoters shows that GR and NF-κB occupancy cycle upon treatment of their respective activators. ChIP-seq of the *IL6* (a), *IL8* (b), and *ICAM1* (c) promoters upon dexamethasone treatment (top) and dexamethasone + TNF-α co-treatment (bottom). Of note, upon dexamethasone treatment, GR and p65 occupancy seem inversely correlated at all three promoters. In some cases, dexamethasone treatment reduces GR occupancy, perhaps ejecting unliganded GR from the promoter (see also Figure 6.4e).

Given that direct interactions between GR and NF-κB response elements would represent a dramatic shift in current models of GR action, it is worthwhile to re-examine previous studies in light of the hypothesis proposed here. Hundreds of mutation studies have been performed on the GR,[865] and their conclusions are quite complex. As we propose, at least three studies have demonstrated that activation of GR affects the DNA binding of NF-κB,[866–868] although this was not observed in at least three other reports.[869–871] Many studies have shown that GR and the p65 subunit of NF-κB interact,[575,759,866,871,872] but the existence of mutants able to bind NF-κB yet deficient in its repression demonstrate that a GR - NF-κB interaction may not be not sufficient for glucocorticoid-mediated repression of NF-κB.[871,872] In line with our results, the DNA binding domain of GR is crucial for repression of NF-κB;[833,866,873] swapping the GR DBD for that of the retinoic acid receptor abolishes NF-κB repression by GR.[759] Interestingly, replacing the C-terminal ligand binding domain of GR with β-galactosidase does not abolish NF-κB repression by the receptor,[874] and later reports demonstrate that coactivator and corepressor recruitment is not necessary for GR-mediated NF-κB repression.[875] This is consistent with our model of competition of DNA binding between GR and NF-κB. Finally, the AATTY motif which we identify as critical for GR binding to NF-κB response elements is found at many pro-inflammatory genes which are responsive to glucocorticoid treatment (Table 6.2).

Interestingly, ChIP-seq studies on the GR have identified the NF-κB motif (at perhaps 5-10% of GR-bound sites) in the presence of GR activation only,[677,831] despite the fact that GR activation does not alter the subcellular localization of NF-κB.[866,867,869] The NF-κB binding motif further appears at approximately 25% of GR-occupied sites after GR and NF-κB activation.[834] This could be explained by a tethering-like mechanism or alternatively by increased chromatin accessibility for GR binding. A recent study suggests the second mechanism may be more likely, as some GR binding sites without NF-κB or AP-1 co-occupancy are only accessible and

bound upon LPS treatment of macrophages.[831] The same study also concluded that tethering is unlikely to be the major mechanism of GR induced gene repression based on equal enrichment of the (+)GRE, a known sequence that binds GR directly, at genes both repressed and activated by GR.[831]

While we propose that direct GR binding to NF-κB response elements is an important mechanism of GC action, it is likely that multiple mechanisms govern GR-mediated suppression of inflammation and possible that multiple mechanisms act on any given gene. Some GR ligands affect multiple functions of the protein: for example, selective glucocorticoid receptor modulators have been reported to differentially affect repression of AP-1 and NF-κB.[876] Newly-discovered DNA binding motifs mediate some of the transcriptional repression by GR, as nGREs may represent a genome-wide class of GR binding sites.[574] However, ChIP-seq studies have not yet validated the widespread occupancy of nGREs by GR,[682] potentially due to their lower affinity for GR compared to dimeric (+)GRE elements (see Chapter 3). Moreover, similar GR-bound DNA sequences are enriched near genes both repressed and activated by GC administration.[831] While this finding indicates that DNA sequence and/or oligomerization state[682] may not be sufficient to predict GR action at a particular site, it also suggests that the current tethering model of protein-protein interactions between GR and NF-κB is insufficient to explain GC-mediated transcriptional responses. Instead, we propose that direct, sequence-specific DNA binding by GR to NF-κB response elements likely mediates anti-inflammatory transcriptional repression by GCs.

6.5 Acknowledgements

We thank Oskar Laur of the Emory Cloning Core for assistance with generating reporter constructs. W.H.H. is supported by an American Heart Association (AHA) predoctoral fellowship (13PRE16920012) and by a U.S. National Institutes of Health

(NIH) training grant to Emory University (5T32GM008602). A.G.H was supported by a Howard Hughes Medical Institute Science Education Program award (#52006923) to Emory University. E.A.O. is supported by NIH grant R01DK095750 from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), by an AHA grant 14GRNT20460124 and a W. M. Keck Foundation Medical Research Grant. D.J.K. is supported by NIH grant R01DK101871 from the NIDDK.

6.6 Methods

6.6.1 Protein expression and purification

DBDs were expressed and purified as previously described in Chapters 2 and 3: residues 417-506 of the human GR (GenBank ADP91252) and residues 593-671 of the human MR (GenBank AAA59571.1) were expressed as an N-terminal 6X-His fusion followed by a TEV protease cleavage site. *E. coli* BL21 (DE3) pLysS cells were induced with 0.3 mM IPTG for 4 hours at 30 °C after reaching an OD_{600} of 0.6. Proteins were purified via affinity chromatography (HisTrap) followed by gel filtration in 100 mM NaCl, 20 mM Tris-HCl pH 7.4, and 5% glycerol. Protein was concentrated to 4 mg/ml, flash frozen in liquid N_2, and stored at -80 °C until use. ^{15}N-GR DBD was expressed in *E. coli* BL21(DE3) pLysS cells with $^{15}NH_4Cl$ as the sole nitrogen source, and was purified as described above. The 6X-His tag was cleaved with TEV protease overnight at 4 °C, passed through an NiNTA column, and the flow through containing purified ^{15}N-GR DBD was collected and verified to be >99% pure by SDS-PAGE.

6.6.2 Reporter gene assays

Reporter gene assays were performed as previously reported described (see Chapter 5): indicated amounts of receptor plasmid (in the pcDNA3.1 vector), 50 ng of a pGL3 plasmid (Firefly luciferase) containing the indicated promoter (see below) constitu-

tively driven by the SV40 promoter and enhancer, and 10 ng of a constitutively active Renilla luciferase under the control of the pRL-TK promoter were transfected into HeLa cells with FuGene HD (Promega) according to the manufacturer's protocol. Twenty-four hours after transfection, cells were treated with indicated amounts of dexamethasone, TNF-α, and/or vehicle. Twenty-four hours following treatment, firefly and Renilla luciferase activities were read using a Biotek Synergy plate reader and the Dual-Glo Luciferase Assay System (Promega), according to the manufacturer's protocol. For figures, firefly divided by Renilla activity is shown, normalized to the control condition. GraphPad Prism was used to graph the data, and statistical tests used are located in the figure legends.

The sequences cloned into the pGL3 reporter vectors are as follows: *IL8*, Chr4: 73,740,106-73,740,505; *ICAM1*, Chr19:10,270,782-10,271,081; *CCL2*, Chr17:34,252,540-34,252,896; *RELB*, Chr19:45,001,228-45,001,545. All numbers correspond to the human GRCh38.p2 genome, accessed through Ensembl.[648] The *TSLP* construct was reported in Chapter 3.5.3.

6.6.3 *In vitro* binding assays

Ten nM of double-stranded, 6-FAM labeled DNA (Integrated DNA Technologies) was incubated with indicated amounts of protein in 100 mM NaCl, 20 mM Tris-HCl pH 7.4, and 5% glycerol. Formation of DBD - DNA complexes was monitored via fluorescence polarization on a Biotek Synergy plate reader at an excitation and emission wavelength of 485 and 528 nm, respectively. Data were graphed and analyzed in Prism 6 (Graphpad Software). For binding to NF-κB response elements, a two-site binding model was used.

Sequences of DNA constructs used for fluorescence polarization assays were: *PLAU*: 5'-(FAM) CCCTGGGAATTTCCTGATA-3' and 5' - TATCAGGAAATTC-CCAGGG - 3'; CCL2: 5'- GAGTGGGAATTTCCACTCA-3' and 5'- TGAGTGGAAATTCCCACT

3'; *IL8*: 5'-(FAM) AATCGTGGAATTTCCTCTG-3' and 5'- CAGAGGAAATTC-CACGATT - 3'. In all cases, (FAM) indicates the position of 6-FAM (fluorescein).

6.6.4 Crystallization and structure analysis

Crystals of the GR DBD - *RELB* NF-κB response element complex were grown in 0.1 M HEPES pH 7.5, 7 μM spermine, and 15% PEG 8000. Crystals of the GR DBD - *PLAU* NF-κB response element complex were grown in 0.05 M sodium cacodylate, 0.05 M spermine, and 16% PEG 400. Crystals of the GR DBD - *CCL2* NF-κB response element complex were grown in 0.1 M sodium malonate, 6% glycerol, and 5% PEG 3350. Crystals of the GR DBD - *IL8* NF-κB response element were grown in 0.1 M HEPES pH 7.5, 7.5% glycerol, and 22% PEG 20000. Crystals of the GR DBD - *ICAM1* NF-κB response element were grown in 0.1 M HEPES pH 7.7, 3% ethylene glycol, and 10% PEG 8000. DNA bases for figures are labeled by their position relative to the RELB-001, PLAU-001, CCL2-001, CXCL8-001, and ICAM-002 transcripts, respectively.[648] All crystals were grown at 20 °C via hanging drop vapor diffusion and flash cooled in mother liquor with the addition of 10-20% glycerol.

All x-ray data were collected at 1.00 Å wavelength at the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID and 22-BM beamlines at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at www.ser-cat.org/members.html. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38. Structures were solved by molecular replacement using PHASER and refined with phenix.refine in the PHENIX suite.[691] Coot was used for visualization model rebuilding and PDBREDO was used for validation and model improvement.[690,812] Figures were generated in MacPyMOL v1.7.0.3 (Schrödinger LLC).

6.6.5 NMR analysis

NMR data were collected on a Bruker 700 MHz (^{1}H frequency) NMR instrument equipped with a QCI cryoprobe. For DNA NMR experiments, the 19 nt *IL8* NF-κB DNA response element duplex was reconstituted in 20 mM phosphate (pH 6.7), 100 mM NaCl, 1 mM TCEP, 10% D_2O buffer to a final concentration of 437 μM, subsequently annealed by denaturing at 95 °C for 3 minutes and equilibrated to room temperature (20-23°C) overnight. Two-dimensional (2D) ^{1}H-detected NOESY was collected at 10 °C and 25 °C using 300-ms mixing time for *IL8* NF-κB DNA before and after adding 0.44:1 or 2.3:1 of ^{15}N-GR DBD. For protein NMR experiments, 2D [^{1}H,^{15}N]-HSQC spectra were collected at 25 °C for free ^{15}N-GR DBD protein or protein complexed with 0.44:1 or 2.3:1 of *IL8* NF-κB DNA duplex; or 1.5x (+)GRE consensus DNA sequence. Chemical shift perturbations were calculated using previously published GR DBD NMR chemical shifts[672] and calculated using the minimum chemical shift perturbation procedure[692] in the NMR analysis program NMRViewJ (OneMoon Scientific, Inc.). The DNA sequence used in NMR experiments was identical in sequence to that used in binding assays, with the fluorescein label excluded: 5' - AATCGTGGAATTTCCTCTG - 3' and 5'- CAGAGGAAATTCCACGATT - 3'.

6.6.6 Sequence alignments

Sequences and sequence alignments were obtained from the Ensembl database.[648] Human sequences are from the GRCh38 genome build. Geneious version 6.1.6 (Biomatters Limited) was used for sequence visualization.

Chapter 7

Conclusions

7.1 Preceding work

Signaling from the hypothalamic-pituitary-adrenal axis is critical for the maintenance of vertebrate homeostasis as well as mediating responses to stress and modulation of the immune system. The glucocorticoid receptor (GR) is the cellular sensor of cortisol, a small molecule that is the end product of hypothalamic-pituitary-adrenal axis signaling. As such, GR is often targeted in pro-inflammatory conditions such as asthma, transplant rejection, and psoriasis. The Nobel Prize in Physiology or Medicine was awarded in 1950 to Edward Calvin Kendall, Tadeus Reichstein, and Philip Showalter Hench for their discovery of endogenous glucocorticoids and their applicability to the treatment of rheumatoid arthritis. The development of glucocorticoids for pharmaceutical use has continued to the present day, although the persistence of severe side effects has remained a challenge in the development of glucocorticoids suitable for prolonged, systemic use (discussed briefly in Chapter 6.2).

To develop glucocorticoids that demonstrate improved side effect profiles, much work has been done to develop 'dissociated' glucocorticoids that retain anti-inflammatory actions while minimizing expression changes in glucocorticoid-controlled genes in processes such as metabolism (for examples, see references 739, 809, 868, 877–879). However, the development such compounds has been disappointing, perhaps due to an incomplete or misguided understanding of actions and mechanisms of the GR.[566] In the years leading up to this dissertation, GR has been discovered to bind to an increasing number of nucleic acid substrates: (+)GRE DNA sequences,[334] nGRE DNA sequences,[574] tRNA,[330] lincRNA,[2] and mRNA.[1,329] Additionally, I also propose the GR binds to NF-κB DNA response elements in Chapter 6. Unfortunately, structural and biochemical understanding of GR's interactions with nucleic acids has badly lagged the identification of such substrates. Thus, I believe the work presented in this dissertation is extremely timely for the cellular and molecular endocrinology field. Second, detailed understanding of the mechanisms by which GR binds to both

RNA and DNA may set a precedent for the many proteins that bind both forms of nucleic acid. In these concluding thoughts, I will discuss both of these points.

7.2 The role of nucleic acid substrates in GR function

The glucocorticoid receptor was originally cloned in 1985,[880,881] and functional work has since proceeded at a rapid pace given the role of GR-targeting compounds in the treatment of numerous pro-inflammatory conditions. The crystal structure of receptor's DNA binding domain (DBD) in complex with (+)GRE DNA was reported in 1991.[334] Thirteen additional DBD - (+)GRE crystal structures were reported in 2009, just as I began this work.[669] However, GR - (+)GRE interactions are only one facet of GR function; interactions of GR with tRNA, mRNA, and lincRNA had at that time been or were soon discovered,[1,2,329,330] and GR was hypothesized to bind to the NF-κB and AP-1 proteins.[575,759,866,871,872] Soon afterwards, in 2011, a new repressive DNA response element, the nGRE, was shown to have a significant role in GR-mediated gene repression.[574] Together, these works have led to a complex model for how GR acts upon various nucleic acid substrates to mediate the intracellular effects of steroid hormones (Figure 7.1). However, no significant structural or biochemical studies had been conducted on the interaction of GR with nGREs or any RNA substrates. The primary goal of this dissertation was to examine the mechanisms by which GR interacts with its less-well-studied nucleic acid substrates, with a particular focus on its interaction with the lincRNA Gas5 as well as repressive DNA response elements such as nGREs. Further, I aimed to relate these interactions and their detailed mechanisms to GR function in a cellular context.

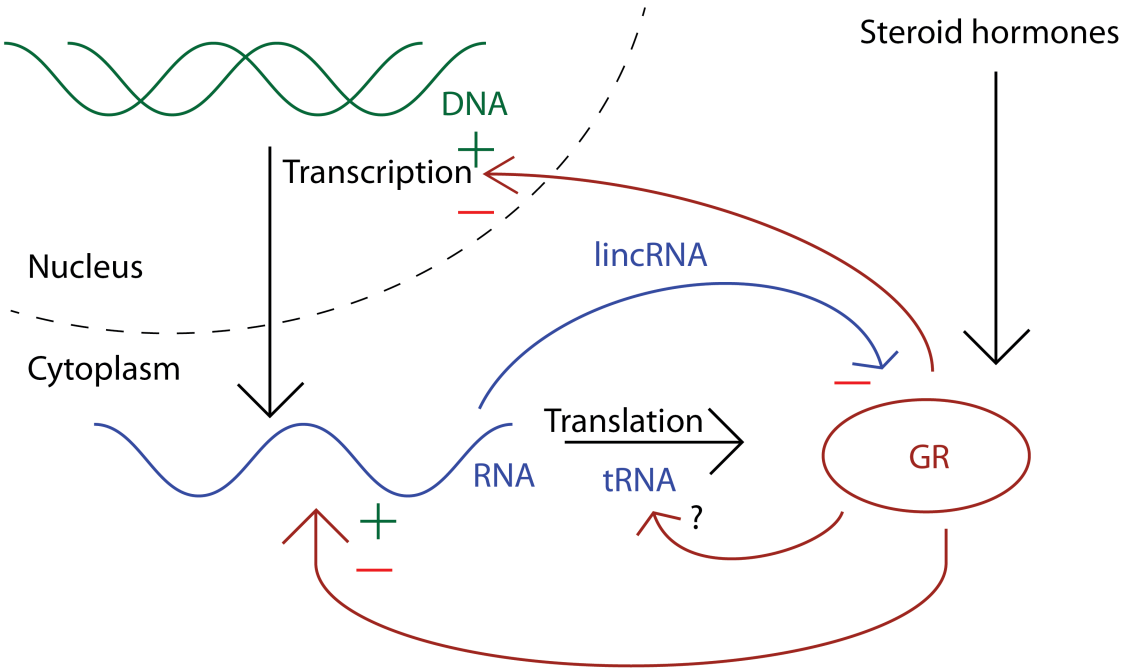


Figure 7.1: Model of GR's interactions with nucleic acids and their effects on cellular function. GR interacts with both RNA and DNA - a feature of many proteins, as discussed in Chapter 1. Specifically, GR itself can both activate or repress transcription based on its DNA binding stoichiometry and orientation and bound sequence (Chapters 3 and 6). Direct, DNA-mediated transcriptional repression is an evolutionarily conserved function of GR, as discussed in Chapter 5 and supported by the MR - (+)GRE DNA crystal structure (Chapter 4). Additionally, GR interacts with multiple RNA substrates, including tRNA,[330] mRNA,[1] and lincRNA.[2] The lincRNA Gas5 inhibits GR's ability for transcriptional activation through sequence-specific binding and serving as a 'decoy' in competition with GR - DNA interactions (Chapter 2).

In Chapter 2, I present the detailed results of the biochemical and structural analysis of the GR - Gas5 interaction. Of particular interest, Gas5 is specific for 3-keto steroid receptors - binding GR and its related SRs - but not the estrogen receptor (Figure 2.2). GR itself recognizes Gas5 in a sequence-specific manner (Figure 2.6). Mutation of the specifically-recognized bases in Gas5 ablates the ability of Gas5 to both repress SR-mediated transcription (Figure 2.6) and induce apoptosis (Figure 2.7). These finding suggests that GR's recognition of other RNA substrates is sequence specific, which is supported by a seemingly specific recognition of tRNAs and mRNAs by GR.[1,330] However, this also implies that recognition of RNA by GR is dictated at the RNA sequence level, and transcriptome-wide experiments such as CLIP-seq may be required to provide a comprehensive catalog of RNA motifs that permit recognition by GR. Additional experiments could include treating cells concomitantly with glucocorticoids and a transcription inhibitor such as actinomycin D. RNA-seq or RT-PCR of such treated cells could provide a list of RNAs targeted by GR in cells. It is likely that RNA binding is responsible for much of GR's cellular effects, and there is currently extraordinarily little understanding of the mechanisms by which GR mediates mRNA turnover.

Fortunately, multiple genome-wide studies have been conducted to identify sites at which GR interacts with DNA (or DNA-bound factors),[565,682,831,834,882] and thus GR's interactions with DNA are more completely understood than those with RNA. However, considerable controversy still exists about how GR interacts with DNA in a cellular context and how these interactions subsequently lead to gene activation or - particularly - gene repression. In Chapter 3, I report the crystallographic and biochemical analysis of GR with the nGRE, a recently-described repressive response element.[574] Our results suggest a model in which monomeric GR is bound at repressive response elements and dimeric GR is bound at response elements that promote transcriptional activation. This model remains controversial; in its support, studies

have shown that GR ligands that inhibit dimer formation are capable of gene repression (but not activation).[737,739] *In vitro* studies also show that full-length GR has little affinity for dimerization when not bound to DNA.[725,883] However, studies using GFP-tagged GR suggest that little monomeric exists within cells treated with glucocorticoids,[682] yet this assertion was disputed recently by a report demonstrating that monomeric ChIP-exo GR footprints are found in cells.[882] I also demonstrate that monomeric DNA binding is a feature of the earliest 3-keto SR (Chapter 5), suggesting that monomeric response element binding may be an ancient feature of GR.

A further point of contention is how extensively GR interacts with DNA-bound factors without binding to DNA itself, a model known as 'tethering'. ChIP-seq studies routinely find AP-1 and NF-κB motifs enriched at genomic GR binding sites. This observation has long been proposed to be a result of tethering, although recent studies[831] as well as my own work (Figure 6.4) have proposed that this explanation may be incorrect. GR is certainly capable of repressing transcription by binding to DNA itself (Chapter 3). In Chapter 6, I demonstrate that GR can bind to NF-κB response elements, implying that direct DNA binding by the GR to these elements - rather than tethering - is the cause of much glucocorticoid-mediated repression. However, this finding will likely be controversial and require further studies for complete verification. For example, ChIP-seq studies with a DNA-binding deficient GR mutant (such as that in Chapter 6.3.2) may be insightful. Even further, transgenic mice with mutations that abrogate DNA binding or force monomerization[682] would provide powerful *in vivo* insights into the importance of GR's DNA binding and oligomerization status for transcriptional repression.

Despite the fact that the functional consequences of GR's binding to both DNA and RNA are not clear at this time, it is indisputable that the ability of GR to bind both diverse sequences within genomic DNA as well as mRNA allows GR to effect both fast, non-genomic effects at a post-transcriptional level as well as deeper,

longer-lasting changes on gene expression at the transcriptional level. Additionally, GR's affinity for the lincRNA Gas5 allows specific cells to abrogate DNA-dependent functions of GR. Such properties make GR an ideal end-effector for the hypothalamic-pituitary axis.

7.3 Generalizability of GR as a DNA/RNA binding protein

The ability of GR to bind both RNA and DNA in order to perform its cellular functions raised the possibility that other proteins exhibit similar bifunctionality to control gene expression. This possibility is addressed in Chapter 1. I performed a thorough bioinformatics and literature search for proteins that bind both DNA and RNA (see Table 1.1 and Figure 1.1). As discussed in Chapter 1.3, approximately 2% of the human proteome is capable of binding DNA and RNA. Strikingly, only one previous review had been written on DNA/RNA binding proteins (DRBPs),[884] which appeared in 2002 and identified fewer than ten DRBPs. At the time of this dissertation a much larger literature on protein - nucleic acid interactions is available, including high-throughput studies designed to detect novel protein - nucleic acid interactions. However, the authors of the 2002 review were particularly prescient in noting that their examples suggested "the possibility that RNA binding by DNA-binding proteins is more common than currently appreciated."[884] Further, their work "motivated [them] to identify artificial RNA sequences that competitively inhibit a DNA-binding transcription factor not known to have a natural RNA partner."[884] With the recent discovery that mammalian cells produce thousands of lincRNAs, it is likely that many natural RNA partners for DNA-binding proteins exist.

A seminal example of this phenomenon is the inhibition of GR by the lincRNA Gas5.[2] The studies presented in Chapter 2 are the most detailed biochemical examination of a protein - lincRNA interaction to date. Notably, all 3-keto SRs are capable of binding to Gas5, mirroring their affinity for (+)GREs (Figure 2.2). The exonic steroid receptor-binding region of Gas5 (the GREM) appeared much later than the 3-keto SRs, indicating that lincRNAs are capable of evolving to acquire protein targets (Figure 2.10). Such evolution need not be widespread; a pseudogene termed

Lethe is a mouse non-coding RNA capable of binding to NF-κB and inhibiting its transcriptional activity,[9] but the *Lethe* gene is not found in other rodent species.[648] Remarkably, Lethe expression is modulated by GR and pro-inflammatory cytokines,[9] indicating that individual species may have separately evolved RNAs for varying protein targets; this prospect greatly increases the complexity of inter-species biological differences.

However, from the relatively few DRBPs that have been studied, I can draw conclusions about general mechanisms of DRBP action, of which three are proposed in Chapter 1.4. Briefly, DNA and RNA binding may be competitive (Chapter 1.4.1), where RNA binding prevents a protein from activating or repressing transcription. A DRBP may bind to both the promoters of its target genes or its mRNA product, allowing multifactorial control of gene expression (Chapter 1.4.2). Finally, a protein may bind DNA and RNA simultaneously for a variety of purposes (Chapter 1.4.3). GR fits squarely into these first two categories, and this dissertation has focused on GR's roles within the first category. This leaves much work to be done, particularly in identifying the relative role (compared to DNA binding) of GR's mRNA binding ability in achieving the anti-inflammatory effects of glucocorticoids. Beyond GR, each of the DRBPs identified here (see Table 1.1) and in the future will likely have its own rules for subcellular localization, oligomerization, recognition of epigenetic modifications, and sequence specificity (or lack thereof) in DNA/RNA binding.

This complexity will require interactions of protein with DNA/RNA to be the subject of intense study for the foreseeable future. However, I believe that this dissertation provides valuable insight into GR as a model protein that binds both DNA and RNA. As demonstrated throughout this work, GR is capable of binding diverse types of nucleic acids, making it a terrific model for DRBP function. Furthermore, its central role in controlling responses to stress, metabolism, and immunity make it a drug target of great importance and subject of intense previous and ongoing study. I have

demonstrated GR recognition of RNA is sequence specific. Moreover, I show that GR recognizes multiple DNA response elements to mediate transcriptional repression in a sequence-specific manner. This recognition is evolutionarily conserved and does not necessarily extend to paralogous proteins. These insights challenge current models of GR action such as tethering and provide a basis for the future study of GR-mediated transrepression as well as the broader study of proteins that recognize both RNA and DNA to modulate gene expression.

Chapter 8

Bibliography

[1] Ishmael, F. T, Fang, X, Houser, K. R, Pearce, K, Abdelmohsen, K, Zhan, M, Gorospe, M, & Stellato, C. (2011) The human glucocorticoid receptor as an RNA-binding protein: global analysis of glucocorticoid receptor-associated transcripts and identification of a target RNA motif. *J. Immunol.* **186**, 1189–98.

[2] Kino, T, Hurt, D. E, Ichijo, T, Nader, N, & Chrousos, G. P. (2010) Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci. Signal.* **3**, ra8.

[3] Davis, B. N, Hilyard, A. C, Nguyen, P. H, Lagna, G, & Hata, A. (2010) Smad proteins bind a conserved RNA sequence to promote microRNA maturation by Drosha. *Mol. Cell* **39**, 373–384.

[4] Shi, L, Godfrey, W. R, Lin, J, Zhao, G, & Kao, P. N. (2007) NF90 regulates inducible IL-2 gene expression in T cells. *J. Exp. Med.* **204**, 971–7.

[5] Shi, L, Qiu, D, Zhao, G, Corthesy, B, Lees-Miller, S, Reeves, W. H, & Kao, P. N. (2007) Dynamic binding of Ku80, Ku70 and NF90 to the IL-2 promoter in vivo in activated T-cells. *Nucleic Acids Res.* **35**, 2302–10.

[6] Xu, B, Yang, W. H, Gerin, I, Hu, C. D, Hammer, G. D, & Koenig, R. J. (2009) Dax-1 and steroid receptor RNA activator (SRA) function as transcriptional coactivators for steroidogenic factor 1 in steroidogenesis. *Mol. Cell. Biol.* **29**, 1719–34.

[7] Lanz, R. B, McKenna, N. J, Onate, S. A, Albrecht, U, Wong, J, Tsai, S. Y, Tsai, M.-J, & OMalley, B. W. (1999) A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell* **97**, 17–27.

[8] Hung, T, Wang, Y, Lin, M. F, Koegel, A. K, Kotake, Y, Grant, G. D, Horlings, H. M, Shah, N, Umbricht, C, Wang, P, Wang, Y, Kong, B, Langerod, A, Borresen-Dale, A.-L, Kim, S. K, van de Vijver, M, Sukumar, S, Whitfield, M. L, Kellis, M, Xiong, Y, Wong, D. J, & Chang, H. Y. (2011) Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat. Genet.* **43**, 621–629.

[9] Rapicavoli, N. A, Qu, K, Zhang, J, Mikhail, M, Laberge, R. M, & Chang, H. Y. (2013) A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. *eLife* **2**, e00762–e00762.

[10] Lebruska, L. L & Maher III, L. J. (1999) Selection and characterization of an RNA decoy for transcription factor NF-κB. *Biochemistry* **38**, 3168–74.

[11] Duss, O, Michel, E, Yulikov, M, Schubert, M, Jeschke, G, & Allain, F. H. T. (2014) Structural basis of the non-coding RNA RsmZ acting as a protein sponge. *Nature* **509**, 588–592.

[12] Binns, D, Dimmer, E, Huntley, R, Barrell, D, O'Donovan, C, & Apweiler, R. (2009) QuickGO: a web-based tool for gene ontology searching. *Bioinformatics* **25**, 3045–6.

[13] Hu, S, Xie, Z, Onishi, A, Yu, X, Jiang, L, Lin, J, Rho, H.-s, Woodard, C, Wang, H, Jeong, J.-S, Long, S, He, X, Wade, H, Blackshaw, S, Qian, J, & Zhu, H. (2009)

Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. *Cell* **139**, 610–622.

[14] Castello, A, Fischer, B, Eichelbaum, K, Horos, R, Beckmann, B, Strein, C, Davey, N, Humphreys, D, Preiss, T, Steinmetz, L, Krijgsveld, J, & Hentze, M. (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–1406.

[15] (2013) Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **42**, D191–D198.

[16] Herbert, A. G, Spitzner, J. R, Lowenhaupt, K, & Rich, A. (1993) Z-DNA binding protein from chicken blood nuclei. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 3339–3342.

[17] Eggington, J. M, Greene, T, & Bass, B. L. (2011) Predicting sites of ADAR editing in double-stranded RNA. *Nat. Commun.* **2**, 319.

[18] Polson, A. G & Bass, B. L. (1994) Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase. *EMBO J.* **13**, 5701–5711.

[19] Placido, D, Brown, B. A, Lowenhaupt, K, Rich, A, & Athanasiadis, A. (2007) A left-handed RNA double helix bound by the Z domain of the RNA-editing enzyme ADAR1. *Structure* **15**, 395–404.

[20] Yang, W, Chendrimada, T. P, Wang, Q, Higuchi, M, Seeburg, P. H, Shiekhattar, R, & Nishikura, K. (2005) Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat. Struct. Mol. Biol.* **13**, 13–21.

[21] Athanasiadis, A. (2012) Zalpha-domains: At the intersection between RNA editing and innate immunity. *Semin. Cell Dev. Biol.* **23**, 275–280.

[22] Schwartz, T, Rould, M. A, Lowenhaupt, K, Herbert, A, & Rich, A. (1999) Crystal Structure of the Zα Domain of the Human Editing Enzyme ADAR1 Bound to Left-Handed Z-DNA. *Science* **284**, 1841–1845.

[23] Barraud, P & Allain, F. H. (2012) ADAR proteins: double-stranded RNA and Z-DNA binding domains. *Curr. Top. Microbiol. Immunol.* **353**, 35–60.

[24] Larijani, M, Petrov, A. P, Kolenchenko, O, Berru, M, Krylov, S. N, & Martin, A. (2006) AID Associates with Single-Stranded DNA with High Affinity and a Long Complex Half-Life in a Sequence-Independent Manner. *Mol. Cell. Biol.* **27**, 20–30.

[25] Nonaka, T, Doi, T, Toyoshima, T, Muramatsu, M, Honjo, T, & Kinoshita, K. (2009) Carboxy-terminal domain of AID required for its mRNA complex formation in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 2747–2751.

[26] Conticello, S. G, Langlois, M, Yang, Z, & Neuberger, M. S. (2007) in *Adv. Immunol.*, eds. Frederick, W. A & Tasuku, H. (Academic Press) Vol. Volume 94, pp. 37–73.

[27] Stavnezer, J. (2011) Complex regulation and function of activation-induced cytidine deaminase. *Trends Immunol.* **32**, 194–201.

[28] Boyko, V, Mudrak, O, Svetlova, M, Negishi, Y, Ariga, H, & Tomilin, N. (1994) A major cellular substrate for protein kinases, annexin II, is a DNA-binding protein. *FEBS Lett.* **345**, 139–142.

[29] Filipenko, N. R, MacLeod, T. J, Yoon, C. S, & Waisman, D. M. (2003) Annexin A2 Is a Novel RNA-binding Protein. *J. Biol. Chem.* **279**, 8723–8731.

[30] Wang, C.-Y & Lin, C.-F. (2014) Annexin A2: Its Molecular Regulation and Cellular Expression in Cancer Development. *Dis. Markers* **2014**, 10.

[31] Krishna, P, Kennedy, B. P, Waisman, D. M, van de Sande, J. H, & McGhee, J. D. (1990) Are many Z-DNA binding proteins actually phospholipid-binding proteins? *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1292–5.

[32] Kijas, A. W, Harris, J. L, Harris, J. M, & Lavin, M. F. (2006) Aprataxin forms a discrete branch in the HIT (histidine triad) superfamily of proteins with both DNA/RNA binding and nucleotide hydrolase activities. *J. Biol. Chem.* **281**, 13939–13948.

[33] Tumbale, P, Williams, J. S, Schellenberg, M. J, Kunkel, T. A, & Williams, R. S. (2014) Aprataxin resolves adenylated RNA-DNA junctions to maintain genome integrity. *Nature* **506**, 111–5.

[34] Ahel, I, Rass, U, El-Khamisy, S. F, Katyal, S, Clements, P. M, McKinnon, P. J, Caldecott, K. W, & West, S. C. (2006) The neurodegenerative disease protein aprataxin resolves abortive DNA ligation intermediates. *Nature* **443**, 713–6.

[35] Zillner, K, Filarsky, M, Rachow, K, Weinberger, M, Lngst, G, & Nmeth, A. (2013) Large-scale organization of ribosomal DNA chromatin is regulated by Tip5. *Nucleic Acids Res.* **41**, 5251–5262.

[36] Mayer, C, Schmitz, K.-M, Li, J, Grummt, I, & Santoro, R. (2006) Intergenic Transcripts Regulate the Epigenetic State of rRNA Genes. *Mol. Cell* **22**, 351–361.

[37] Guetg, C, Scheifele, F, Rosenthal, F, Hottiger, M, & Santoro, R. (2012) Inheritance of Silent rDNA Chromatin Is Mediated by PARP1 via Noncoding RNA. *Mol. Cell* **45**, 790–800.

[38] Strohner, R, Nemeth, A, Jansa, P, Hofmann-Rohrer, U, Santoro, R, Längst, G, & Grummt, I. (2001) NoRC–a novel member of mammalian ISWI-containing chromatin remodeling machines. *EMBO J.* **20**, 4892–4900.

[39] Kasof, G. M, Goyal, L, & White, E. (1999) Btf, a Novel Death-Promoting Transcriptional Repressor That Interacts with Bcl-2-Related Proteins. *Mol. Cell. Biol.* **19**, 4390–4404.

[40] Liu, H, Lu, Z. G, Miki, Y, & Yoshida, K. (2007) Protein Kinase C Induces Transcription of the TP53 Tumor Suppressor Gene by Controlling Death-Promoting Factor Btf in the Apoptotic Response to DNA Damage. *Mol. Cell. Biol.* **27**, 8480–8491.

[41] Baltz, A, Munschauer, M, Schwanhusser, B, Vasile, A, Murakawa, Y, Schueler, M, Youngs, N, Penfold-Brown, D, Drew, K, Milek, M, Wyler, E, Bonneau, R, Selbach, M, Dieterich, C, & Landthaler, M. (2012) The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Mol. Cell* **46**, 674–690.

[42] Varia, S, Potabathula, D, Deng, Z, Bubulya, A, & Bubulya, P. A. (2013) Btf and TRAP150 have distinct roles in regulating subcellular mRNA distribution. *Nucleus* **4**, 229–40.

[43] Paull, T. T, Cortez, D, Bowers, B, Elledge, S. J, & Gellert, M. (2001) Direct DNA binding by Brca1. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6086–6091.

[44] Kawai, S & Amano, A. (2012) BRCA1 regulates microRNA biogenesis via the DROSHA microprocessor complex. *J. Cell Biol.* **197**, 201–208.

[45] Sturdy, A, Naseem, R, & Webb, M. (2004) Purification and characterisation of a soluble N-terminal fragment of the breast cancer susceptibility protein BRCA1. *J. Mol. Biol.* **340**, 469–75.

[46] Roy, R, Chun, J, & Powell, S. N. (2012) BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat. Rev. Cancer* **12**, 68–78.

[47] Nehls, P, Keck, T, Greferath, R, Spiess, E, Glaser, T, Rothbarth, K, Stammer, H, & Werner, D. (1998) cDNA cloning, recombinant expression and characterization of polypetides with exceptional DNA affinity. *Nucleic Acids Res.* **26**, 1160–6.

[48] Schilders, G, van Dijk, E, & Pruijn, G. J. (2007) C1D and hMtr4p associate with the human exosome subunit PM/Scl-100 and are involved in pre-rRNA processing. *Nucleic Acids Res.* **35**, 2564–72.

[49] Zamir, I, Dawson, J, Lavinsky, R. M, Glass, C. K, Rosenfeld, M. G, & Lazar, M. A. (1997) Cloning and characterization of a corepressor and potential component of the nuclear hormone receptor repression complex. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 14400–14405.

[50] Li, G, Liu, J. H, Abu-Asab, M, Masabumi, S, & Maru, Y. (2010) XPB Induces C1D Expression to Counteract UV-Induced Apoptosis. *Mol. Cancer Res.* **8**, 885–895.

[51] Hou, H, Wang, F, Zhang, W, Wang, D, Li, X, Bartlam, M, Yao, X, & Rao, Z. (2010) Structure-Functional Analyses of CRHSP-24 Plasticity and Dynamics in Oxidative Stress Response. *J. Biol. Chem.* **286**, 9623–9635.

[52] Pfeiffer, J. R, McAvoy, B. L, Fecteau, R. E, Deleault, K. M, & Brooks, S. A. (2010) CARHSP1 Is Required for Effective Tumor Necrosis Factor Alpha mRNA Stabilization and Localizes to Processing Bodies and Exosomes. *Mol. Cell. Biol.* **31**, 277–286.

[53] Fan, Y, Guo, Y, Hamblin, M, Chang, L, Zhang, J, & Chen, Y. E. (2011) Inhibition of Gluconeogenic Genes by Calcium-regulated Heat-stable Protein 1 via Repression of Peroxisome Proliferator-activated Receptor . *J. Biol. Chem.* **286**, 40584–40594.

[54] Zhao, T, Heyduk, T, Allis, C. D, & Eissenberg, J. C. (2000) Heterochromatin protein 1 binds to nucleosomes and DNA in vitro. *J. Biol. Chem.* **275**, 28332–8.

[55] Deng, Z, Norseen, J, Wiedmer, A, Riethman, H, & Lieberman, P. M. (2009) TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Mol. Cell* **35**, 403–413.

[56] Keller, C, Adaixo, R, Stunnenberg, R, Woolcock, K, Hiller, S, & Bhler, M. (2012) HP1Swi6 Mediates the Recognition and Destruction of Heterochromatic RNA Transcripts. *Mol. Cell* **47**, 215–227.

[57] Muchardt, C, Guilleme, M, Seeler, J. S, Trouche, D, Dejean, A, & Yaniv, M. (2002) Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1. *EMBO Rep.* **3**, 975–981.

[58] Eissenberg, J. C & Elgin, S. C. (2014) HP1a: a structural chromosomal protein regulating transcription. *Trends Genet.* **30**, 103–110.

[59] Kaller, M, Foldesi, B, & Nellen, W. (2007) Localization and organization of protein factors involved in chromosome inheritance in Dictyostelium discoideum. *Biol. Chem.* **388**, 355–65.

[60] Forzati, F, Federico, A, Pallante, P, Abbate, A, Esposito, F, Malapelle, U, Sepe, R, Palma, G, Troncone, G, Scarf, xF, Marzia, Arra, C, Fedele, M, & Fusco, A. (2012) CBX7 is a tumor suppressor in mice and humans. *J. Clin. Invest.* **122**, 612–623.

[61] Yap, K. L, Li, S, Munoz-Cabello, A. M, Raguz, S, Zeng, L, Mujtaba, S, Gil, J, Walsh, M. J, & Zhou, M. M. (2010) Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol. Cell* **38**, 662–74.

[62] Aguilo, F, Zhou, M. M, & Walsh, M. J. (2011) Long Noncoding RNA, Polycomb, and the Ghosts Haunting INK4b-ARF-INK4a Expression. *Cancer Res.* **71**, 5365–5369.

[63] Sugimoto, K, Yata, H, Muro, Y, & Himeno, M. (1994) Human Centromere Protein C (CENP-C) is a DNA-binding protein which possesses a novel DNA-binding motif. *J. Biochem.* **116**, 877–881.

[64] Du, Y, Topp, C. N, & Dawe, R. K. (2010) DNA binding of centromere protein C (CENPC) is stabilized by single-stranded RNA. *PLoS Genet.* **6**, e1000835.

[65] Kwon, M.-S, Hori, T, Okada, M, & Fukagawa, T. (2007) CENP-C is involved in chromosome segregation, mitotic checkpoint function, and kinetochore assembly. *Mol. Biol. Cell* **18**, 2155–2168.

[66] Chun, Y, Park, B, Koh, W, Lee, S, Cheon, Y, Kim, R, Che, L, & Lee, S. (2011) New centromeric component CENP-W is an RNA-associated nuclear matrix protein that interacts with Nucleophosmin/B23 Protein. *J. Biol. Chem.* **286**, 42758–42769.

[67] Topp, C. N, Zhong, C. X, & Dawe, R. K. (2004) Centromere-encoded RNAs are integral components of the maize kinetochore. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15986–91.

[68] Kato, H, Jiang, J, Zhou, B. R, Rozendaal, M, Feng, H, Ghirlando, R, Xiao, T. S, Straight, A. F, & Bai, Y. (2013) A conserved mechanism for centromeric nucleosome recognition by centromere protein CENP-C. *Science* **340**, 1110–3.

[69] Lum, L. S, Sultzman, L. A, Kaufman, R. J, Linzer, D. I, & Wu, B. J. (1990) A cloned human CCAAT-box-binding factor stimulates transcription from the human hsp70 promoter. *Mol. Cell. Biol.* **10**, 6709–6717.

[70] Rajavashisth, T, Taylor, A, Andalibi, A, Svenson, K, & Lusis, A. (1989) Identification of a zinc finger protein that binds to the sterol regulatory element. *Science* **245**, 640–643.

[71] Pellizzoni, L, Lotti, F, Maras, B, & Pierandrei-Amaldi, P. (1997) Cellular nucleic acid binding protein binds a conserved region of the 5' UTR of Xenopus laevis ribosomal protein mRNAs. *J. Mol. Biol.* **267**, 264–75.

[72] Calcaterra, N. B, Armas, P, Weiner, A. M, & Borgognone, M. (2010) CNBP: a multifunctional nucleic acid chaperone involved in cell death and proliferation control. *IUBMB Life* **62**, 707–14.

[73] Armas, P, Nasif, S, & Calcaterra, N. B. (2008) Cellular nucleic acid binding protein binds G-rich single-stranded nucleic acids and may function as a nucleic acid chaperone. *J. Cell. Biochem.* **103**, 1013–36.

[74] Jacquemin-Sablon, H, Triqueneaux, G, Deschamps, S, le Maire, M, Doniger, J, & Dautry, F. (1994) Nucleic acid binding and intracellular localization of unr, a protein with five cold shock domains. *Nucleic Acids Res.* **22**, 2643–2650.

[75] Chang, T. C, Yamashita, A, Chen, C. Y, Yamashita, Y, Zhu, W, Durdan, S, Kahvejian, A, Sonenberg, N, & Shyu, A. B. (2004) UNR, a new partner of poly(A)-binding protein, plays a key role in translationally coupled mRNA turnover mediated by the c-fos major coding-region determinant. *Genes Dev.* **18**, 2010–23.

[76] Dormoy-Raclet, V, Markovits, J, Malato, Y, Huet, S, Lagarde, P, Montaudon, D, Jacquemin-Sablon, A, & Jacquemin-Sablon, H. (2007) Unr, a cytoplasmic RNA-binding protein with cold-shock domains, is involved in control of apoptosis in ES and HuH7 cells. *Oncogene* **26**, 2595–605.

[77] Swinburne, I. A, Meyer, C. A, Liu, X. S, Silver, P. A, & Brodsky, A. S. (2006) Genomic localization of RNA binding proteins reveals links between pre-mRNA processing and transcription. *Genome Res.* **16**, 912–21.

[78] Cumming, S. A, Repellin, C. E, McPhillips, M, Radford, J. C, Clements, J. B, & Graham, S. V. (2002) The Human Papillomavirus Type 31 Late 3 untranslated

region contains a complex bipartite negative regulatory element. *J. Virol.* **76**, 5993–6003.

[79] Aragaki, M, Takahashi, K, Akiyama, H, Tsuchiya, E, Kondo, S, Nakamura, Y, & Daigo, Y. (2011) Characterization of a cleavage stimulation factor, 3' pre-RNA, subunit 2, 64 kDa (CSTF2) as a therapeutic target for lung cancer. *Clin. Cancer Res.* **17**, 5889–900.

[80] Ruepp, M. D, Schweingruber, C, Kleinschmidt, N, & Schumperli, D. (2011) Interactions of CstF-64, CstF-77, and symplekin: implications on localisation and function. *Mol. Biol. Cell* **22**, 91–104.

[81] Garbelli, A, Beermann, S, Di Cicco, G, Dietrich, U, & Maga, G. (2011) A motif unique to the human DEAD-box protein DDX3 is important for nucleic acid binding, ATP hydrolysis, RNA/DNA unwinding and HIV-1 replication. *PLoS ONE* **6**, e19810.

[82] Botlagunta, M, Vesuna, F, Mironchik, Y, Raman, A, Lisok, A, Winnard, P., J, Mukadam, S, Van Diest, P, Chen, J. H, Farabaugh, P, Patel, A. H, & Raman, V. (2008) Oncogenic role of DDX3 in breast cancer biogenesis. *Oncogene* **27**, 3912–22.

[83] Soto-Rifo, R, Rubilar, P. S, Limousin, T, de Breyne, S, Décimo, D, & Ohlmann, T. (2012) DEAD-box protein DDX3 associates with eIF4F to promote translation of selected mRNAs. *EMBO J.* **31**, 3745–56.

[84] Soto-Rifo, R & Ohlmann, T. (2013) The role of the DEAD-box RNA helicase DDX3 in mRNA metabolism. *Wiley Interdiscip. Rev. RNA* **4**, 369–85.

[85] Drewett, V, Molina, H, Millar, A, Muller, S, von Hesler, F, & Shaw, P. E. (2001) DNA-bound transcription factor complexes analysed by mass-spectrometry: binding of novel proteins to the human c-fos SRE and related sequences. *Nucleic Acids Res.* **29**, 479–87.

[86] Enukashvily, N, Donev, R, Sheer, D, & Podgornaya, O. (2005) Satellite DNA binding and cellular localisation of RNA helicase P68. *J. Cell Sci.* **118**, 611–22.

[87] Arun, G, Akhade, V. S, Donakonda, S, & Rao, M. R. (2012) mrhl RNA, a long noncoding RNA, negatively regulates Wnt signaling through its protein partner Ddx5/p68 in mouse spermatogonial cells. *Mol. Cell. Biol.* **32**, 3140–52.

[88] Rossler, O. G. (2001) Rearrangement of structured RNA via branch migration structures catalysed by the highly related DEAD-box proteins p68 and p72. *Nucleic Acids Res.* **29**, 2088–2096.

[89] Devany, M, Kotharu, N. P, & Matsuo, H. (2004) Solution NMR structure of the C-terminal domain of the human protein DEK. *Protein Sci.* **13**, 2252–9.

[90] Waldmann, T, Baack, M, Richter, N, & Gruss, C. (2003) Structure-specific binding of the proto-oncogene protein DEK to DNA. *Nucleic Acids Res.* **31**, 7003–10.

[91] McGarvey, T, Rosonina, E, McCracken, S, Li, Q, Arnaout, R, Mientjes, E, Nickerson, J. A, Awrey, D, Greenblatt, J, Grosveld, G, & Blencowe, B. J. (2000) The acute myeloid leukemia-associated protein, DEK, forms a splicing-dependent interaction with exon-product complexes. *J. Cell Biol.* **150**, 309–20.

[92] Kavanaugh, G. M, Wise-Draper, T. M, Morreale, R. J, Morrison, M. A, Gole, B, Schwemberger, S, Tichy, E. D, Lu, L, Babcock, G. F, Wells, J. M, Drissi, R, Bissler, J. J, Stambrook, P. J, Andreassen, P. R, Wiesmuller, L, & Wells, S. I. (2011) The human DEK oncogene regulates DNA damage response signaling and repair. *Nucleic Acids Res.* **39**, 7465–76.

[93] Vaughn, J. P, Creacy, S. D, Routh, E. D, Joyner-Butt, C, Jenkins, G. S, Pauli, S, Nagamine, Y, & Akman, S. A. (2005) The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. *J. Biol. Chem.* **280**, 38117–20.

[94] Creacy, S. D, Routh, E. D, Iwamoto, F, Nagamine, Y, Akman, S. A, & Vaughn, J. P. (2008) G4 Resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates. *J. Biol. Chem.* **283**, 34626–34634.

[95] Tran, H, Schilling, M, Wirbelauer, C, Hess, D, & Nagamine, Y. (2004) Facilitation of mRNA deadenylation and decay by the exosome-bound, DExH protein RHAU. *Mol. Cell* **13**, 101–11.

[96] Iwamoto, F, Stadler, M, Chalupnkov, K, Oakeley, E, & Nagamine, Y. (2008) Transcription-dependent nucleolar cap localization and possible nuclear function of DExH RNA helicase RHAU. *Exp. Cell Res.* **314**, 1378–1391.

[97] Chalupnkov, K, Lattmann, S, Selak, N, Iwamoto, F, Fujiki, Y, & Nagamine, Y. (2008) Recruitment of the RNA Helicase RHAU to stress granules via a unique RNA-binding domain. *J. Biol. Chem.* **283**, 35186–35198.

[98] Zhong, X & Safa, A. R. (2004) RNA helicase A in the MEF1 transcription factor complex up-regulates the MDR1 gene in multidrug-resistant cancer cells. *J. Biol. Chem.* **279**, 17134–17141.

[99] Manojlovic, Z & Stefanovic, B. (2012) A novel role of RNA helicase A in regulation of translation of type I collagen mRNAs. *RNA* **18**, 321–34.

[100] Fu, Q & Yuan, Y. A. (2013) Structural insights into RISC assembly facilitated by dsRNA-binding domains of human RNA helicase A (DHX9). *Nucleic Acids Res.* **41**, 3457–70.

[101] Feng, J, Bi, C, Clark, B. S, Mady, R, Shah, P, & Kohtz, J. D. (2006) The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes Dev.* **20**, 1470–1484.

[102] Sunwoo, J. B, Kim, S, Yang, L, Naik, T, Higuchi, D. A, Rubenstein, J. L, & Yokoyama, W. M. (2008) Distal-less homeobox transcription factors regulate development and maturation of natural killer cells. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10877–82.

[103] Kraus, P & Lufkin, T. (2006) Dlx homeobox gene control of mammalian limb and craniofacial development. *Am. J. Med. Genet.* **140**, 1366–74.

[104] de Melo, J, Zhou, Q. P, Zhang, Q, Zhang, S, Fonseca, M, Wigle, J. T, & Eisenstat, D. D. (2008) Dlx2 homeobox gene transcriptional regulation of Trkb neurotrophin receptor expression during mouse retinal development. *Nucleic Acids Res.* **36**, 872–84.

[105] Frauer, C, Rottach, A, Meilinger, D, Bultmann, S, Fellinger, K, Hasenoder, S, Wang, M, Qin, W, Soding, J, Spada, F, & Leonhardt, H. (2011) Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. *PLoS ONE* **6**, e16627.

[106] Di Ruscio, A, Ebralidze, A. K, Benoukraf, T, Amabile, G, Goff, L. A, Terragni, J, Figueroa, M. E, De Figueiredo Pontes, L. L, Alberich-Jorda, M, Zhang, P, Wu, M, D'Alo, F, Melnick, A, Leone, G, Ebralidze, K. K, Pradhan, S, Rinn, J. L, & Tenen, D. G. (2013) DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* **503**, 371–6.

[107] Pradhan, M, Estève, P.-O, Chin, H. G, Samaranayke, M, Kim, G.-D, & Pradhan, S. (2008) CXXC Domain of Human DNMT1 Is Essential for Enzymatic Activity. *Biochemistry* **47**, 10000–10009.

[108] Jurkowska, R. Z, Jurkowski, T. P, & Jeltsch, A. (2011) Structure and function of mammalian DNA methyltransferases. *ChemBioChem* **12**, 206–222.

[109] Goll, M. G, Kirpekar, F, Maggert, K. A, Yoder, J. A, Hsieh, C.-L, Zhang, X, Golic, K. G, Jacobsen, S. E, & Bestor, T. H. (2006) Methylation of tRNAAsp by the DNA Methyltransferase Homolog Dnmt2. *Science* **311**, 395–398.

[110] Holz-Schietinger, C & Reich, N. O. (2012) RNA modulation of the human DNA methyltransferase 3A. *Nucleic Acids Res.* **40**, 8550–8557.

[111] Song, J, Rechkoblit, O, Bestor, T. H, & Patel, D. J. (2011) Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* **331**, 1036–40.

[112] Yamashita, N, Shimazaki, N, Ibe, S, Kaneko, R, Tanabe, A, Toyomoto, T, Fujita, K, Hasegawa, T, Toji, S, Tamai, K, Yamamoto, H, & Koiwai, O. (2001) Terminal deoxynucleotidyltransferase directly interacts with a novel nuclear protein that is homologous to p65. *Genes Cells* **6**, 641–652.

[113] Defossez, P.-A, Kubota, T, Koiwai, O, Hori, K, Watanabe, N, & Koiwai, K. (2013) TdIF1 recognizes a specific DNA sequence through Its helix-turn-helix and AT-hook motifs to regulate gene transcription. *PLoS ONE* **8**, e66710.

[114] Nakamura, T, Arai, Y, Umehara, H, Masuhara, M, Kimura, T, Taniguchi, H, Sekimoto, T, Ikawa, M, Yoneda, Y, Okabe, M, Tanaka, S, Shiota, K, & Nakano, T. (2007) PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nat. Cell Biol.* **9**, 64–71.

[115] Payer, B, Saitou, M, Barton, S. C, Thresher, R, Dixon, J. P. C, Zahn, D, Colledge, W. H, Carlton, M. B. L, Nakano, T, & Surani, M. A. (2003) Stella is a maternal effect gene required for normal early development in mice. *Curr. Biol.* **13**, 2110–2117.

[116] Yuan, Y, Li, D. M, & Sun, H. (1998) PIR1, a novel phosphatase that exhibits high affinity to RNA ribonucleoprotein complexes. *J. Biol. Chem.* **273**, 20347–20353.

[117] Deshpande, T, Takagi, T, Hao, L, Buratowski, S, & Charbonneau, H. (1999) Human PIR1 of the protein-tyrosine phosphatase superfamily has RNA 5'-triphosphatase and diphosphatase activities. *J. Biol. Chem.* **274**, 16590–16594.

[118] Ray, R & Miller, D. M. (1991) Cloning and Characterization of a Human C-Myc Promoter-Binding Protein. *Mol. Cell. Biol.* **11**, 2154–2161.

[119] Hernández-Pérez, L, Depardón, F, Fernández-Ram'ırez, F, Sánchez-Trujillo, A, Bermudez-Crúz, R. M, Dangott, L, & Montañez, C. (2011) α-Enolase binds to RNA. *Biochimie* **93**, 1520–1528.

[120] Pancholi, V. (2001) Multifunctional α-enolase: its role in diseases. *Cell. Mol. Life Sci.* **58**, 902–920.

[121] Feo, S, Arcuri, D, Piddini, E, Passantino, R, & Giallongo, A. (2000) ENO1 gene product binds to the c-myc promoter and acts as a transcriptional repressor: Relationship with myc promoter-binding protein 1 (MBP-1). *FEBS Lett.* **473**, 47–52.

[122] Citterio, E, Van Den Boom, V, Schnitzler, G, Kanaar, R, Bonte, E, Kingston, R. E, Hoeijmakers, J. H. J, & Vermeulen, W. (2000) ATP-Dependent Chromatin Remodeling by the Cockayne Syndrome B DNA Repair-Transcription-Coupling Factor. *Mol. Cell. Biol.* **20**, 7643–7653.

[123] Selby, C. P & Sancar, A. (1997) Human transcription-repair coupling factor CSB/ERCC6 is a DNA-stimulated ATPase but is not a helicase and does not disrupt the ternary transcription complex of stalled RNA polymerase II. *J. Biol. Chem.* **272**, 1885–90.

[124] Berquist, B. R & Wilson III, D. M. (2009) Nucleic acid binding activity of human Cockayne syndrome B protein and identification of Ca^{2+} as a novel metal cofactor. *J. Mol. Biol.* **391**, 820–832.

[125] Orren, D. (1996) The human CSB (ERCC6) gene corrects the transcription-coupled repair defect in the CHO cell mutant UV61. *Nucleic Acids Res.* **24**, 3317–3322.

[126] Newman, J. C, Bailey, A. D, & Weiner, A. M. (2006) Cockayne syndrome group B protein (CSB) plays a general role in chromatin maintenance and remodeling. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 9613–9618.

[127] Schwabe, J. W. R, Chapman, L, Finch, J. T, & Rhodes, D. (1993) The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: How receptors discriminate between their response elements. *Cell* **75**, 567–578.

[128] Feldman, M, Kallos, J, & Hollander, V. P. (1981) RNA inhibits estrogen receptor binding to DNA. *J. Biol. Chem.* **256**, 1145–8.

[129] Katzenellenbogen, B. S, Choi, I, Delage-Mourroux, R, Ediger, T. R, Martini, P. G. V, Montano, M, Sun, J, Weis, K, & Katzenellenbogen, J. A. (2000) Molecular mechanisms of estrogen action: selective ligands and receptor pharmacology. *J. Steroid Biochem. Mol. Biol.* **74**, 279–285.

[130] Edwards, D. P. (2005) Regulation of signal transduction pathways by estrogen and progesterone. *Annu. Rev. Physiol.* **67**, 335–376.

[131] Duncan, R, Bazar, L, Michelotti, G, Tomonaga, T, Krutzsch, H, Avigan, M, & Levens, D. (1994) A sequence-specific, single-strand binding protein activates the far upstream element of c-myc and defines a new DNA-binding motif. *Genes Dev.* **8**, 465–80.

[132] Zheng, Y & Miskimins, W. K. (2011) Far upstream element binding protein 1 activates translation of p27Kip1 mRNA through its internal ribosomal entry site. *Int. J. Biochem. Cell Biol.* **43**, 1641–1648.

[133] Chien, H. L, Liao, C. L, & Lin, Y. L. (2011) FUSE binding protein 1 interacts with untranslated regions of Japanese encephalitis virus RNA and negatively regulates viral replication. *J. Virol.* **85**, 4698–4706.

[134] Zhang, J & Chen, Q. M. (2013) Far upstream element binding protein 1: a commander of transcription, translation and beyond. *Oncogene* **32**, 2907–16.

[135] Davis-Smyth, T, Duncan, R. C, Zheng, T, Michelotti, G, & Levens, D. (1996) The far upstream element-binding proteins comprise an ancient family of single-strand DNA-binding transactivators. *J. Biol. Chem.* **271**, 31679–31687.

[136] Gherzi, R, Chen, C. Y, Trabucchi, M, Ramos, A, & Briata, P. (2010) The role of KSRP in mRNA decay and microRNA precursor maturation. *Wiley Interdiscip. Rev. RNA* **1**, 230–9.

[137] Braddock, D. T, Louis, J. M, Baber, J. L, Levens, D, & Clore, G. M. (2002) Structure and dynamics of KH domains from FBP bound to single-stranded DNA. *Nature* **415**, 1051–6.

[138] Nicastro, G, García-Mayoral, M. F, Hollingworth, D, Kelly, G, Martin, S. R, Briata, P, Gherzi, R, & Ramos, A. (2012) Noncanonical G recognition mediates KSRP regulation of let-7 biogenesis. *Nat. Struct. Mol. Biol.* **19**, 1282–6.

[139] Baechtold, H, Kuroda, M, Sok, J, Ron, D, Lopez, B. S, & Akhmedov, A. T. (1999) Human 75-kDa DNA-pairing protein Is identical to the pro-oncoprotein TLS/FUS and is able to promote D-loop formation. *J. Biol. Chem.* **274**, 34337–34342.

[140] Wang, X, Arai, S, Song, X, Reichart, D, Du, K, Pascual, G, Tempst, P, Rosenfeld, M. G, Glass, C. K, & Kurokawa, R. (2008) Induced ncRNAs allosterically modify RNA-binding proteins *in cis* to inhibit transcription. *Nature* **454**, 126–30.

[141] Fujii, R & Takumi, T. (2005) TLS facilitates transport of mRNA encoding an actin-stabilizing protein to dendritic spines. *J. Cell Sci.* **118**, 5755–5765.

[142] Fiesel, F. C & Kahle, P. J. (2011) TDP-43 and FUS/TLS: cellular functions and implications for neurodegeneration. *FEBS J.* **278**, 3550–3568.

[143] Lanson Jr., N. A & Pandey, U. B. (2012) FUS-related proteinopathies: Lessons from animal models. *Brain Res.* **1462**, 44–60.

[144] Takahama, K, Kino, K, Arai, S, Kurokawa, R, & Oyoshi, T. (2008) Identification of RNA binding specificity for the TET-family proteins (Nucleic Acids Symp Ser (Oxf)).

[145] Costa, M, Ochem, A, Falaschi, A, & Staub, A. (1999) Human DNA helicase VIII: A DNA and RNA helicase corresponding to the G3BP protein, an element of the Ras transduction pathway. *Nucleic Acids Res.* **27**, 817–821.

[146] Taniuchi, K, Nishimori, I, & Hollingsworth, M. A. (2011) The N-terminal domain of G3BP enhances cell motility and invasion by posttranscriptional regulation of BART. *Mol. Cancer Res.* **9**, 856–866.

[147] Tourrière, H, Gallouzi, I. E, Chebli, K, Capony, J. P, Mouaikel, J, van der Geer, P, & Tazi, J. (2001) RasGAP-associated endoribonuclease G3Bp: selective RNA degradation and phosphorylation-dependent localization. *Mol. Cell. Biol.* **21**, 7747–60.

[148] Demarse, N. A, Ponnusamy, S, Spicer, E. K, Apohan, E, Baatz, J. E, Ogretmen, B, & Davies, C. (2009) Direct binding of glyceraldehyde 3-phosphate dehydrogenase to telomeric DNA protects telomeres against chemotherapy-induced rapid degradation. *J. Mol. Biol.* **394**, 789–803.

[149] Singh, R & Green, M. (1993) Sequence-specific binding of transfer RNA by glyceraldehyde-3-phosphate dehydrogenase. *Science* **259**, 365–368.

[150] Nagy, E & Rigby, W. F. C. (1995) Glyceraldehyde-3-phosphate dehydrogenase selectively binds AU-rich RNA in the NAD(+)-binding region (Rossmann fold). *J. Biol. Chem.* **270**, 2755–2763.

[151] Tristan, C, Shahani, N, Sedlak, T. W, & Sawa, A. (2011) The diverse functions of GAPDH: Views from different subcellular compartments. *Cell. Signal.* **23**, 317–323.

[152] Nicholls, C, Li, H, & Liu, J. P. (2012) GAPDH: a common enzyme with uncommon functions. *Clin. Exp. Pharmacol. Physiol.* **39**, 674–9.

[153] Zheng, L, Roeder, R. G, & Luo, Y. (2003) S phase activation of the histone H2B promoter by OCA-S, a coactivator complex that contains GAPDH as a key component. *Cell* **114**, 255–66.

[154] Grosse, F, Nasheuer, H. P, Scholtissek, S, & Schomburg, U. (1986) Lactate dehydrogenase and glyceraldehyde-phosphate dehydrogenase are single-stranded DNA-binding proteins that affect the DNA-polymerase-alpha-primase complex. *Eur. J. Biochem.* **160**, 459–67.

[155] Hanas, J. S, Hocker, J. R, Cheng, Y.-G, Lerner, M. R, Brackett, D. J, Lightfoot, S. A, Hanas, R. J, Madhusudhan, K. T, & Moreland, R. J. (2002) cDNA cloning, DNA binding, and evolution of mammalian transcription factor IIIA. *Gene* **282**, 43–52.

[156] Seifart, K. H, Wang, L, Waldschmidt, R, Jahn, D, & Wingender, E. (1989) Purification of human transcription factor IIIA and its interaction with a chemically synthesized gene encoding human 5 S rRNA. *J. Biol. Chem.* **264**, 1702–9.

[157] Moorefield, B & Roeder, R. G. (1994) Purification and characterization of human transcription factor IIIA. *J. Biol. Chem.* **269**, 20857–65.

[158] Clemens, K, Wolf, V, McBryant, S, Zhang, P, Liao, X, Wright, P, & Gottesfeld, J. (1993) Molecular basis for specific recognition of both RNA and DNA by a zinc finger protein. *Science* **260**, 530–533.

[159] Nolte, R. T, Conlin, R. M, Harrison, S. C, & Brown, R. S. (1998) Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA complex. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2938–43.

[160] Lu, D, Searles, M. A, & Klug, A. (2003) Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. *Nature* **426**, 96–100.

[161] Zhuma, T, Tyrrell, R, Sekkali, B, Skavdis, G, Saveliev, A, Tolaini, M, Roderick, K, Norton, T, Smerdon, S, Sedgwick, S, Festenstein, R, & Kioussis, D. (1999) Human HMG box transcription factor HBP1: a role in hCD2 LCR function. *EMBO J.* **18**, 6396–406.

[162] de Chiara, C, Menon, R. P, Adinolfi, S, de Boer, J, Ktistaki, E, Kelly, G, Calder, L, Kioussis, D, & Pastore, A. (2005) The AXH domain adopts alternative folds: the solution structure of HBP1 AXH. *Structure* **13**, 743–753.

[163] Tevosian, S. G, Shih, H. H, Mendelson, K. G, Sheppard, K. A, Paulson, K. E, & Yee, A. S. (1997) HBP1: a HMG box transcriptional repressor that is targeted by the retinoblastoma family. *Genes Dev.* **11**, 383–396.

[164] Zhang, X, Kim, J, Ruthazer, R, McDevitt, M. A, Wazer, D. E, Paulson, K. E, & Yee, A. S. (2006) The HBP1 transcriptional repressor participates in RAS-induced premature senescence. *Mol. Cell. Biol.* **26**, 8252–66.

[165] Pan, K, Chen, Y, Roth, M, Wang, W, Wang, S, Yee, A. S, & Zhang, X. (2013) HBP1-mediated transcriptional regulation of DNA methyltransferase 1 and its impact on cell senescence. *Mol. Cell. Biol.* **33**, 887–903.

[166] Yang, J & Everett, A. D. (2007) Hepatoma-derived growth factor binds DNA through the N-terminal PWWP domain. *BMC Mol. Biol.* **8**, 101.

[167] Thirant, C, Galan-Moya, E.-M, Dubois, L. G, Pinte, S, Chafey, P, Broussard, C, Varlet, P, Devaux, B, Soncin, F, Gavard, J, Junier, M.-P, & Chneiweiss, H. (2012) Differential proteomic analysis of human glioblastoma and neural stem cells reveals HDGF as a novel angiogenic secreted factor. *Stem Cells* **30**, 845–853.

[168] Zhao, J, Yu, H, Lin, L, Tu, J, Cai, L, Chen, Y, Zhong, F, Lin, C, He, F, & Yang, P. (2011) Interactome study suggests multiple cellular functions of hepatoma-derived growth factor (HDGF). *J. Proteomics* **75**, 588–602.

[169] Bremer, S, Klein, K, Sedlmaier, A, Abouzied, M, Gieselmann, V, & Franken, S. (2013) Hepatoma-derived growth factor and nucleolin exist in the same ribonucleoprotein complex. *BMC Biochem.* **14**, 2.

[170] Hughes, E. N, Engelsberg, B. N, & Billings, P. C. (1992) Purification of nuclear proteins that bind to cisplatin-damaged DNA. Identity with high mobility group proteins 1 and 2. *J. Biol. Chem.* **267**, 13520–13527.

[171] Bell, A. J, Chauhan, S, Woodson, S. A, & Kallenbach, N. R. (2008) Interactions of recombinant HMGB proteins with branched RNA substrates. *Biochem. Biophys. Res. Commun.* **377**, 262–267.

[172] Fang, P, Schachner, M, & Shen, Y. Q. (2012) HMGB1 in development and diseases of the central nervous system. *Mol. Neurobiol.* **45**, 499–506.

[173] Yanai, H, Ban, T, Wang, Z, Choi, M. K, Kawamura, T, Negishi, H, Nakasato, M, Lu, Y, Hangai, S, Koshiba, R, Savitsky, D, Ronfani, L, Akira, S, Bianchi, M. E, Honda, K, Tamura, T, Kodama, T, & Taniguchi, T. (2009) HMGB proteins function as universal sentinels for nucleic-acid-mediated innate immune responses. *Nature* **462**, 99–103.

[174] Stott, K, Tang, G. S, Lee, K. B, & Thomas, J. O. (2006) Structure of a complex of tandem HMG boxes and DNA. *J. Mol. Biol.* **360**, 90–104.

[175] Fiset, S & Chabot, B. (2001) hnRNP A1 may interact simultaneously with telomeric DNA and the human telomerase RNA in vitro. *Nucleic Acids Res.* **29**, 2268–75.

[176] Lau, J. S, Baumeister, P, Kim, E, Roy, B, Hsieh, T.-Y, Lai, M, & Lee, A. S. (2000) Heterogeneous nuclear ribonucleoproteins as regulators of gene expression through interactions with the human thymidine kinase promoter. *J. Cell. Biochem.* **79**, 395–406.

[177] Burd, C. G & Dreyfuss, G. (1994) RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J.* **13**, 1197–204.

[178] Guil, S & Cceres, J. F. (2007) The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nat. Struct. Mol. Biol.* **14**, 591–596.

[179] Jean-Philippe, J, Paz, S, & Caputi, M. (2013) hnRNP A1: The Swiss army knife of gene expression. *Int. J. Mol. Sci.* **14**, 18999–19024.

[180] Moran-Jones, K, Wayman, L, Kennedy, D. D, Reddel, R. R, Sara, S, Snee, M. J, & Smith, R. (2005) hnRNP A2, a potential ssDNA/RNA molecular adapter at the telomere. *Nucleic Acids Res.* **33**, 486–496.

[181] McKay, S. J & Cooke, H. (1992) hnRNP A2/B1 binds specifically to single stranded vertebrate telomeric repeat TTAGGGn. *Nucleic Acids Res.* **20**, 6461–4.

[182] Mayeda, A, Munroe, S. H, Caceres, J. F, & Krainer, A. R. (1994) Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *EMBO J.* **13**, 5483–95.

[183] Ding, J, Hayashi, M. K, Zhang, Y, Manche, L, Krainer, A. R, & Xu, R. M. (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.* **13**, 1102–15.

[184] Abdul-Manan, N & Williams, K. R. (1996) hnRNP A1 binds promiscuously to oligoribonucleotides: utilization of random and homo-oligonucleotides to discriminate sequence from base-specific binding. *Nucleic Acids Res.* **24**, 4063–70.

[185] Vassileva, M. T & Matunis, M. J. (2004) SUMO modification of heterogeneous nuclear ribonucleoproteins. *Mol. Cell. Biol.* **24**, 3623–3632.

[186] Shetty, S. (2005) Regulation of urokinase receptor mRNA stability by hnRNP C in lung epithelial cells. *Mol. Cell. Biochem.* **272**, 107–18.

[187] Holcik, M, Gordon, B. W, & Korneluk, R. G. (2003) The internal ribosome entry site-mediated translation of antiapoptotic protein XIAP is modulated by the heterogeneous nuclear ribonucleoproteins C1 and C2. *Mol. Cell. Biol.* **23**, 280–288.

[188] Chaudhury, A, Chander, P, & Howe, P. H. (2010) Heterogeneous nuclear ribonucleoproteins (hnRNPs) in cellular processes: Focus on hnRNP E1's multifunctional regulatory roles. *RNA* **16**, 1449–1462.

[189] Lee, S. Y, Park, J. H, Kim, S, Park, E. J, Yun, Y, & Kwon, J. (2005) A proteomics approach for the identification of nucleophosmin and heterogeneous nu-

clear ribonucleoprotein C1/C2 as chromatin-binding proteins in response to DNA double-strand breaks. *Biochem. J.* **388**, 7–15.

[190] Hossain, M. N, Fuji, M, Miki, K, Endoh, M, & Ayusawa, D. (2007) Downregulation of hnRNP C1/C2 by siRNA sensitizes HeLa cells to various stresses. *Mol. Cell. Biochem.* **296**, 151–7.

[191] Swanson, M. S & Dreyfuss, G. (1988) Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. *Mol. Cell. Biol.* **8**, 2237–41.

[192] Gorlach, M, Burd, C. G, & Dreyfuss, G. (1994) The determinants of RNA-binding specificity of the heterogeneous nuclear ribonucleoprotein C proteins. *J. Biol. Chem.* **269**, 23074–8.

[193] Tolnay, M, Vereshchagina, L. A, & Tsokos, G. C. (1999) Heterogeneous nuclear ribonucleoprotein D0B is a sequence-specific DNA-binding protein. *Biochem. J.* **338**, 417–25.

[194] Dempsey, L. A, Sun, H, Hanakahi, L. A, & Maizels, N. (1999) G4 DNA binding by LR1 and its subunits, nucleolin and hnRNP D: A role for G-G pairing in immunoglobulin switch recombination. *J. Biol. Chem.* **274**, 1066–1071.

[195] Kajita, Y, Nakayama, J, Aizawa, M, & Ishikawa, F. (1995) The UUAG-specific RNA binding protein, heterogeneous nuclear ribonucleoprotein D0. Common modular structure and binding properties of the 2xRBD-Gly family. *J. Biol. Chem.* **270**, 22167–75.

[196] Lin, S, Wang, W, Wilson, G. M, Yang, X, Brewer, G, Holbrook, N. J, & Gorospe, M. (2000) Down-regulation of cyclin D1 expression by prostaglandin A(2) is mediated by enhanced cyclin D1 mRNA turnover. *Mol. Cell. Biol.* **20**, 7903–7913.

[197] Pautz, A, Linker, K, Altenhofer, S, Heil, S, Schmidt, N, Art, J, Knauer, S, Stauber, R, Sadri, N, Pont, A, Schneider, R. J, & Kleinert, H. (2009) Similar regulation of human inducible nitric-oxide synthase expression by different isoforms of the RNA-binding protein AUF1. *J. Biol. Chem.* **284**, 2755–66.

[198] Eversole, A & Maizels, N. (2000) *In vitro* properties of the conserved mammalian protein hnRNP D suggest a role in telomere maintenance. *Mol. Cell. Biol.* **20**, 5425–32.

[199] Tsuchiya, N, Kamei, D, Takano, A, Matsui, T, & Yamada, M. (1998) Cloning and characterization of a cDNA encoding a novel heterogeneous nuclear ribonucleoprotein-like protein and its expression in myeloid leukemia cells. *J. Biochem.* **123**, 499–507.

[200] Enokizono, Y, Konishi, Y, Nagata, K, Ouhashi, K, Uesugi, S, Ishikawa, F, & Katahira, M. (2005) Structure of hnRNP D complexed with single-stranded telomere DNA and unfolding of the quadruplex by heterogeneous nuclear ribonucleoprotein D. *J. Biol. Chem.* **280**, 18862–70.

[201] Dempsey, L. A, Hanakahi, L. A, & Maizels, N. (1998) A specific isoform of hnRNP D interacts with DNA in the LR1 heterodimer: canonical RNA binding motifs in a sequence-specific duplex DNA binding protein. *J. Biol. Chem.* **273**, 29224–9.

[202] Takimoto, M, Tomonaga, T, Matunis, M, Avigan, M, Krutzsch, H, Dreyfuss, G, & Levens, D. (1993) Specific binding of heterogeneous ribonucleoprotein particle protein K to the human c-myc promoter, *in vitro*. *J. Biol. Chem.* **268**, 18249–58.

[203] Ostrowski, J, Schullery, D. S, Denisenko, O. N, Higaki, Y, Watts, J, Aebersold, R, Stempka, L, Gschwendt, M, & Bomsztyk, K. (2000) Role of tyrosine phosphorylation in the regulation of the interaction of heterogenous nuclear ribonucleoprotein K protein with its protein and RNA partners. *J. Biol. Chem.* **275**, 3619–28.

[204] Huarte, M, Guttman, M, Feldser, D, Garber, M, Koziol, M. J, Kenzelmann-Broz, D, Khalil, A. M, Zuk, O, Amit, I, Rabani, M, Attardi, L. D, Regev, A, Lander, E. S, Jacks, T, & Rinn, J. L. (2010) A Large Intergenic Noncoding RNA Induced by p53 Mediates Global Gene Repression in the p53 Response. *Cell* **142**, 409–419.

[205] Bomsztyk, K, Denisenko, O, & Ostrowski, J. (2004) hnRNP K: one protein multiple processes. *BioEssays* **26**, 629–38.

[206] Dejgaard, K & Leffers, H. (1996) Characterisation of the nucleic-acid-binding activity of KH domains. Different properties of different domains. *Eur. J. Biochem.* **241**, 425–431.

[207] Matunis, M. J, Michael, W. M, & Dreyfuss, G. (1992) Characterization and primary structure of the poly(C)-binding heterogeneous nuclear ribonucleoprotein complex K protein. *Mol. Cell. Biol.* **12**, 164–71.

[208] Braddock, D. T, Baber, J. L, Levens, D, & Clore, G. M. (2002) Molecular basis of sequence-specific single-stranded DNA recognition by KH domains: solution structure of a complex between hnRNP K KH3 and single-stranded DNA. *EMBO J.* **21**, 3476–85.

[209] Shih, S. C & Claffey, K. P. (1999) Regulation of human vascular endothelial growth factor mRNA stability in hypoxia by heterogeneous nuclear ribonucleoprotein L. *J. Biol. Chem.* **274**, 1359–65.

[210] Hung, L. H, Heiner, M, Hui, J, Schreiner, S, Benes, V, & Bindereif, A. (2008) Diverse roles of hnRNP L in mammalian mRNA processing: a combined microarray and RNAi analysis. *RNA* **14**, 284–96.

[211] Hui, J, Hung, L.-H, Heiner, M, Schreiner, S, Neumller, N, Reither, G, Haas, S. A, & Bindereif, A. (2005) Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. *EMBO J.* **24**, 1988–1998.

[212] Kang, H. J, Kendrick, S, Hecht, S. M, & Hurley, L. H. (2014) The transcriptional complex between the BCL2 i-motif and hnRNP LL is a molecular switch for control of gene expression that can be modulated by small molecules. *J. Am. Chem. Soc.* **136**, 4172–85.

[213] Smith, S. A, Ray, D, Cook, K. B, Mallory, M. J, Hughes, T. R, & Lynch, K. W. (2013) Paralogs hnRNP L and hnRNP LL exhibit overlapping but distinct RNA binding constraints. *PLoS ONE* **8**, e80701.

[214] Fackelmayer, F. O & Richter, A. (1994) Purification of two isoforms of hnRNP-U and characterization of their nucleic acid binding activity. *Biochemistry* **33**, 10416–10422.

[215] Yugami, M, Kabe, Y, Yamaguchi, Y, Wada, T, & Handa, H. (2007) hnRNP-U enhances the expression of specific genes by stabilizing mRNA. *FEBS Lett.* **581**, 1–7.

[216] Hasegawa, Y, Brockdorff, N, Kawano, S, Tsutui, K, Tsutui, K, & Nakagawa, S. (2010) The Matrix Protein hnRNP U Is Required for Chromosomal Localization of Xist RNA. *Dev. Cell* **19**, 469–476.

[217] Hacisuleyman, E, Goff, L. A, Trapnell, C, Williams, A, Henao-Mejia, J, Sun, L, McClanahan, P, Hendrickson, D. G, Sauvageau, M, Kelley, D. R, Morse, M, Engreitz, J, Lander, E. S, Guttman, M, Lodish, H. F, Flavell, R, Raj, A, & Rinn, J. L. (2014) Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat. Struct. Mol. Biol.* **21**, 198–206.

[218] Xiao, R, Tang, P, Yang, B, Huang, J, Zhou, Y, Shao, C, Li, H, Sun, H, Zhang, Y, & Fu, X. D. (2012) Nuclear matrix factor hnRNP U/SAF-A exerts a global control of alternative splicing by regulating U2 snRNP maturation. *Mol. Cell* **45**, 656–68.

[219] Kiledjian, M & Dreyfuss, G. (1992) Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO J.* **11**, 2655–64.

[220] Hasegawa, Y & Nakagawa, S. (2011) Revisiting the function of nuclear scaffold/matrix binding proteins in X chromosome inactivation. *RNA Biol.* **8**, 735–739.

[221] Kukalev, A, Nord, Y, Palmberg, C, Bergman, T, & Percipalle, P. (2005) Actin and hnRNP U cooperate for productive transcription by RNA polymerase II. *Nat. Struct. Mol. Biol.* **12**, 238–44.

[222] Fackelmayer, F. O, Dahm, K, Renz, A, Ramsperger, U, & Richter, A. (1994) Nucleic-acid-binding properties of hnRNP-U/SAF-A, a nuclear-matrix protein which binds DNA and RNA in vivo and in vitro. *Eur. J. Biochem.* **221**, 749–757.

[223] Polo, S, Blackford, A, Chapman, J. , Baskcomb, L, Gravel, S, Rusch, A, Thomas, A, Blundred, R, Smith, P, Kzhyshkowska, J, Dobner, T, Taylor, A. , Turnell, A, Stewart, G, Grand, R, & Jackson, S. (2012) Regulation of DNA-end resection by hnRNPU-like proteins promotes DNA double-strand break signaling and repair. *Mol. Cell* **45**, 505–516.

[224] Guenther, U. P, Handoko, L, Laggerbauer, B, Jablonka, S, Chari, A, Alzheimer, M, Ohmer, J, Plottner, O, Gehring, N, Sickmann, A, von Au, K, Schuelke, M, & Fischer, U. (2009) IGHMBP2 is a ribosome-associated helicase inactive in the neuromuscular disorder distal SMA type 1 (DSMA1). *Hum. Mol. Genet.* **18**, 1288–1300.

[225] de Planell-Saguer, M, Schroeder, D. G, Rodicio, M. C, Cox, G. A, & Mourelatos, Z. (2009) Biochemical and genetic evidence for a role of IGHMBP2 in the translational machinery. *Hum. Mol. Genet.* **18**, 2115–26.

[226] Kerr, D & Khalili, K. (1991) A recombinant cDNA derived from human brain encodes a DNA binding protein that stimulates transcription of the human neurotropic virus JCV. *J. Biol. Chem.* **266**, 15876–15881.

[227] Grohmann, K, Varon, R, Stolz, P, Schuelke, M, Janetzki, C, Bertini, E, Bushby, K, Muntoni, F, Ouvrier, R, Van Maldergem, L, Goemans, N. M. L. A, Lochmller, H, Eichholz, S, Adams, C, Bosch, F, Grattan-Smith, P, Navarro, C, Neitzel, H, Polster, T, Topalolu, H, Steglich, C, Guenther, U. P, Zerres, K, Rudnik-Scheneborn, S, & Hbner, C. (2003) Infantile spinal muscular atrophy with respiratory distress type 1 (SMARD1). *Ann. Neurol.* **54**, 719–724.

[228] Lim, S. C, Bowler, M. W, Lai, T. F, & Song, H. (2012) The Ighmbp2 helicase structure reveals the molecular basis for disease-causing mutations in DMSA1. *Nucleic Acids Res.* **40**, 11009–22.

[229] Satoh, M, Shaheen, V. M, Kao, P. N, Okano, T, Shaw, M, Yoshida, H, Richards, H. B, & Reeves, W. H. (1999) Autoantibodies define a family of proteins with conserved double-stranded RNA-binding domains as well as DNA binding activity. *J. Biol. Chem.* **274**, 34598–34604.

[230] Liao, H. J, Kobayashi, R, & Mathews, M. B. (1998) Activities of adenovirus virus-associated RNAs: purification and characterization of RNA binding proteins. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8514–9.

[231] Langland, J. O, Kao, P. N, & Jacobs, B. L. (1999) Nuclear factor-90 of activated T-cells: A double-stranded RNA-binding protein and substrate for the double-stranded RNA-dependent protein kinase, PKR. *Biochemistry* **38**, 6361–6368.

[232] Masuda, K, Kuwano, Y, Nishida, K, Rokutan, K, & Imoto, I. (2013) NF90 in posttranscriptional gene regulation and microRNA biogenesis. *Int. J. Mol. Sci.* **14**, 17111–17121.

[233] Kuwano, Y, Pullmann, R, Marasa, B. S, Abdelmohsen, K, Lee, E. K, Yang, X, Martindale, J. L, Zhan, M, & Gorospe, M. (2009) NF90 selectively represses the translation of target mRNAs bearing an AU-rich signature motif. *Nucleic Acids Res.* **38**, 225–238.

[234] Wong, G, Mller, O, Clark, R, Conroy, L, Moran, M. F, Polakis, P, & McCormick, F. (1992) Molecular cloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. *Cell* **69**, 551–558.

[235] Tremblay, G. A & Richard, S. (2006) mRNAs associated with the Sam68 RNA binding protein. *RNA Biol.* **3**, 90–3.

[236] Najib, S, Martín-Romero, C, González-Yanes, C, & Sánchez-Margalet, V. (2005) Role of Sam68 as an adaptor protein in signal transduction. *Cell. Mol. Life Sci.* **62**, 36–43.

[237] Taylor, S. J, Anafi, M, Pawson, T, & Shalloway, D. (1995) Functional interaction between c-Src and its mitotic target, Sam 68. *J. Biol. Chem.* **270**, 10120–10124.

[238] Venables, J. P, Vernet, C, Chew, S. L, Elliott, D. J, Cowmeadow, R. B, Wu, J, Cooke, H. J, Artzt, K, & Eperon, I. C. (1999) T-STAR/ETOILE: a novel relative of SAM68 that interacts with an RNA- binding protein implicated in spermatogenesis. *Hum. Mol. Genet.* **8**, 959–969.

[239] Cote, J, Boisvert, F. M, Boulanger, M. C, Bedford, M. T, & Richard, S. (2003) Sam68 RNA binding protein is an in vivo substrate for protein arginine N-methyltransferase 1. *Mol. Biol. Cell* **14**, 274–87.

[240] Angulo, J. F, Rouer, E, Mazin, A, Mattei, M.-g, Tissier, A, Horellou, P, Benarous, R, & Devoret, R. (1991) Identification and expression of the cDNA of KIN17, a zinc-finger gene located on mouse chromosome 2, encoding a new DNA-binding protein. *Nucleic Acids Res.* **19**, 5117–5123.

[241] Pinon-Lataillade, G, Masson, C, Bernardino-Sgherri, J, Henriot, V, Mauffrey, P, Frobert, Y, Araneda, S, & Angulo, J. F. (2004) KIN17 encodes an RNA-binding protein and is expressed during mouse spermatogenesis. *J. Cell Sci.* **117**, 3691–3702.

[242] le Maire, A, Schiltz, M, Stura, E. A, Pinon-Lataillade, G, Couprie, J, Moutiez, M, Gondry, M, Angulo, J. F, & Zinn-Justin, S. (2006) A Tandem of SH3-like Domains Participates in RNA Binding in KIN17, a Human Protein Activated in Response to Genotoxics. *J. Mol. Biol.* **364**, 764–776.

[243] Miccoli, L, Frouin, I, Novac, O, Di Paola, D, Harper, F, Zannis-Hadjopoulos, M, Maga, G, Biard, D. S, & Angulo, J. F. (2005) The human stress-activated protein kin17 belongs to the multiprotein DNA replication complex and associates in vivo with mammalian replication origins. *Mol. Cell. Biol.* **25**, 3814–30.

[244] Masson, C, Menaa, F, Pinon-Lataillade, G, Frobert, Y, Chevillard, S, Radicella, J. P, Sarasin, A, & Angulo, J. F. (2003) Global genome repair is required to activate KIN17, a UVC-responsive gene involved in DNA replication. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 616–621.

[245] Feldmann, H & Winnacker, E. L. (1993) A putative homologue of the human autoantigen Ku from Saccharomyces cerevisiae. *J. Biol. Chem.* **268**, 12895–900.

[246] Martin, S. G, Laroche, T, Suka, N, Grunstein, M, & Gasser, S. M. (1999) Relocalization of telomeric Ku and SIR proteins in response to DNA strand breaks in yeast. *Cell* **97**, 621–33.

[247] Stellwagen, A. E, Haimberger, Z. W, Veatch, J. R, & Gottschling, D. E. (2003) Ku interacts with telomerase RNA to promote telomere addition at native and broken chromosome ends. *Genes Dev.* **17**, 2384–2395.

[248] Walker, J. R, Corpina, R. A, & Goldberg, J. (2001) Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* **412**, 607–614.

[249] Downs, J. A & Jackson, S. P. (2004) A means to a DNA end: the many roles of Ku. *Nat. Rev. Mol. Cell Biol.* **5**, 367–378.

[250] Pfingsten, J, Goodrich, K, Taabazuing, C, Ouenzar, F, Chartrand, P, & Cech, T. (2012) Mutually Exclusive Binding of Telomerase RNA and DNA by Ku Alters Telomerase Recruitment Model. *Cell* **148**, 922–932.

[251] Calissano, P, Volont, C, Biocca, S, & Cattaneo, A. (1985) Synthesis and content of a DNA-binding protein with lactic dehydrogenase activity are reduced by nerve growth factor in the neoplastic cell line PC12. *Exp. Cell Res.* **161**, 117–129.

[252] Cattaneo, A, Biocca, S, Corvaja, N, & Calissano, P. (1985) Nuclear localization of a lactic dehydrogenase with single-stranded DNA-binding properties. *Exp. Cell Res.* **161**, 130–140.

[253] Pioli, P. A, Hamilton, B. J, Connolly, J. E, Brewer, G, & Rigby, W. F. (2002) Lactate dehydrogenase is an AU-rich element-binding protein that directly interacts with AUF1. *J. Biol. Chem.* **277**, 35738–35745.

[254] Le, A, Cooper, C. R, Gouw, A. M, Dinavahi, R, Maitra, A, Deck, L. M, Royer, R. E, Vander Jagt, D. L, Semenza, G. L, & Dang, C. V. (2010) Inhibition of lactate dehydrogenase A induces oxidative stress and inhibits tumor progression. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2037–2042.

[255] Chung, F. Z, Tsujibo, H, Bhattacharyya, U, Sharief, F. S, & Li, S. S. (1985) Genomic organization of human lactate dehydrogenase-A gene. *Biochem. J.* **231**, 537–41.

[256] Mayr, F, Schutz, A, Doge, N, & Heinemann, U. (2012) The Lin28 cold-shock domain remodels pre-let-7 microRNA. *Nucleic Acids Res.* **40**, 7492–506.

[257] Viswanathan, S. R, Daley, G. Q, & Gregory, R. I. (2008) Selective blockade of microRNA processing by Lin28. *Science* **320**, 97–100.

[258] Qiu, C, Ma, Y, Wang, J, Peng, S, & Huang, Y. (2010) Lin28-mediated post-transcriptional regulation of Oct4 expression in human embryonic stem cells. *Nucleic Acids Res.* **38**, 1240–8.

[259] Huang, Y. (2012) A mirror of two faces: Lin28 as a master regulator of both miRNA and mRNA. *Wiley Interdiscip. Rev. RNA* **3**, 483–494.

[260] Zhu, H, Shyh-Chang, N, Segr, A, Shinoda, G, Shah, S, Einhorn, W, Takeuchi, A, Engreitz, J, Hagan, J, Kharas, M, Urbach, A, Thornton, J, Triboulet, R, Gregory, R, Altshuler, D, & Daley, G. (2011) The Lin28/let-7 axis regulates glucose metabolism. *Cell* **147**, 81–94.

[261] Piskounova, E, Viswanathan, S. R, Janas, M, LaPierre, R. J, Daley, G. Q, Sliz, P, & Gregory, R. I. (2008) Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28. *J. Biol. Chem.* **283**, 21310–4.

[262] Nam, Y, Chen, C, Gregory, R. I, Chou, J. J, & Sliz, P. (2011) Molecular basis for interaction of let-7 microRNAs with Lin28. *Cell* **147**, 1080–91.

[263] Liu, T, Lu, B, Lee, I, Ondrovicova, G, Kutejova, E, & Suzuki, C. K. (2004) DNA and RNA binding by the mitochondrial lon protease is regulated by nucleotide and protein substrate. *J. Biol. Chem.* **279**, 13902–10.

[264] Chen, S. H, Suzuki, C. K, & Wu, S. H. (2008) Thermodynamic characterization of specific interactions between the human Lon protease and G-quartet DNA. *Nucleic Acids Res.* **36**, 1273–87.

[265] Wang, N, Gottesman, S, Willingham, M. C, Gottesman, M. M, & Maurizi, M. R. (1993) A human mitochondrial ATP-dependent protease that is highly homologous to bacterial Lon protease. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 11247–51.

[266] Lu, B, Yadav, S, Shah, P. G, Liu, T, Tian, B, Pukszta, S, Villaluna, N, Kutejova, E, Newlon, C. S, Santos, J. H, & Suzuki, C. K. (2007) Roles for the human ATP-dependent Lon protease in mitochondrial DNA maintenance. *J. Biol. Chem.* **282**, 17363–74.

[267] Fu, G. K & Markovitz, D. M. (1998) The human LON protease binds to mitochondrial promoters in a single-stranded, site-specific, strand-specific manner. *Biochemistry* **37**, 1905–9.

[268] Liu, T, B., L, Ondrovicov, G, Kutejov, E, & Suzuki, C. K. (2004) DNA and RNA binding by the mitochondrial lon protease is regulated by nucleotide and protein substrate. *J. Biol. Chem.* **279**, 13902–13910.

[269] Tsuchiya, N, Fukuda, H, Sugimura, T, Nagao, M, & Nakagama, H. (2002) LRP130, a protein containing nine pentatricopeptide repeat motifs, interacts with a single-stranded cytosine-rich sequence of mouse hypervariable minisatellite Pc-1. *Eur. J. Biochem.* **269**, 2927–33.

[270] Mili, S & Pinol-Roma, S. (2003) LRP130, a pentatricopeptide motif protein with a noncanonical RNA-binding domain, is bound in vivo to mitochondrial and nuclear RNAs. *Mol. Cell. Biol.* **23**, 4972–4982.

[271] Tsuchiya, N, Fukuda, H, Nakashima, K, Nagao, M, Sugimura, T, & Nakagama, H. (2004) LRP130, a single-stranded DNA/RNA-binding protein, localizes at the

outer nuclear and endoplasmic reticulum membrane, and interacts with mRNA in vivo. *Biochem. Biophys. Res. Commun.* **317**, 736–743.

[272] Gohil, V. M, Nilsson, R, Belcher-Timme, C. A, Luo, B, Root, D. E, & Mootha, V. K. (2010) Mitochondrial and nuclear genomic responses to loss of LRPPRC expression. *J. Biol. Chem.* **285**, 13742–7.

[273] Xu, F, Morin, C, Mitchell, G, Ackerley, C, & Robinson, B. H. (2004) The role of the LRPPRC (leucine-rich pentatricopeptide repeat cassette) gene in cytochrome oxidase assembly: mutation causes lowered levels of COX (cytochrome c oxidase) I and COX III mRNA. *Biochem. J.* **382**, 331–6.

[274] Liu, L, Sanosaka, M, Lei, S, Bestwick, M. L, Frey, J. H, Surovtseva, Y. V, Shadel, G. S, & Cooper, M. P. (2011) LRP130 protein remodels mitochondria and stimulates fatty acid oxidation. *J. Biol. Chem.* **286**, 41253–41264.

[275] Rikiyama, T, Curtis, J, Oikawa, M, Zimonjic, D. B, Popescu, N, Murphy, B. A, Wilson, M. A, & Johnson, A. C. (2003) GCF2: expression and molecular analysis of repression. *Biochim. Biophys. Acta* **1629**, 15–25.

[276] Wilson, S. A, Brown, E. C, Kingsman, A. J, & Kingsman, S. M. (1998) TRIP: A novel double stranded RNA binding protein which interacts with the leucine rich repeat of Flightless I. *Nucleic Acids Res.* **26**, 3460–3467.

[277] Shi, L, Song, L, Fitzgerald, M, Maurer, K, Bagashev, A, & Sullivan, K. E. (2014) Noncoding RNAs and LRRFIP1 regulate TNF expression. *J. Immunol.* **192**, 3057–67.

[278] Broz, P & Monack, D. M. (2013) Newly described pattern recognition receptors team up against intracellular pathogens. *Nat. Rev. Immunol.* **13**, 551–565.

[279] Khachigian, L. M, Santiago, F. S, Rafty, L. A, Chan, O. L, Delbridge, G. J, Bobik, A, Collins, T, & Johnson, A. C. (1999) GC factor 2 represses platelet-derived growth factor A-chain gene transcription and is itself induced by arterial injury. *Circ. Res.* **84**, 1258–67.

[280] Galvo, T. C & Thomas, J. O. (2005) Structure-specific binding of MeCP2 to four-way junction DNA through its methyl CpG-binding domain. *Nucleic Acids Res.* **33**, 6603–6609.

[281] Melln, M, Ayata, P, Dewell, S, Kriaucionis, S, & Heintz, N. (2012) MeCP2 Binds to 5hmC Enriched within Active Genes and Accessible Chromatin in the Nervous System. *Cell* **151**, 1417–1430.

[282] Jeffery, L & Nakielny, S. (2004) Components of the DNA methylation system of chromatin control are RNA-binding proteins. *J. Biol. Chem.* **279**, 49479–49487.

[283] Baker, S, Chen, L, Wilkins, A, Yu, P, Lichtarge, O, & Zoghbi, H. (2013) An AT-hook domain in MeCP2 determines the clinical course of Rett syndrome and related disorders. *Cell* **152**, 984–996.

[284] Na, E. S, Nelson, E. D, Kavalali, E. T, & Monteggia, L. M. (2013) The impact of MeCP2 loss- or gain-of-function on synaptic plasticity. *Neuropsychopharmacology* **38**, 212–219.

[285] Bedogni, F, Rossi, R. L, Galli, F, Cobolli Gigli, C, Gandaglia, A, Kilstrup-Nielsen, C, & Landsberger, N. (2014) Rett syndrome and the urge of novel approaches to study MeCP2 functions and mechanisms of action. *Neurosci. Biobehav. Rev.*

[286] Ho, K. L, McNae, I. W, Schmiedeberg, L, Klose, R. J, Bird, A. P, & Walkinshaw, M. D. (2008) MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol. Cell* **29**, 525–31.

[287] Kang, H. S, Beak, J. Y, Kim, Y.-S, Petrovich, R. M, Collins, J. B, Grissom, S. F, & Jetten, A. M. (2006) NABP1, a novel ROR-regulated gene encoding a single-stranded nucleic-acid-binding protein. *Biochem. J.* **397**, 89–99.

[288] Huang, J, Gong, Z, Ghosal, G, & Chen, J. (2009) SOSS complexes participate in the maintenance of genomic stability. *Mol. Cell* **35**, 384–393.

[289] Yotov, W. V & St-Arnaud, R. (1996) Differential splicing-in of a proline-rich exon converts alphaNAC into a muscle-specific transcription factor. *Genes Dev.* **10**, 1763–1772.

[290] Beatrix, B, Sakai, H, & Wiedmann, M. (2000) The alpha and beta subunit of the nascent polypeptide-associated complex have distinct functions. *J. Biol. Chem.* **275**, 37838–37845.

[291] Liu, Y, Hu, Y, Li, X, Niu, L, & Teng, M. (2010) The crystal structure of the human nascent polypeptide-associated complex domain reveals a nucleic acid-binding region on the NACA subunit. *Biochemistry* **49**, 2890–2896.

[292] Jafarov, T, Alexander, J. W, & St-Arnaud, R. (2012) alphaNAC interacts with histone deacetylase corepressors to control Myogenin and Osteocalcin gene expression. *Biochim. Biophys. Acta* **1819**, 1208–16.

[293] Wiedmann, B, Sakai, H, Davis, T. A, & Wiedmann, M. (1994) A protein complex required for signal-sequence-specific sorting and translocation. *Nature* **370**, 434–40.

[294] Akhouayri, O, Quelo, I, & St-Arnaud, R. (2005) Sequence-specific DNA binding by the NAC coactivator is required for potentiation of c-Jun-dependent transcription of the osteocalcin gene. *Mol. Cell. Biol.* **25**, 3452–60.

[295] Lv, J, Liu, H, Wang, Q, Tang, Z, Hou, L, & Zhang, B. (2003) Molecular cloning of a novel human gene encoding histone acetyltransferase-like protein involved in transcriptional activation of hTERT. *Biochem. Biophys. Res. Commun.* **311**, 506–513.

[296] Kong, R, Zhang, L, Hu, L, Peng, Q, Han, W, Du, X, & Ke, Y. (2010) hALP, a novel transcriptional U three protein (t-UTP), activates RNA polymerase I transcription by binding and acetylating the upstream binding factor (UBF). *J. Biol. Chem.* **286**, 7139–7148.

[297] Shen, Q, Zheng, X, McNutt, M. A, Guang, L, Sun, Y, Wang, J, Gong, Y, Hou, L, & Zhang, B. (2009) NAT10, a nucleolar protein, localizes to the midbody and regulates cytokinesis and acetylation of microtubules. *Exp. Cell Res.* **315**, 1653–67.

[298] Grinstein, E, Du, Y, Santourlidis, S, Christ, J, Uhrberg, M, & Wernet, P. (2007) Nucleolin regulates gene expression in CD34-positive hematopoietic cells. *J. Biol. Chem.* **282**, 12439–12449.

[299] Westmark, C. J. (2000) Up-regulation of nucleolin mRNA and protein in peripheral blood mononuclear cells by extracellular-regulated kinase. *J. Biol. Chem.* **276**, 1119–1126.

[300] Abdelmohsen, K, Tominaga, K, Lee, E. K, Srikantan, S, Kang, M.-J, Kim, M. M, Selimyan, R, Martindale, J. L, Yang, X, Carrier, F, Zhan, M, Becker, K. G, & Gorospe, M. (2011) Enhanced translation by Nucleolin via G-rich elements in coding and non-coding regions of target mRNAs. *Nucleic Acids Res.* **39**, 8513–8530.

[301] Abdelmohsen, K & Gorospe, M. (2012) RNA-binding protein nucleolin in disease. *RNA Biol.* **9**, 799–808.

[302] Tuteja, N, Huang, N. W, Skopac, D, Tuteja, R, Hrvatic, S, Zhang, J, Pongor, S, Joseph, G, Faucher, C, Amalric, F, & Falaschi, A. (1995) Human DNA helicase IV is nucleolin, an RNA helicase modulated by phosphorylation. *Gene* **160**, 143–148.

[303] Allain, F. H, Bouvet, P, Dieckmann, T, & Feigon, J. (2000) Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO. J.* **19**, 6870–81.

[304] Escalante, C. R, Shen, L, Thanos, D, & Aggarwal, A. K. (2002) Structure of NF-κB p50/p65 heterodimer bound to the PRDII DNA element from the interferon-β promoter. *Structure* **10**, 383–391.

[305] Wurster, S. E & Maher, L. J. (2008) Selection and characterization of anti-NF-κB p65 RNA aptamers. *RNA* **14**, 1037–1047.

[306] Hayden, M. S & Ghosh, S. (2012) NF-κB, the first quarter-century: remarkable progress and outstanding questions. *Genes Dev.* **26**, 203–234.

[307] Huang, D.-B, Vu, D, Cassiday, L. A, Zimmerman, J. M, Maher, L. J, & Ghosh, G. (2003) Crystal structure of NF-κB $(p50)_2$ complexed to a high-affinity RNA aptamer. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9268–9273.

[308] Müller, C. W, Rey, F. A, Sodeoka, M, Verdine, G. L, & Harrison, S. C. (1995) Structure of the NF-κB p50 homodimer bound to DNA. *Nature* **373**, 311–317.

[309] Song, Z, Krishna, S, Thanos, D, Strominger, J. L, & Ono, S. J. (1994) A novel cysteine-rich sequence-specific DNA-binding protein interacts with the conserved X-box motif of the human major histocompatibility complex class II genes via a repeated Cys-His domain and functions as a transcriptional repressor. *J. Exp. Med.* **180**, 1763–74.

[310] Katzenellenbogen, R. A, Vliet-Gregg, P, Xu, M, & Galloway, D. A. (2009) NFX1-123 Increases hTERT Expression and Telomerase Activity Posttranscriptionally in Human Papillomavirus Type 16 E6 Keratinocytes. *J. Virol.* **83**, 6446–6456.

[311] Xu, M, Luo, W, Elzi, D. J, Grandori, C, & Galloway, D. A. (2008) NFX1 interacts with mSin3A/histone deacetylase to repress hTERT transcription in keratinocytes. *Mol. Cell. Biol.* **28**, 4819–28.

[312] Lorick, K. L, Jensen, J. P, Fang, S, Ong, A. M, Hatakeyama, S, & Weissman, A. M. (1999) RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11364–9.

[313] Nardini, M, Gnesutta, N, Donati, G, Gatta, R, Forni, C, Fossati, A, Vonrhein, C, Moras, D, Romier, C, Bolognesi, M, & Mantovani, R. (2013) Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. *Cell* **152**, 132–143.

[314] Imbriano, C, Gnesutta, N, & Mantovani, R. (2012) The NF-Y/p53 liaison: Well beyond repression. *Biochim. Biophys. Acta* **1825**, 131–139.

[315] Ronchi, A, Bellorini, M, Mongelli, N, & Mantovani, R. (1995) CCAAT-box binding protein NF-Y (CBF, CP1) recognizes the minor groove and distorts DNA. *Nucleic Acids Res.* **23**, 4565–72.

[316] Nourbakhsh, M & Hauser, H. (1999) Constitutive silencing of IFN-β promoter is mediated by NRF (NF-κB-repressing factor), a nuclear inhibitor of NF-κB. *EMBO J.* **18**, 6415–6425.

[317] Niedick, I, Froese, N, Oumard, A, P., M. P, Nourbakhsh, M, Hauser, H, & Kster, M. (2004) Nucleolar localization and mobility analysis of the NF-κB repressing factor NRF. *J. Cell Sci.* **117**, 3447–3458.

[318] Dreikhausen, U, Hiebenthal-Millow, K, Bartels, M, Resch, K, & Nourbakhsh, M. (2005) NF-κB-Repressing Factor Inhibits Elongation of Human Immunodeficiency Virus Type 1 Transcription by DRB Sensitivity-Inducing Factor. *Mol. Cell. Biol.* **25**, 7473–7483.

[319] Lee, K. Y, Ho, S. C, Chan, Y. F, Wang, C. H, Huang, C. D, Liu, W. T, Lin, S. M, Lo, Y. L, Chang, Y. L, Kuo, L. W, & Kuo, H. P. (2012) Reduced nuclear factor-κB repressing factor: a link toward systemic inflammation in COPD. *Eur. Respir. J.* **40**, 863–873.

[320] Basu, A, Dong, B, Krainer, A. R, & Howe, C. C. (1997) The intracisternal A-particle proximal enhancer-binding protein activates transcription and is identical to the RNA- and DNA-binding protein p54nrb/NonO. *Mol. Cell. Biol.* **17**, 677–86.

[321] Zhang, Z & Carmichael, G. G. (2001) The Fate of dsRNA in the NucleusA p54nrb-Containing Complex Mediates the Nuclear Retention of Promiscuously A-to-I Edited RNAs. *Cell* **106**, 465–475.

[322] Izumi, H, McCloskey, A, Shinmyozu, K, & Ohno, M. (2014) p54nrb/NonO and PSF promote U snRNA nuclear export by accelerating its export complex assembly. *Nucleic Acids Res.* **42**, 3998–4007.

[323] Nakagawa, S & Hirose, T. (2012) Paraspeckle nuclear bodies- useful uselessness? *Cell. Mol. Life Sci.* **69**, 3027–3036.

[324] Yadav, S. P, Hao, H, Yang, H. J, Kautzmann, M. A. I, Brooks, M, Nellissery, J, Klocke, B, Seifert, M, & Swaroop, A. (2013) The transcription-splicing protein NonO/p54nrb and three NonO-interacting proteins bind to distal enhancer region and augment rhodopsin expression. *Hum. Mol. Genet.* **23**, 2132–2144.

[325] Zazopoulos, E, Lalli, E, Stocco, D. M, & Sassone-Corsi, P. (1997) DNA binding and transcriptional repression by DAX-1 blocks steroidogenesis. *Nature* **390**, 311–5.

[326] Lalli, E, Ohe, K, Hindelang, C, & Sassone-Corsi, P. (2000) Orphan receptor DAX-1 is a shuttling RNA binding protein associated with polyribosomes via mRNA. *Mol. Cell. Biol.* **20**, 4910–4921.

[327] El-Khairi, R, Martinez-Aguayo, A, Ferraz-de Souza, B, Lin, L, & Achermann, J. C. (2011) Role of DAX-1 (NR0B1) and steroidogenic factor-1 (NR5A1) in human adrenal function. *Endocr. Dev.* **20**, 38–46.

[328] Jantzen, H.-M, Strhle, U, Gloss, B, Stewart, F, Schmid, W, Boshart, M, Miksicek, R, & Schtz, G. (1987) Cooperativity of glucocorticoid response elements located far upstream of the tyrosine aminotransferase gene. *Cell* **49**, 29–38.

[329] Dhawan, L, Liu, B, Blaxall, B. C, & Taubman, M. B. (2007) A novel role for the glucocorticoid receptor in the regulation of monocyte chemoattractant protein-1 mRNA stability. *J. Biol. Chem.* **282**, 10146–10152.

[330] Ali, M & Vedeckis, W. (1987) The glucocorticoid receptor protein binds to transfer RNA. *Science* **235**, 467–470.

[331] Freedman, L. P, Luisi, B. F, Korszun, Z. R, Basavappa, R, Sigler, P. B, & Yamamoto, K. R. (1988) The function and structure of the metal coordination sites within the glucocorticoid receptor DNA binding domain. *Nature* **334**, 543–546.

[332] Kadmiel, M & Cidlowski, J. A. (2013) Glucocorticoid receptor signaling in health and disease. *Trends Pharmacol. Sci.* **34**, 518–530.

[333] Hudson, W. H, Youn, C, & Ortlund, E. A. (2013) The structural basis of direct glucocorticoid-mediated transrepression. *Nat. Struct. Mol. Biol.* **20**, 53–8.

[334] Luisi, B. F, Xu, W. X, Otwinowski, Z, Freedman, L. P, Yamamoto, K. R, & Sigler, P. (1991) Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* **352**, 497–505.

[335] Hudson, W. H, Pickard, M. R, de Vera, I. M. S, Mourtada-Maarabouni, M, Kuiper, E. G, Conn, G. L, Kojetin, D. J, Williams, G. T, & Ortlund, E. A. (2014) Conserved, sequence-specific lincRNA-protein interactions drive transcriptional repression and direct cell fate. *Nat. Commun.* **5**.

[336] Kabe, Y, Goto, M, Shima, D, Imai, T, Wada, T, Morohashi, K.-i, Shirakawa, M, Hirose, S, & Handa, H. (1999) The role of human MBF1 as a transcriptional coactivator. *J. Biol. Chem.* **274**, 34196–34202.

[337] Halvorson, L. M, Ito, M, Jameson, J. L, & Chin, W. W. (1998) Steroidogenic factor-1 and early growth response protein 1 act through two composite DNA binding sites to regulate luteinizing hormone β-subunit gene expression. *J. Biol. Chem.* **273**, 14712–14720.

[338] Wang, C.-Y, Chen, W.-Y, Lai, P.-Y, & Chung, B.-C. (2013) Distinct functions of steroidogenic factor-1 (NR5A1) in the nucleus and the centrosome. *Mol. Cell. Endocrinol.* **371**, 148–153.

[339] Little, T. H, Zhang, Y, Matulis, C. K, Weck, J, Zhang, Z, Ramachandran, A, Mayo, K. E, & Radhakrishnan, I. (2006) Sequence-specific deoxyribonucleic acid (DNA) recognition by steroidogenic factor 1: a helix at the carboxy terminus of the DNA binding domain is necessary for complex stability. *Mol. Endocrinol.* **20**, 831–43.

[340] Wilson, T. E, Fahrner, T. J, & Milbrandt, J. (1993) The orphan receptors NGFI-B and steroidogenic factor 1 establish monomer binding as a third paradigm of nuclear receptor-DNA interaction. *Mol. Cell. Biol.* **13**, 5794–804.

[341] Sakita-Suto, S, Kanda, A, Suzuki, F, Sato, S, Takata, T, & Tatsuka, M. (2007) Aurora-B regulates RNA methyltransferase NSUN2. *Mol. Biol. Cell* **18**, 1107–1117.

[342] Hussain, S, Benavente, S. B, Nascimento, E, Dragoni, I, Kurowski, A, Gillich, A, Humphreys, P, & Frye, M. (2009) The nucleolar RNA methyltransferase Misu (NSun2) is required for mitotic spindle stability. *J. Cell Biol.* **186**, 27–40.

[343] Brzezicha, B, Schmidt, M, Makalowska, I, Jarmolowski, A, Pienkowska, J, & Szweykowska-Kulinska, Z. (2006) Identification of human tRNA:m5C methyltransferase catalysing intron-dependent m5C formation in the first position of the anticodon of the Formula. *Nucleic Acids Res.* **34**, 6034–6043.

[344] Hussain, S, Sajini, A, Blanco, S, Dietmann, S, Lombard, P, Sugimoto, Y, Paramor, M, Gleeson, J, Odom, D, Ule, J, & Frye, M. (2013) NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. *Cell Rep.* **4**, 255–261.

[345] Tuorto, F, Liebers, R, Musch, T, Schaefer, M, Hofmann, S, Kellner, S, Frye, M, Helm, M, Stoecklin, G, & Lyko, F. (2012) RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nat. Struct. Mol. Biol.* **19**, 900–905.

[346] Cabart, P, Chew, H. K, & Murphy, S. (2004) BRCA1 cooperates with NUFIP and P-TEFb to activate transcription by RNA polymerase II. *Oncogene* **23**, 5316–5329.

[347] Bardoni, B, Schenck, A, & Louis Mandel, J. (1999) A novel RNA-binding nuclear protein that interacts with the fragile X mental retardation (FMR1) protein. *Hum. Mol. Genet.* **8**, 2557–2566.

[348] McKeegan, K. S, Debieux, C. M, & Watkins, N. J. (2009) Evidence that the AAA+ proteins TIP48 and TIP49 bridge interactions between 15.5K and the re-

lated NOP56 and NOP58 proteins during box C/D snoRNP biogenesis. *Mol. Cell. Biol.* **29**, 4971–4981.

[349] Huambachano, O, Herrera, F, Rancourt, A, & Satoh, M. S. (2011) Double-stranded DNA binding domain of poly(ADP-ribose) polymerase-1 and molecular insight into the regulation of its activity. *J. Biol. Chem.* **286**, 7149–7160.

[350] Lavrik, O. I, Prasad, R, Sobol, R. W, Horton, J. K, Ackerman, E. J, & Wilson, S. H. (2001) Photoaffinity labeling of mouse fibroblast enzymes by a base excision repair intermediate. Evidence for the role of poly(ADP-ribose) polymerase-1 in DNA repair. *J. Biol. Chem.* **276**, 25541–8.

[351] Mangerich, A & Bürkle, A. (2012) Pleiotropic cellular functions of PARP1 in longevity and aging: genome maintenance meets inflammation. *Oxid. Med. Cell. Longev.* **2012**, 19.

[352] Pines, A, Mullenders, L. H, van Attikum, H, & Luijsterburg, M. S. (2013) Touching base with PARPs: moonlighting in the repair of UV lesions and double-strand breaks. *Trends Biochem. Sci.* **38**, 321–330.

[353] Hassler, M & Ladurner, A. G. (2012) Towards a structural understanding of PARP1 activation and related signalling ADP-ribosyl-transferases. *Curr. Opin. Struct. Biol.* **22**, 721–729.

[354] Langelier, M. F, Planck, J. L, Roy, S, & Pascal, J. M. (2011) Crystal structures of poly(ADP-ribose) polymerase-1 (PARP-1) zinc fingers bound to DNA: structural and functional insights into DNA-dependent PARP-1 activity. *J. Biol. Chem.* **286**, 10690–701.

[355] Aasheim, H.-C, Loukianova, T, Deggerdal, A, & Smeland, E. B. (1994) Tissue specific expression and cDNA structure of a human transcript encoding a nucleic acid

binding [oligo(dC)] protein related to the pre-mRNA binding protein K. *Nucleic Acids Res.* **22**, 959–964.

[356] Leffers, H, Dejgaard, K, & Celis, J. E. (1995) Characterisation of two major cellular poly(rC)-binding human proteins, each containing three K-homologous (KH) domains. *Eur. J. Biochem.* **230**, 447–453.

[357] Choi, H. S, Hwang, C. K, Song, K. Y, Law, P.-Y, Wei, L.-N, & Loh, H. H. (2009) Poly(C)-binding proteins as transcriptional regulators of gene expression. *Biochem. Biophys. Res. Commun.* **380**, 431–436.

[358] Du, Z, Lee, J. K, Tjhen, R, Li, S, Pan, H, Stroud, R. M, & James, T. L. (2005) Crystal structure of the first KH domain of human poly(C)-binding protein-2 in complex with a C-rich strand of human telomeric DNA at 1.7 . *J. Biol. Chem.* **280**, 38823–38830.

[359] Du, Z, Lee, J. K, Fenn, S, Tjhen, R, Stroud, R. M, & James, T. L. (2007) X-ray crystallographic and NMR studies of protein-protein and protein-nucleic acid interactions involving the KH domains from human poly(C)-binding protein-2. *RNA* **13**, 1043–51.

[360] Jindal, H. K & Vishwanatha, J. K. (1990) Functional identity of a primer recognition protein as phosphoglycerate kinase. *J. Biol. Chem.* **265**, 6540–6543.

[361] Shetty, S, Muniyappa, H, Halady, P. K. S, & Idell, S. (2004) Regulation of urokinase receptor expression by phosphoglycerate kinase. *Am. J. Respir. Cell Mol. Biol.* **31**, 100–106.

[362] Michelson, A. M, Markham, A. F, & Orkin, S. H. (1983) Isolation and DNA sequence of a full-length cDNA clone for human X chromosome-encoded phosphoglycerate kinase. *Proc. Natl. Acad. Sci. U.S.A.* **80**, 472–6.

[363] Yin, S, Fan, X, Yu, S, Li, C, & Sy, M. S. (2008) Binding of recombinant but not endogenous prion protein to DNA causes DNA internalization and expression in mammalian cells. *J. Biol. Chem.* **283**, 25446–25454.

[364] Cordeiro, Y, Machado, F, Juliano, L, Juliano, M. A, Brentani, R. R, Foguel, D, & Silva, J. L. (2001) DNA converts cellular prion protein into the beta-sheet conformation and inhibits prion peptide aggregation. *J. Biol. Chem.* **276**, 49400–9.

[365] Gabus, C. (2001) The Prion Protein Has RNA Binding and Chaperoning Properties Characteristic of Nucleocapsid Protein NCp7 of HIV-1. *J. Biol. Chem.* **276**, 19301–19309.

[366] Lima, L. M. T. R, Cordeiro, Y, Tinoco, L. W, Marques, A. F, Oliveira, C. L. P, Sampath, S, Kodali, R, Choi, G, Foguel, D, Torriani, I, Caughey, B, & Silva, J. L. (2006) Structural insights into the interaction between prion protein and nucleic acid. *Biochemistry* **45**, 9180–9187.

[367] Biasini, E, Turnbaugh, J. A, Unterberger, U, & Harris, D. A. (2012) Prion protein at the crossroads of physiology and disease. *Trends Neurosci.* **35**, 92–103.

[368] Deleault, N. R, Lucassen, R. W, & Supattapone, S. (2003) RNA molecules stimulate prion protein conversion. *Nature* **425**, 717–20.

[369] Brunel, F, Zakin, M. M, Buc, H, & Buckle, M. (1996) The polypyrimidine tract binding (PTB) protein interacts with single-stranded DNA in a sequence-specific manner. *Nucleic Acids Res.* **24**, 1608–1615.

[370] Lin, J. C & Tarn, W. Y. (2005) Exon selection in α-tropomyosin mRNA is regulated by the antagonistic action of RBM4 and PTB. *Mol. Cell. Biol.* **25**, 10111–10121.

[371] Keppetipola, N, Sharma, S, Li, Q, & Black, D. L. (2012) Neuronal regulation of pre-mRNA splicing by polypyrimidine tract binding proteins, PTBP1 and PTBP2. *Crit. Rev. Biochem. Mol. Biol.* **47**, 360–378.

[372] Cheung, H. C, Hai, T, Zhu, W, Baggerly, K. A, Tsavachidis, S, Krahe, R, & Cote, G. J. (2009) Splicing factors PTBP1 and PTBP2 promote proliferation and migration of glioma cell lines. *Brain* **132**, 2277–2288.

[373] Oberstrass, F. C, Auweter, S. D, Erat, M, Hargous, Y, Henning, A, Wenter, P, Reymond, L, Amir-Ahmady, B, Pitsch, S, Black, D. L, & Allain, F. H. (2005) Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science* **309**, 2054–7.

[374] Bergemann, A. D, Ma, Z. W, & Johnson, E. M. (1992) Sequence of cDNA comprising the human pur gene and sequence-specific single-stranded-DNA-binding properties of the encoded protein. *Mol. Cell. Biol.* **12**, 5673–5682.

[375] Wortman, M. J, Johnson, E. M, & Bergemann, A. D. (2005) Mechanism of DNA binding and localized strand separation by Pur and comparison with Pur family member, Pur. *Biochim. Biophys. Acta* **1743**, 64–78.

[376] Graebsch, A, Roche, S, & Niessing, D. (2009) X-ray structure of Pur- reveals a Whirly-like fold and an unusual nucleic-acid binding surface. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 18521–18526.

[377] Gupta, M, Sueblinvong, V, Raman, J, Jeevanandam, V, & Gupta, M. P. (2003) Single-stranded DNA-binding proteins PUR and PUR bind to a purine-rich negative regulatory element of the alpha-myosin heavy chain gene and control transcriptional and translational regulation of the gene expression: Implications in the repression of alpha-myosin heavy chain during heart failure. *J. Biol. Chem.* **278**, 44935–44948.

[378] Kovalenko, O. V, Golub, E. I, Bray-Ward, P, Ward, D. C, & Radding, C. M. (1997) A novel nucleic acid-binding protein that interacts with human rad51 recombinase. *Nucleic Acids Res.* **25**, 4946–4953.

[379] Wiese, C, Dray, E, Groesser, T, San Filippo, J, Shi, I, Collins, D. W, Tsai, M.-S, Williams, G. J, Rydberg, B, Sung, P, & Schild, D. (2007) Promotion of homologous recombination and genomic stability by RAD51AP1 via RAD51 recombinase enhancement. *Mol. Cell* **28**, 482–490.

[380] Dunlop, M. H, Dray, E, Zhao, W, San Filippo, J, Tsai, M. S, Leung, S. G, Schild, D, Wiese, C, & Sung, P. (2012) Mechanistic insights into RAD51-associated protein 1 (RAD51AP1) action in homologous DNA repair. *J. Biol. Chem.* **287**, 12343–12347.

[381] Modesti, M, Budzowska, M, Baldeyron, C, Demmers, J. A. A, Ghirlando, R, & Kanaar, R. (2007) RAD51AP1 is a structure-specific DNA binding protein that stimulates joint molecule formation during RAD51-mediated homologous recombination. *Mol. Cell* **28**, 468–481.

[382] Rastinejad, F, Wagner, T, Zhao, Q, & Khorasanizadeh, S. (2000) Structure of the RXR-RAR DNA-binding complex on the retinoic acid response element DR1. *EMBO J.* **19**, 1045–54.

[383] Poon, M. M & Chen, L. (2008) Retinoic acid-gated sequence-specific translational control by RAR. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20303–20308.

[384] Tang, X.-H & Gudas, L. J. (2011) Retinoids, retinoic acid receptors, and cancer. *Annu. Rev. Pathol.* **6**, 345–364.

[385] Rhinn, M & Dolle, P. (2012) Retinoic acid signalling during development. *Development* **139**, 843–858.

[386] Wright, C. F, Oswald, B, & Dellis, S. (2001) Vaccinia virus late transcription is activated in vitro by cellular heterogeneous nuclear ribonucleoproteins. *J. Biol. Chem.* **276**, 40680–40686.

[387] Gerber, A. P, Pilotte, J, Dupont-Versteegden, E. E, & Vanderklish, P. W. (2011) Widespread regulation of miRNA biogenesis at the Dicer step by the cold-inducible RNA-binding protein, RBM3. *PLoS ONE* **6**, e28446.

[388] Liu, Y, Hu, W, Murakawa, Y, Yin, J, Wang, G, Landthaler, M, & Yan, J. (2013) Cold-induced RNA-binding proteins regulate circadian gene expression by controlling alternative polyadenylation. *Sci. Rep.* **3**, 2054.

[389] Lleonart, M. E. (2010) A new generation of proto-oncogenes: Cold-inducible RNA binding proteins. *Biochim. Biophys. Acta* **1805**, 43–52.

[390] Sutherland, L. C, Rintala-Maki, N. D, White, R. D, & Morin, C. D. (2005) RNA binding motif (RBM) proteins: a novel family of apoptosis modulators? *J. Cell. Biochem.* **94**, 5–24.

[391] Timmer, T, Terpstra, P, van den Berg, A, Veldhuis, P. M, Ter Elst, A, Voutsinas, G, Hulsbeek, M. M, Draaijers, T. G, Looman, M. W, Kok, K, Naylor, S. L, & Buys, C. H. (1999) A comparison of genomic structures and expression patterns of two closely related flanking genes in a critical lung cancer region at 3p21.3. *Eur. J. Hum. Genet.* **7**, 478–86.

[392] Shu, L, Yan, W, & Chen, X. (2006) RNPC1, an RNA-binding protein and a target of the p53 family, is required for maintaining the stability of the basal and stress-induced p21 transcript. *Genes Dev.* **20**, 2961–72.

[393] Negishi, Y, Nishita, Y, Sagusa, Y, Kakizaki, I, Galli, I, Kihara, F, Tamai, K, Miyajima, N, Iguchi-Ariga, S. M, & Ariga, H. (1994) Identification and cDNA

cloning of single-stranded DNA binding proteins that interact with the region upstream of the human c-myc gene. *Oncogene* **9**, 1133–1143.

[394] Fritz, D & Stefanovic, B. (2007) RNA-binding protein RBMS3 is expressed in activated hepatic stellate cells and liver fibrosis and increases expression of transcription factor Prx1. *J. Mol. Biol.* **371**, 585–595.

[395] Penkov, D, Ni, R, Else, C, Piol-Roma, S, Ramirez, F, & Tanaka, S. (2000) Cloning of a human gene closely related to the genes coding for the c-myc single-strand binding proteins. *Gene* **243**, 27–36.

[396] Kimura, K, Saga, H, Hayashi, K, Obata, H, Chimori, Y, Ariga, H, & Sobue, K. (1998) c-Myc gene single-strand binding protein-1, MSSP-1, suppresses transcription of alpha-smooth muscle actin gene in chicken visceral smooth muscle cells. *Nucleic Acids Res.* **26**, 2420–5.

[397] Hemmerich, P, Bosbach, S, Mikecz, A, & Krawinkel, U. (1997) Human ribosomal protein L7 binds RNA with an alpha-helical arginine-rich and lysine-rich domain. *Eur. J. Biochem.* **245**, 549–556.

[398] Neumann, F, Hemmerich, P, von Mikecz, A, Peter, H. H, & Krawinkel, U. (1995) Human ribosomal protein L7 inhibits cell-free translation in reticulocyte lysates and affects the expression of nuclear proteins upon stable transfection into Jurkat T-lymphoma cells. *Nucleic Acids Res.* **23**, 195–202.

[399] Neumann, F & Krawinkel, U. (1997) Constitutive expression of human ribosomal protein L7 arrests the cell cycle in G1 and induces apoptosis in Jurkat T-lymphoma cells. *Exp. Cell Res.* **230**, 252–61.

[400] Morita, T, Sato, T, Nyunoya, H, Tsujimoto, A, Takahara, J, Irino, S, & Shimotohno, K. (1993) Isolation of a cDNA clone encoding DNA-binding protein

(TAXREB107) that binds specifically to domain C of the tax-responsive enhancer element in the long terminal repeat of human T-cell leukemia virus type I. *AIDS Res. Hum. Retroviruses* **9**, 115–121.

[401] Haeusler, A. R, Donnelly, C. J, Periz, G, Simko, E. A, Shaw, P. G, Kim, M. S, Maragakis, N. J, Troncoso, J. C, Pandey, A, Sattler, R, Rothstein, J. D, & Wang, J. (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507**, 195–200.

[402] de Jong, R. N, Truffault, V, Diercks, T, Ab, E, Daniels, M. A, Kaptein, R, & Folkers, G. E. (2008) Structure and DNA binding of the human Rtf1 Plus3 domain. *Structure* **16**, 149–59.

[403] Eydmann, T, Sommariva, E, Inagawa, T, Mian, S, Klar, A. J, & Dalgaard, J. Z. (2008) Rtf1-mediated eukaryotic site-specific replication termination. *Genetics* **180**, 27–39.

[404] Dermody, J. L & Buratowski, S. (2010) Leo1 subunit of the yeast paf1 complex binds RNA and contributes to complex recruitment. *J. Biol. Chem.* **285**, 33671–33679.

[405] Piro, A. S, Mayekar, M. K, Warner, M. H, Davis, C. P, & Arndt, K. M. (2012) Small region of Rtf1 protein can substitute for complete Paf1 complex in facilitating global histone H2B ubiquitylation in yeast. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 10837–10842.

[406] Squazzo, S. L, Costa, P. J, Lindstrom, D, Kumer, K. E, Simic, R, Jennings, J. L, Link, A. J, Ardnt, K. M, & Hartzog, G. A. (2002) The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. *EMBO J.* **21**, 1764–1774.

[407] Kaur, G, Jalagadugula, G, Mao, G, & Rao, A. K. (2010) RUNX1/core binding factor A2 regulates platelet 12-lipoxygenase gene (ALOX12): studies in human RUNX1 haplodeficiency. *Blood* **115**, 3128–35.

[408] Fukunaga, J, Nomura, Y, Tanaka, Y, Amano, R, Tanaka, T, Nakamura, Y, Kawai, G, Sakamoto, T, & Kozu, T. (2013) The Runt domain of AML1 (RUNX1) binds a sequence-conserved RNA motif that mimics a DNA element. *RNA* **19**, 927–936.

[409] Barton, J. L, Bunka, D. H, Knowling, S. E, Lefevre, P, Warren, A. J, Bonifer, C, & Stockley, P. G. (2009) Characterization of RNA aptamers that disrupt the RUNX1-CNF/DNA complex. *Nucleic Acids Res.* **37**, 6818–30.

[410] Kagoshima, H, Shigesada, K, Satake, M, Ito, Y, Miyoshi, H, Ohki, M, Pepling, M, & Gergen, P. (1993) The runt domain identifies a new family of heterometric transcriptional regulators. *Trends Genet.* **9**, 338–341.

[411] Ichikawa, M, Yoshimi, A, Nakagawa, M, Nishimoto, N, Watanabe-Okochi, N, & Kurokawa, M. (2013) A role for RUNX1 in hematopoiesis and myeloid leukemia. *Int. J. Hematol.* **97**, 726–34.

[412] Lutterbach, B, Westendorf, J. J, Linggi, B, Isaac, S, Seto, E, & Hiebert, S. W. (2000) A mechanism of repression by acute myeloid leukemia-1, the target of multiple chromosomal translocations in acute leukemia. *J. Biol. Chem.* **275**, 651–6.

[413] Tahirov, T. H, Inoue-Bungo, T, Morii, H, Fujikawa, A, Sasaki, M, Kimura, K, Shiina, M, Sato, K, Kumasaka, T, Yamamoto, M, Ishii, S, & Ogata, K. (2001) Structural analyses of DNA recognition by the AML1/Runx-1 Runt domain and its allosteric control by CBFβ. *Cell* **104**, 755–67.

[414] Matias, P. M, Gorynia, S, Donner, P, & Carrondo, M. A. (2006) Crystal structure of the human AAA+ protein RuvBL1. *J. Biol. Chem.* **281**, 38918–38929.

[415] Jha, S & Dutta, A. (2009) RVB1/RVB2: Running rings around molecular biology. *Mol. Cell* **34**, 521–533.

[416] Izumi, N, Yamashita, A, & Ohno, S. (2012) Integrated regulation of PIKK-mediated stress responses by AAA+ proteins RUVBL1 and RUVBL2. *Nucleus* **3**, 29–43.

[417] Tanaka, T, Tanaka, K, Ogawa, S, Kurokawa, M, Mitani, K, Nishida, J, Shibata, Y, Yazaki, Y, & Hirai, H. (1995) An acute myeloid leukemia gene, AML1, regulates hemopoietic myeloid cell differentiation and transcriptional activation antagonistically by two alternative spliced forms. *EMBO J.* **14**, 341–50.

[418] Papin, C, Humbert, O, Kalashnikova, A, Eckert, K, Morera, S, Ks, E, & Grigoriev, M. (2010) 3'- to 5' DNA unwinding by TIP49b proteins. *FEBS J.* **277**, 2705–2714.

[419] Newman, D. R, Kuhn, J. F, Shanab, G. M, & Maxwell, E. S. (2000) Box C/D snoRNA-associated proteins: two pairs of evolutionarily ancient proteins and possible links to replication and transcription. *RNA* **6**, 861–79.

[420] Gorynia, S, Bandeiras, T. M, Pinho, F. G, McVey, C. E, Vonrhein, C, Round, A, Svergun, D. I, Donner, P, Matias, P. M, & Carrondo, M. A. (2011) Structural and functional insights into a dodecameric molecular machine The RuvBL1/RuvBL2 complex. *J. Struct. Biol.* **176**, 279–291.

[421] Townson, S. M, Kang, K, Lee, A. V, & Oesterreich, S. (2004) Structure-function analysis of the estrogen receptor corepressor scaffold attachment factor-B1: identification of a potent transcriptional repression domain. *J. Biol. Chem.* **279**, 26074–26081.

[422] Kipp, M, Gohring, F, Ostendorp, T, van Drunen, C. M, van Driel, R, Przybylski, M, & Fackelmayer, F. O. (2000) SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA. *Mol. Cell. Biol.* **20**, 7480–9.

[423] Oesterreich, S, Zhang, Q, Hopp, T, Fuqua, S. A, Michaelis, M, Zhao, H. H, Davie, J. R, Osborne, C. K, & Lee, A. V. (2000) Tamoxifen-bound estrogen receptor (ER) strongly interacts with the nuclear matrix protein HET/SAF-B, a novel inhibitor of ER-mediated transactivation. *Mol. Endocrinol.* **14**, 369–81.

[424] Oesterreich, S. (2003) Scaffold attachment factors SAFB1 and SAFB2: Innocent bystanders or critical players in breast tumorigenesis? *J. Cell. Biochem.* **90**, 653–661.

[425] Akhmedov, A. T & Lopez, B. S. (2000) Human 100-kDa homologous DNA-pairing protein is the splicing factor PSF and promotes DNA strand invasion. *Nucleic Acids Res.* **28**, 3022–3030.

[426] Patton, J. G, Porro, E. B, Galceran, J, Tempst, P, & Nadal-Ginard, B. (1993) Cloning and characterization of PSF, a novel pre-mRNA splicing factor. *Genes Dev.* **7**, 393–406.

[427] Mathur, M, Tucker, P. W, & Samuels, H. H. (2001) PSF is a novel corepressor that mediates its effect through Sin3A and the DNA binding domain of nuclear hormone receptors. *Mol. Cell. Biol.* **21**, 2298–311.

[428] Landeras-Bueno, S, Jorba, N, Perez-Cidoncha, M, & Ortin, J. (2011) The splicing factor proline-glutamine rich (SFPQ/PSF) is involved in influenza virus transcription. *PLoS Pathog.* **7**, e1002397.

[429] Straub, T, Grue, P, Uhse, A, Lisby, M, Knudsen, B. R, Tange, T. O, Westergaard, O, & Boege, F. (1998) The RNA-splicing factor PSF/p54 controls DNA-topoisomerase I activity by a direct interaction. *J. Biol. Chem.* **273**, 26261–4.

[430] BabuRajendran, N, Palasingam, P, Narasimhan, K, Sun, W, Prabhakar, S, Jauch, R, & Kolatkar, P. R. (2010) Structure of Smad1 MH1/DNA complex reveals

distinctive rearrangements of BMP and TGF-β effectors. *Nucleic Acids Res.* **38**, 3477–3488.

[431] Davis, B. N, Hilyard, A. C, Lagna, G, & Hata, A. (2008) SMAD proteins control DROSHA-mediated microRNA maturation. *Nature* **454**, 56–61.

[432] Morikawa, M, Koinuma, D, Miyazono, K, & Heldin, C. H. (2012) Genome-wide mechanisms of Smad binding. *Oncogene* **32**, 1609–1615.

[433] Lorson, C. (1998) The domain encoded by exon 2 of the survival motor neuron protein mediates nucleic acid binding. *Hum. Mol. Genet.* **7**, 1269–1275.

[434] Anderton, R. S, Meloni, B. P, Mastaglia, F. L, & Boulos, S. (2013) Spinal muscular atrophy and the antiapoptotic role of survival of motor neuron (SMN) protein. *Mol. Neurobiol.* **47**, 821–32.

[435] Fallini, C, Bassell, G. J, & Rossoll, W. (2012) Spinal muscular atrophy: The role of SMN in axonal mRNA regulation. *Brain Res.* **1462**, 81–92.

[436] Sun, C. T, Lo, W. Y, Wang, I. H, Lo, Y. H, Shiou, S. R, Lai, C. K, & Ting, L. P. (2001) Transcription repression of human hepatitis B virus genes by negative regulatory element-binding protein/SON. *J. Biol. Chem.* **276**, 24059–24067.

[437] Ahn, E.-Y, DeKelver, R, Lo, M.-C, Nguyen, T, Matsuura, S, Boyapati, A, Pandit, S, Fu, X.-D, & Zhang, D.-E. (2011) SON controls cell-cycle progression by coordinated regulation of RNA splicing. *Mol. Cell* **42**, 185–198.

[438] Lu, X, Goke, J, Sachs, F, Jacques, P. E, Liang, H, Feng, B, Bourque, G, Bubulya, P. A, & Ng, H. H. (2013) SON connects the splicing-regulatory network with pluripotency in human embryonic stem cells. *Nat. Cell Biol.* **15**, 1141–52.

[439] Reményi, A. (2003) Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev.* **17**, 2048–2059.

[440] Ng, S.-Y, Bogu, G. K, Soh, B. S, & Stanton, L. W. (2013) The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis. *Mol. Cell* **51**, 349–359.

[441] Ng, S.-Y, Johnson, R, & Stanton, L. W. (2012) Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J.* **31**, 522–533.

[442] Masui, S, Nakatake, Y, Toyooka, Y, Shimosato, D, Yagi, R, Takahashi, K, Okochi, H, Okuda, A, Matoba, R, Sharov, A. A, Ko, M. S. H, & Niwa, H. (2007) Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat. Cell Biol.* **9**, 625–635.

[443] Wei, F, Scholer, H. R, & Atchison, M. L. (2007) Sumoylation of Oct4 enhances its stability, DNA binding, and transactivation. *J. Biol. Chem.* **282**, 21551–60.

[444] Tsuruzoe, S, Ishihara, K, Uchimura, Y, Watanabe, S, Sekita, Y, Aoto, T, Saitoh, H, Yuasa, Y, Niwa, H, Kawasuji, M, Baba, H, & Nakao, M. (2006) Inhibition of DNA binding of Sox2 by the SUMO conjugation. *Biochem. Biophys. Res. Commun.* **351**, 920–6.

[445] Newberry, E. P, Latifi, T, & Towler, D. A. (1999) The RRM domain of MINT, a novel Msx2 binding protein, recognizes and regulates the rat osteocalcin promoter. *Biochemistry* **38**, 10678–10690.

[446] Shi, Y, Downes, M, Xie, W, Kao, H. Y, Ordentlich, P, Hon, M, & Evans, R. M. (2001) Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. *Genes Dev.* **15**, 1140–1151.

[447] Yabe, D, Fukuda, H, Aoki, M, Yamada, S, Takebayashi, S, Shinkura, R, Yamamoto, N, & Honjo, T. (2007) Generation of a conditional knockout allele for mammalian Spen protein Mint/SHARP. *Genesis* **45**, 300–6.

[448] Sierra, O. L, Cheng, S. L, Loewy, A. P, Charlton-Kachigian, N, & Towler, D. A. (2004) MINT, the Msx2 interacting nuclear matrix target, enhances Runx2-dependent activation of the osteocalcin fibroblast growth factor response element. *J. Biol. Chem.* **279**, 32913–23.

[449] Lacroix, L, Linard, H, Labourier, E, Djavaheri-Mergny, M, Lacoste, J, Leffers, H, Tazi, J, Hlne, C, & Mergny, J. L. (2000) Identification of two human nuclear proteins that recognise the cytosine-rich strand of human telomeres in vitro. *Nucleic Acids Res.* **28**, 1564–1575.

[450] Mayeda, A, Screaton, G. R, Chandler, S. D, Fu, X. D, & Krainer, A. R. (1999) Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements. *Mol. Cell. Biol.* **19**, 1853–63.

[451] Ezponda, T, Pajares, M. J, Agorreta, J, Echeveste, J. I, Lopez-Picazo, J. M, Torre, W, Pio, R, & Montuenga, L. M. (2010) The oncoprotein SF2/ASF promotes non-small cell lung cancer survival by enhancing survivin expression. *Clin. Cancer Res.* **16**, 4113–4125.

[452] Clery, A, Sinha, R, Anczukow, O, Corrionero, A, Moursy, A, Daubner, G. M, Valcarcel, J, Krainer, A. R, & Allain, F. H. (2013) Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E2802–11.

[453] Wang, X, Juan, L, Lv, J, Wang, K, Sanford, J. R, & Liu, Y. (2011) Predicting sequence and structural specificities of RNA binding regions recognized by splicing factor SRSF1. *BMC Genomics* **12 Suppl 5**, S8.

[454] Curth, U, Urbanke, C, Greipel, J, Gerberding, H, Tiranti, V, & Zeviani, M. (1994) Single-stranded-DNA-binding proteins from human mitochondria and Escherichia coli have analogous physicochemical properties. *Eur. J. Biochem.* **221**, 435–443.

[455] Wanrooij, S, Fuste, J. M, Farge, G, Shi, Y, Gustafsson, C. M, & Falkenberg, M. (2008) Human mitochondrial RNA polymerase primes lagging-strand DNA synthesis in vitro. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 11122–7.

[456] Robertson, G, Hirst, M, Bainbridge, M, Bilenky, M, Zhao, Y, Zeng, T, Euskirchen, G, Bernier, B, Varhol, R, Delaney, A, Thiessen, N, Griffith, O. L, He, A, Marra, M, Snyder, M, & Jones, S. (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* **4**, 651–657.

[457] Peyman, J. A. (1999) Repression of major histocompatibility complex genes by a human trophoblast ribonucleic acid. *Biol. Reprod.* **60**, 23–31.

[458] Chen, X, Vinkemeier, U, Zhao, Y, Jeruzalmi, D, Darnell, J. E., J, & Kuriyan, J. (1998) Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA. *Cell* **93**, 827–39.

[459] Au-Yeung, N, Mandhana, R, & Horvath, C. M. (2013) Transcriptional regulation by STAT1 and STAT2 in the interferon JAK-STAT pathway. *Jak-Stat* **2**, e23931.

[460] Stark, G & Darnell Jr, J. (2012) The JAK-STAT pathway at twenty. *Immunity* **36**, 503–514.

[461] Becker, S, Groner, B, & Mller, C. W. (1998) Three-dimensional structure of the Stat3 homodimer bound to DNA. *Nature* **394**, 145–151.

[462] Wang, P, Xue, Y, Han, Y, Lin, L, Wu, C, Xu, S, Jiang, Z, Xu, J, Liu, Q, & Cao, X. (2014) The STAT3-Binding Long Noncoding RNA lnc-DC Controls Human Dendritic Cell Differentiation. *Science* **344**, 310–313.

[463] Ehret, G. B, Reichenbach, P, Schindler, U, Horvath, C. M, Fritz, S, Nabholz, M, & Bucher, P. (2001) DNA binding specificity of different STAT proteins" Comparison of in vitro specificity with natural target sites. *J. Biol. Chem.* **276**, 6675–88.

[464] Kaiser, K, Stelzer, G, & Meisterernst, M. (1995) The coactivator p15 (PC4) initiates transcriptional activation during TFIIA-TFIID-promoter complex formation. *EMBO J.* **14**, 3520–7.

[465] Ge, H & Roeder, R. G. (1994) Purification, cloning, and characterization of a human coactivator, PC4, that mediates transcriptional activation of class II genes. *Cell* **78**, 513–523.

[466] Brandsen, J, Werten, S, van der Vliet, P. C, Meisterernst, M, Kroon, J, & Gros, P. (1997) C-terminal domain of transcription cofactor PC4 reveals dimeric ssDNA binding site. *Nat. Struct. Biol.* **4**, 900–3.

[467] Conesa, C & Acker, J. (2010) Sub1/PC4 a chromatin associated protein with multiple functions in transcription. *RNA Biol.* **7**, 287–90.

[468] Werten, S & Moras, D. (2006) A global transcription cofactor bound to juxtaposed strands of unwound DNA. *Nat. Struct. Mol. Biol.* **13**, 181–2.

[469] Shu, Z, Vijayakumar, S, Chen, C.-F, Chen, P.-L, & Lee, W.-H. (2004) Purified human SUV3p exhibits multiple-substrate unwinding activity upon conformational change. *Biochemistry* **43**, 4781–4790.

[470] Szczesny, R. J, Wojcik, M. A, Borowski, L. S, Szewczyk, M. J, Skrok, M. M, Golik, P, & Stepien, P. P. (2013) Yeast and human mitochondrial helicases. *Biochim. Biophys. Acta* **1829**, 842–853.

[471] Jedrzejczak, R, Wang, J, Dauter, M, Szczesny, R. J, Stepien, P. P, & Dauter, Z. (2011) Human Suv3 protein reveals unique features among SF2 helicases. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 988–96.

[472] Magoulas, C, Zatsepina, O. V, Jordan, P. W. H, Jordan, E. G, & Fried, M. (1998) The SURF-6 protein is a component of the nucleolar matrix and has a high binding capacity for nucleic acids in vitro. *Eur. J. Cell Biol.* **75**, 174–183.

[473] Romanova, L. G, Anger, M, Zatsepina, O. V, & Schultz, R. M. (2006) Implication of nucleolar protein SURF6 in ribosome biogenesis and preimplantation mouse development. *Biol. Reprod.* **75**, 690–6.

[474] Kirmizis, A, Bartley, S. M, Kuzmichev, A, Margueron, R, Reinberg, D, Green, R, & Farnham, P. J. (2004) Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. *Genes Dev.* **18**, 1592–1605.

[475] Kotake, Y, Nakagawa, T, Kitagawa, K, Suzuki, S, Liu, N, Kitagawa, M, & Xiong, Y. (2011) Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* **30**, 1956–62.

[476] Cao, R & Zhang, Y. (2004) SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Mol. Cell* **15**, 57–67.

[477] Pasini, D, Bracken, A. P, Jensen, M. R, Denchi, E. L, & Helin, K. (2004) Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J.* **23**, 4061–4071.

[478] Bertolotti, A, Lutz, Y, Heard, D. J, Chambon, P, & Tora, L. (1996) hTAF(II)68, a novel RNA/ssDNA-binding protein with homology to the pro-oncoproteins TLS/FUS and EWS is associated with both TFIID and RNA polymerase II. *EMBO J.* **15**, 5022–31.

[479] Hoell, J. I, Larsson, E, Runge, S, Nusbaum, J. D, Duggimpudi, S, Farazi, T. A, Hafner, M, Borkhardt, A, Sander, C, & Tuschl, T. (2011) RNA targets of wild-type and mutant FET family proteins. *Nat. Struct. Mol. Biol.* **18**, 1428–31.

[480] Schatz, N, Brandlein, S, Ruckl, K, Hensel, F, & Vollmers, H. P. (2010) Diagnostic and therapeutic potential of a human antibody cloned from a cancer patient that binds to a tumor-specific variant of transcription factor TAF15. *Cancer Res.* **70**, 398–408.

[481] Ibrahim, F, Maragkakis, M, Alexiou, P, Maronski, M. A, Dichter, M. A, & Mourelatos, Z. (2013) Identification of in vivo, conserved, TAF15 RNA binding sites reveals the impact of TAF15 on the neuronal transcriptome. *Cell Rep.* **3**, 301–8.

[482] Kovar, H. (2011) Dr. Jekyll and Mr. Hyde: The two faces of the FUS/EWS/TAF15 protein family. *Sarcoma* **2011**, 837474.

[483] Buratti, E & Baralle, F. E. (2001) Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. *J. Biol. Chem.* **276**, 36337–36343.

[484] Ou, S. H, Wu, F, Harrich, D, Garca-Martnez, L. F, & Gaynor, R. B. (1995) Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs. *J. Virol.* **69**, 3584–96.

[485] Buratti, E, Dork, T, Zuccato, E, Pagani, F, Romano, M, & Baralle, F. E. (2001) Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *EMBO J.* **20**, 1774–84.

[486] Kawahara, Y & Mieda-Sato, A. (2012) TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 3347–52.

[487] Ayala, Y. M, Pantano, S, D'Ambrogio, A, Buratti, E, Brindisi, A, Marchetti, C, Romano, M, & Baralle, F. E. (2005) Human, Drosophila, and C.elegans TDP43: nucleic acid binding properties and splicing regulatory function. *J. Mol. Biol.* **348**, 575–88.

[488] Lee, E. B, Lee, V. M. Y, & Trojanowski, J. Q. (2011) Gains or losses: molecular mechanisms of TDP43-mediated neurodegeneration. *Nat. Rev. Neurosci.*

[489] Kuo, P. H, Doudeva, L. G, Wang, Y. T, Shen, C. K. J, & Yuan, H. S. (2009) Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res.* **37**, 1799–1808.

[490] Lukavsky, P. J, Daujotyte, D, Tollervey, J. R, Ule, J, Stuani, C, Buratti, E, Baralle, F. E, Damberger, F. F, & Allain, F. H. T. (2013) Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat. Struct. Mol. Biol.* **20**, 1443–1449.

[491] Hiroi, Y, Kudoh, S, Monzen, K, Ikeda, Y, Yazaki, Y, Nagai, R, & Komuro, I. (2001) Tbx5 associates with Nkx2-5 and synergistically promotes cardiomyocyte differentiation. *Nat. Genet.* **28**, 276–280.

[492] Fan, C, Chen, Q, & Wang, Q. K. (2009) Functional role of transcriptional factor TBX5 in pre-mRNA splicing and Holt-Oram syndrome via association with SC35. *J. Biol. Chem.* **284**, 25653–25663.

[493] Hatcher, C. J & McDermott, D. A. (2006) Using the TBX5 transcription factor to grow and sculpt the heart. *Am. J. Med. Genet.* **140A**, 1414–1418.

[494] Stirnimann, C. U, Ptchelkine, D, Grimm, C, & Muller, C. W. (2010) Structural basis of TBX5-DNA recognition: the T-box domain in its DNA-bound and -unbound form. *J. Mol. Biol.* **400**, 71–81.

[495] van Beest, M, Dooijes, D, van De Wetering, M, Kjaerulff, S, Bonvin, A, Nielsen, O, & Clevers, H. (2000) Sequence-specific high mobility group box factors recognize 10-12-base pair minor groove motifs. *J. Biol. Chem.*

[496] Lee, S. K, Park, M. W, Yang, E. G, Yu, J, & Jeong, S. (2005) An RNA aptamer that binds to the β-catenin interaction domain of TCF-1 protein. *Biochem. Biophys. Res. Commun.* **327**, 294–299.

[497] Ma, J, Wang, R, Fang, X, & Sun, Z. (2012) beta-catenin/TCF-1 pathway in T cell development and differentiation. *J. Neuroimmune. Pharmacol.* **7**, 750–62.

[498] Staal, F. J. T, Meeldijk, J, Moerer, P, Jay, P, van de Weerdt, B. C. M, Vainio, S, Nolan, G. P, & Clevers, H. (2001) Wnt signaling is required for thymocyte development and activates Tcf-1 mediated transcription. *Eur. J. Immunol.* **31**, 285–293.

[499] Yi, F, Pereira, L, Hoffman, J. A, Shy, B. R, Yuen, C. M, Liu, D. R, & Merrill, B. J. (2011) Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal. *Nat. Cell Biol.* **13**, 762–770.

[500] Park, M. W, Choi, K. H, & Jeong, S. (2005) Inhibition of the DNA binding by the TCF-1 binding RNA aptamer. *Biochem. Biophys. Res. Commun.* **330**, 11–17.

[501] Broccoli, D, Smogorzewska, A, Chong, L, & de Lange, T. (1997) Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nat. Genet.* **17**, 231–5.

[502] Court, R, Chapman, L, Fairall, L, & Rhodes, D. (2005) How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures. *EMBO Rep.* **6**, 39–45.

[503] Verdun, R. E & Karlseder, J. (2007) Replication and protection of telomeres. *Nature* **447**, 924–931.

[504] Okamoto, K, Bartocci, C, Ouzounov, I, Diedrich, J. K, Yates Iii, J. R, & Denchi, E. L. (2013) A two-step mechanism for TRF2-mediated chromosome-end protection. *Nature* **494**, 502–505.

[505] Sap, J, Munoz, A, Damm, K, Goldberg, Y, Ghysdael, J, Leutz, A, Beug, H, & Vennstrom, B. (1986) The c-erb-A protein is a high-affinity receptor for thyroid hormone. *Nature* **324**, 635–40.

[506] Xu, B & Koenig, R. J. (2004) An RNA-binding domain in the thyroid hormone receptor enhances transcriptional activation. *J. Biol. Chem.* **279**, 33051–33056.

[507] Brent, G. A. (2012) Mechanisms of thyroid hormone action. *J. Clin. Invest.* **122**, 3035–3043.

[508] Rastinejad, F, Perlmann, T, Evans, R. M, & Sigler, P. B. (1995) Structural determinants of nuclear receptor assembly on DNA direct repeats. *Nature* **375**, 203–11.

[509] Chen, Y & Young, M. A. (2010) Structure of a thyroid hormone receptor DNA-binding domain homodimer bound to an inverted palindrome DNA response element. *Mol. Endocrinol.* **24**, 1650–1664.

[510] McAlinden, A, Liang, L, Mukudai, Y, Imamura, T, & Sandell, L. J. (2007) Nuclear protein TIA-1 regulates COL2A1 alternative splicing and interacts with precursor mRNA and genomic DNA. *J. Biol. Chem.* **282**, 24444–24454.

[511] Anderson, P & Kedersha, N. (2002) Visibly stressed: the role of eIF2, TIA-1, and stress granules in protein translation. *Cell Stress Chaperones* **7**, 213–21.

[512] Suswam, E. A, Li, Y. Y, Mahtani, H, & King, P. H. (2005) Novel DNA-binding properties of the RNA-binding protein TIAR. *Nucleic Acids Res.* **33**, 4507–4518.

[513] Kang, M. R, Muller, M. T, & Chung, I. K. (2003) Telomeric DNA damage by topoisomerase I: A possible mechanism for cell killing by camptothecin. *J. Biol. Chem.* **279**, 12535–12541.

[514] Redinbo, M. R, Champoux, J. J, & Hol, W. G. J. (2000) Novel insights into catalytic mechanism from a crystal structure of human topoisomerase I in complex with DNA. *Biochemistry* **39**, 6832–6840.

[515] Arimondo, P. B, Riou, J. F, Mergny, J. L, Tazi, J, Sun, J. S, Garestier, T, & Helene, C. (2000) Interaction of human DNA topoisomerase I with G-quartet structures. *Nucleic Acids Res.* **28**, 4832–8.

[516] Takahashi, H, Sawa, H, Hasegawa, H, Sata, T, Hall, W. W, & Kurata, T. (2002) Binding and dissociation of human topoisomerase I with hairpin-loop RNAs: implications for the regulation of HIV-1 replication. *Biochem. Biophys. Res. Commun.* **297**, 593–599.

[517] Iyama, T & Wilson 3rd, D. M. (2013) DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair* **12**, 620–636.

[518] Pommier, Y, Leo, E, Zhang, H, & Marchand, C. (2010) DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chem. Biol.* **17**, 421–433.

[519] Wang, Y, Knudsen, B. R, Bjergbk, L, Westergaard, O, & Andersen, A. H. (1999) Stimulated activity of human topoisomerases II and II on RNA-containing substrates. *J. Biol. Chem.* **274**, 22839–22846.

[520] Stewart, L, Redinbo, M. R, Qiu, X, Hol, W. G, & Champoux, J. J. (1998) A model for the mechanism of human topoisomerase I. *Science* **279**, 1534–41.

[521] Oberosler, P, Hloch, P, Ramsperger, U, & Stahl, H. (1993) p53-catalyzed annealing of complementary single-stranded nucleic acids. *EMBO J.* **12**, 2389–96.

[522] Steinmeyer, K & Deppert, W. (1988) DNA binding properties of murine p53. *Oncogene* **3**, 501–7.

[523] Zotchev, S. B, Protopopova, M, & Selivanova, G. (2000) p53 C-terminal interaction with DNA ends and gaps has opposing effect on specific DNA binding by the core. *Nucleic Acids Res.* **28**, 4005–12.

[524] Mosner, J, Mummenbrauer, T, Bauer, C, Sczakiel, G, Grosse, F, & Deppert, W. (1995) Negative feedback regulation of wild-type p53 biosynthesis. *EMBO J.* **14**, 4442–9.

[525] Shetty, S, Shetty, P, Idell, S, Velusamy, T, Bhandary, Y. P, & Shetty, R. S. (2008) Regulation of plasminogen activator inhibitor-1 expression by tumor suppressor protein p53. *J. Biol. Chem.* **283**, 19570–19580.

[526] Cho, Y, Gorina, S, Jeffrey, P. D, & Pavletich, N. P. (1994) Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science* **265**, 346–55.

[527] Brady, C. A & Attardi, L. D. (2010) p53 at a glance. *J. Cell Sci.* **123**, 2527–2532.

[528] Yoshida, Y, Izumi, H, Torigoe, T, Ishiguchi, H, Yoshida, T, Itoh, H, & Kohno, K. (2004) Binding of RNA to p53 regulates its oligomerization and DNA-binding activity. *Oncogene* **23**, 4371–4379.

[529] Riley, K. J & Maher III, L. J. (2007) p53 RNA interactions: new clues in an old mystery. *RNA* **13**, 1825–33.

[530] Ho, W. C, Fitzgerald, M. X, & Marmorstein, R. (2006) Structure of the p53 core domain dimer bound to DNA. *J. Biol. Chem.* **281**, 20494–502.

[531] Aoki, K, Suzuki, K, Ishida, R, & Kasai, M. (1999) The DNA binding activity of Translin is mediated by a basic region in the ring-shaped structure conserved in evolution. *FEBS Lett.* **443**, 363–366.

[532] Jacob, E, Pucshansky, L, Zeruya, E, Baran, N, & Manor, H. (2004) The human protein translin specifically binds single-stranded microsatellite repeats, d(GT)n, and G-strand telomeric repeats, d(TTAGGG)n: a study of the binding parameters. *J. Mol. Biol.* **344**, 939–950.

[533] Morales, C. R, Wu, X. Q, & Hecht, N. B. (1998) The DNA/RNA-binding protein, TB-RBP, moves from the nucleus to the cytoplasm and through intercellular bridges in male germ cells. *Dev. Biol.* **201**, 113–123.

[534] Wang, J, Boja, E. S, Oubrahim, H, & Chock, P. B. (2004) Testis brain ribonucleic acid-binding protein/translin possesses both single-stranded and double-stranded ribonuclease activities. *Biochemistry* **43**, 13424–13431.

[535] Yu, Z & Hecht, N. B. (2008) The DNA/RNA-binding protein, translin, binds microRNA122a and increases its in vivo stability. *J. Androl.* **29**, 572–9.

[536] Jaendling, A & McFarlane, R. J. (2010) Biological roles of translin and translin-associated factor-X: RNA metabolism comes to the fore. *Biochem. J.* **429**, 225–234.

[537] Li, Z, Wu, Y, & Baraban, J. M. (2008) The Translin/Trax RNA binding complex: clues to function in the nervous system. *Biochim. Biophys. Acta* **1779**, 479–85.

[538] Parizotto, E. A, Lowe, E. D, & Parker, J. S. (2013) Structural basis for duplex RNA recognition and cleavage by Archaeoglobus fulgidus C3PO. *Nat. Struct. Mol. Biol.* **20**, 380–6.

[539] Komuro, A, Saeki, M, & Kato, S. (1999) Association of two nuclear proteins, Npw38 and NpwBP, via the interaction between the WW domain and a novel proline-rich motif containing glycine and arginine. *J. Biol. Chem.* **274**, 36513–36519.

[540] Craggs, G, Finan, P. M, Lawson, D, Wingfield, J, Perera, T, Gadher, S, Totty, N. F, & Kellie, S. (2001) A nuclear SH3 domain-binding protein that colocalizes with mRNA splicing factors and intermediate filament-containing perinuclear networks. *J. Biol. Chem.* **276**, 30552–30560.

[541] Caricasole, A, Duarte, A, Larsson, S. H, Hastie, N. D, Little, M, Holmes, G, Todorov, I, & Ward, A. (1996) RNA binding by the Wilms tumor suppressor zinc finger proteins. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 7562–6.

[542] Stoll, R, Lee, B. M, Debler, E. W, Laity, J. H, Wilson, I. A, Dyson, H. J, & Wright, P. E. (2007) Structure of the Wilms tumor suppressor protein zinc finger domain bound to DNA. *J. Mol. Biol.* **372**, 1227–45.

[543] Chau, Y.-Y & Hastie, N. D. (2012) The role of Wt1 in regulating mesenchyme in cancer, development, and tissue homeostasis. *Trends Genet.* **28**, 515–524.

[544] Kolluri, R, Torrey, T. A, & Kinniburgh, A. J. (1992) A CT promoter element binding protein: definition of a double-strand and a novel single-strand DNA binding motif. *Nucleic Acids Res.* **20**, 111–116.

[545] Ise, T, Nagatani, G, Imamura, T, Kato, K, Takano, H, Nomoto, M, Izumi, H, Ohmori, H, Okamoto, T, Ohga, T, Uchiumi, T, Kuwano, M, & Kohno, K. (1999) Transcription factor Y-box binding protein 1 binds preferentially to cisplatin-modified DNA and interacts with proliferating cell nuclear antigen. *Cancer Res.* **59**, 342–6.

[546] Stickeler, E, Fraser, S. D, Honig, A, Chen, A. L, Berget, S. M, & Cooper, T. A. (2001) The RNA binding protein YB-1 binds A/C-rich exon enhancers and stimulates splicing of the CD44 alternative exon v4. *EMBO J.* **20**, 3821–30.

[547] Izumi, H, Imamura, T, Nagatani, G, Ise, T, Murakami, T, Uramoto, H, Torigoe, T, Ishiguchi, H, Yoshida, H, Yoshida, Y, Nomoto, M, Okamoto, T, Uchiumi, T, Kuwano, M, Funa, K, & Kohno, K. (2001) Y box-binding protein-1 binds preferentially to single-stranded nucleic acids and exhibits 3'->5' exonuclease activity. *Nucleic Acids Res.* **29**, 1200–1207.

[548] Lasham, A, Print, C. G, Woolley, A. G, Dunn, S. E, & Braithwaite, A. W. (2013) YB-1: oncoprotein, prognostic marker and therapeutic target? *Biochem. J.* **449**, 11–23.

[549] Ohmori, M, Shimura, H, Shimura, Y, & Kohn, L. D. (1996) A Y-box protein is a suppressor factor that decreases thyrotropin receptor gene expression. *Mol. Endocrinol.* **10**, 76–89.

[550] Yang, J, Medvedev, S, Reddi, P. P, Schultz, R. M, & Hecht, N. B. (2005) The DNA/RNA-binding protein MSY2 marks specific transcripts for cytoplasmic storage in mouse male germ cells. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 1513–8.

[551] Coles, L. S, Diamond, P, Occhiodoro, F, Vadas, M. A, & Shannon, M. F. (1996) Cold shock domain proteins repress transcription from the GM-CSF promoter. *Nucleic Acids Res.* **24**, 2311–7.

[552] Giorgini, F, Davies, H. G, & Braun, R. E. (2001) MSY2 and MSY4 bind a conserved sequence in the 3' untranslated region of protamine 1 mRNA in vitro and in vivo. *Mol. Cell. Biol.* **21**, 7010–9.

[553] Li, W. W, Hsiung, Y, Zhou, Y, Roy, B, & Lee, A. S. (1997) Induction of the mammalian GRP78/BiP gene by Ca2+ depletion and formation of aberrant proteins: activation of the conserved stress-inducible grp core promoter element by the human nuclear factor YY1. *Mol. Cell. Biol.* **17**, 54–60.

[554] Jeon, Y & Lee, J. (2011) YY1 Tethers Xist RNA to the Inactive X Nucleation Center. *Cell* **146**, 119–133.

[555] Atchison, M. L. (2014) Function of YY1 in Long-Distance DNA Interactions. *Front. Immunol.* **5**, 45.

[556] Houbaviy, H. B, Usheva, A, Shenk, T, & Burley, S. K. (1996) Cocrystal structure of YY1 bound to the adeno-associated virus P5 initiator. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13577–82.

[557] Hwang, E. S, Choi, A, & Ho, I.-C. (2002) Transcriptional Regulation of GATA-3 by an Intronic Regulatory Region and Fetal Liver Zinc Finger Protein 1. *J. Immunol.* **169**, 248–253.

[558] Hwang, E. S & Ho, I. C. (2002) Regulation of thymocyte homeostasis by Fliz1. *Immunology* **106**, 464–9.

[559] Arranz, V, Harper, F, Florentin, Y, Puvion, E, Kress, M, & Ernoult-Lange, M. (1997) Human and mouse MOK2 proteins are associated with nuclear ribonucleoprotein components and bind specifically to RNA and DNA through their zinc finger domains. *Mol. Cell. Biol.* **17**, 2116–26.

[560] Arranz, V, Dreuillet, C, Crisanti, P, Tillit, J, Kress, M, & Ernoult-Lange, M. (2001) The zinc finger transcription factor, MOK2, negatively modulates expression of the interphotoreceptor retinoid-binding protein gene, IRBP. *J. Biol. Chem.* **276**, 11963–9.

[561] Inagaki, H, Matsushima, Y, Nakamura, K, Ohshima, M, Kadowaki, T, & Kitagawa, Y. (1996) A large DNA-binding nuclear protein with RNA recognition motif and serine/arginine-rich domain. *J. Biol. Chem.* **271**, 12525–12531.

[562] Meruvu, S, Hugendubler, L, & Mueller, E. (2011) Regulation of adipocyte differentiation by the zinc finger protein ZNF638. *J. Biol. Chem.* **286**, 26516–23.

[563] Huang da, W, Sherman, B. T, & Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57.

[564] Huang da, W, Sherman, B. T, & Lempicki, R. A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13.

[565] Reddy, T. E, Pauli, F, Sprouse, R. O, Neff, N. F, Newberry, K. M, Garabedian, M. J, & Myers, R. M. (2009) Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. *Genome Res.* **19**, 2163–2171.

[566] Clark, A. R & Belvisi, M. G. (2012) Maps and legends: The quest for dissociated ligands of the glucocorticoid receptor. *Pharmacol. Ther.* **134**, 54–67.

[567] Kanai, A, Tani, H, Torimura, M, & Akimitsu, N. (2013) The RNA degradation pathway regulates the function of GAS5 a non-coding RNA in mammalian cells. *PLoS ONE* **8**, e55684.

[568] Gros, C, Fahy, J, Halby, L, Dufau, I, Erdmann, A, Gregoire, J.-M, Ausseil, F, Visp, S, & Arimondo, P. B. (2012) DNA methylation inhibitors in cancer: Recent and future approaches. *Biochimie* **94**, 2280–2296.

[569] Fletcher, L, Rider, C. C, & Taylor, C. B. (1976) Enolase isoenzymes. III. Chromatographic and immunological characteristics of rat brain enolase. *Biochim. Biophys. Acta* **452**, 245–52.

[570] Anderson, S. L, Minard, K. I, & McAlister-Henn, L. (2000) Allosteric inhibition of NAD+-specific isocitrate dehydrogenase by a mitochondrial mRNA. *Biochemistry* **39**, 5623–5629.

[571] Hentze, M. W & Preiss, T. (2010) The REM phase of gene regulation. *Trends Biochem. Sci.* **35**, 423–426.

[572] Auphan, N, DiDonato, J, Rosette, C, Helmberg, A, & Karin, M. (1995) Immunosuppression by glucocorticoids: inhibition of NF-kappa B activity through induction of I kappa B synthesis. *Science* **270**, 286–290.

[573] Jonat, C, Rahmsdorf, H. J, Park, K. K, Cato, A. C, Gebel, S, Ponta, H, & Herrlich, P. (1990) Antitumor promotion and antiinflammation: down-modulation of AP-1 (Fos/Jun) activity by glucocorticoid hormone. *Cell* **62**, 1189–204.

[574] Surjit, M, Ganti, K. P, Mukherji, A, Ye, T, Hua, G, Metzger, D, Li, M, & Chambon, P. (2011) Widespread negative response elements mediate direct repression by agonist-liganded glucocorticoid receptor. *Cell* **145**, 224–241.

[575] Ray, A & Prefontaine, K. E. (1994) Physical association and functional antagonism between the p65 subunit of transcription factor NF-B and the glucocorticoid receptor. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 752–756.

[576] Yang-Yen, H.-F, Chambard, J.-C, Sun, Y.-L, Smeal, T, Schmidt, T. J, Drouin, J, & Karin, M. (1990) Transcriptional interference between c-Jun and the glucocorticoid receptor: Mutual inhibition of DNA binding due to direct protein-protein interaction. *Cell* **62**, 1205–1215.

[577] Filipowicz, W, Bhattacharyya, S. N, & Sonenberg, N. (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.* **2008**, 102–114.

[578] Kim, V. N, Han, J, & Siomi, M. C. (2009) Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.* **10**, 126–139.

[579] Massagué, J & Wotton, D. (2000) Transcriptional control by the TGF-β/Smad signaling system. *EMBO J.* **19**, 1745–1754.

[580] Kumarswamy, R, Volkmann, I, & Thum, T. (2011) Regulation and function of miRNA-21 in health and disease. *RNA Biol.* **8**, 706–713.

[581] Kao, P. N, Chen, L, Brock, G, Ng, J, Kenny, J, Smith, A. J, & Corthesy, B. (1994) Cloning and expression of cyclosporin A- and FK506-sensitive nuclear factor of activated T-cells: NF45 and NF90. *J. Biol. Chem.* **269**, 20691–9.

[582] Shim, J, Lim, H, R.Yates Iii, J, & Karin, M. (2002) Nuclear export of NF90 is required for interleukin-2 mRNA stabilization. *Mol. Cell* **10**, 1331–1344.

[583] Sakamoto, S, Aoki, K, Higuchi, T, Todaka, H, Morisawa, K, Tamaki, N, Hatano, E, Fukushima, A, Taniguchi, T, & Agata, Y. (2009) The NF90-NF45 complex functions as a negative regulator in the microRNA processing pathway. *Mol. Cell. Biol.* **29**, 3754–3769.

[584] Malek, T. R & Castro, I. (2010) Interleukin-2 Receptor Signaling: At the Interface between Tolerance and Immunity. *Immunity* **33**, 153–165.

[585] Dienz, O & Rincon, M. (2009) The effects of IL-6 on CD4 T cell responses. *Clim. Immunol.* **130**, 27–33.

[586] Iliopoulos, D, Hirsch, H. A, & Struhl, K. (2009) An epigenetic switch involving NF-κB, Lin28, Let-7 MicroRNA, and IL6 links inflammation to cell transformation. *Cell* **139**, 693–706.

[587] Takamizawa, J, Konishi, H, Yanagisawa, K, Tomida, S, Osada, H, Endoh, H, Harano, T, Yatabe, Y, Nagino, M, Nimura, Y, Mitsudomi, T, & Takahashi, T. (2004) Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res.* **64**, 3753–3756.

[588] Guo, N. L, Wan, Y. W, Tosun, K, Lin, H, Msiska, Z, Flynn, D. C, Remick, S. C, Vallyathan, V, Dowlati, A, Shi, X, Castranova, V, Beer, D. G, & Qian, Y. (2008) Confirmation of gene expression-based prediction of survival in non-small cell lung cancer. *Clin. Cancer Res.* **14**, 8213–8220.

[589] Sarkar, A & Hochedlinger, K. (2013) The Sox family of transcription factors: versatile regulators of stem and progenitor cell fate. *Cell Stem Cell* **12**, 15–30.

[590] Bond, A. M, VanGompel, M. J. W, Sametsky, E. A, Clark, M. F, Savage, J. C, Disterhoft, J. F, & Kohtz, J. D. (2009) Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. *Nat. Neurosci.* **12**, 1020–1027.

[591] Wassenegger, M, Heimes, S, Riedel, L, & Snger, H. L. (1994) RNA-directed de novo methylation of genomic sequences in plants. *Cell* **76**, 567–576.

[592] Schmitz, K. M, Mayer, C, Postepska, A, & Grummt, I. (2010) Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev.* **24**, 2264–2269.

[593] Morris, K. V, Chan, S. W, Jacobsen, S. E, & Looney, D. J. (2004) Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* **305**, 1289–1292.

[594] Sun, B. K, Deaton, A. M, & Lee, J. T. (2006) A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. *Mol. Cell* **21**, 617–628.

[595] Zhao, X, Patton, J. R, Davis, S. L, Florence, B, Ames, S. J, & Spanjaard, R. A. (2004) Regulation of nuclear receptor activity by a pseudouridine synthase through posttranscriptional modification of steroid receptor RNA activator. *Mol. Cell* **15**, 549–558.

[596] Watanabe, M, Yanagisawa, J, Kitagawa, H, Takeyama, K, Ogawa, S, Arao, Y, Suzawa, M, Kobayashi, Y, Yano, T, Yoshikawa, H, Masuhiro, Y, & Kato, S. (2001) A subfamily of RNA-binding DEAD-box proteins acts as an estrogen receptor coactivator through the N-terminal activation domain (AF-1) with an RNA coactivator, SRA. *EMBO J.* **20**, 1341–1352.

[597] Deblois, G & Giguere, V. (2003) Ligand-independent coactivation of ER AF-1 by steroid receptor RNA activator (SRA) via MAPK activation. *J. Steroid Biochem. Mol. Biol.* **85**, 123–31.

[598] van Steensel, B, Smogorzewska, A, & de Lange, T. (1998) TRF2 protects human telomeres from end-to-end fusions. *Cell* **92**, 401–413.

[599] Chesnokov, I. N. (2007) in *Int. Rev. Cytol.*, ed. Kwang, W. J. (Academic Press) Vol. Volume 256, pp. 69–109.

[600] Deng, Z, Dheekollu, J, Broccoli, D, Dutta, A, & Lieberman, P. M. (2007) The origin recognition complex localizes to telomere repeats and prevents telomere-circle formation. *Curr. Biol.* **17**, 1989–1995.

[601] Maris, C, Dominguez, C, & Allain, F. H. T. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **272**, 2118–2131.

[602] Bahadur, R. P, Zacharias, M, & Janin, J. (2008) Dissecting protein-RNA recognition sites. *Nucleic Acids Res.* **36**, 2705–16.

[603] Allers, J & Shamoo, Y. (2001) Structure-based analysis of protein-RNA interactions using the program ENTANGLE. *J. Mol. Biol.* **311**, 75–86.

[604] Deo, R. C, Bonanno, J. B, Sonenberg, N, & Burley, S. K. (1999) Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* **98**, 835–845.

[605] Lu, X. J & Olson, W. K. (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.* **3**, 1213–1227.

[606] Barrick, J. E, Sudarsan, N, Weinberg, Z, Ruzzo, W. L, & Breaker, R. R. (2005) 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. *RNA* **11**, 774–784.

[607] Didier, D. K, Schiffenbauer, J, Woulfe, S. L, Zacheis, M, & Schwartz, B. D. (1988) Characterization of the cDNA encoding a protein binding to the major histocompatibility complex class II Y box. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7322–6.

[608] Skabkina, O. V, Lyabin, D. N, Skabkin, M. A, & Ovchinnikov, L. P. (2005) YB-1 autoregulates translation of its own mRNA at or prior to the step of 40S ribosomal subunit joining. *Mol. Cell. Biol.* **25**, 3317–23.

[609] Evdokimova, V, Ruzanov, P, Imataka, H, Raught, B, Svitkin, Y, Ovchinnikov, L. P, & Sonenberg, N. (2001) The major mRNA-associated protein YB-1 is a potent 5' cap-dependent mRNA stabilizer. *EMBO J.* **20**, 5491–5502.

[610] Marenstein, D. R, Ocampo, M. T, Chan, M. K, Altamirano, A, Basu, A. K, Boorstein, R. J, Cunningham, R. P, & Teebor, G. W. (2001) Stimulation of human endonuclease III by Y box-binding protein 1 (DNA-binding protein B). Interaction between a base excision repair enzyme and a transcription factor. *J. Biol. Chem.* **276**, 21242–9.

[611] de Souza-Pinto, N. C, Mason, P. A, Hashiguchi, K, Weissman, L, Tian, J, Guay, D, Lebel, M, Stevnsner, T. V, Rasmussen, L. J, & Bohr, V. A. (2009) Novel DNA mismatch-repair activity involving YB-1 in human mitochondria. *DNA Repair* **8**, 704–719.

[612] Koike, K, Uchiumi, T, Ohga, T, Toh, S, Wada, M, Kohno, K, & Kuwano, M. (1997) Nuclear translocation of the Y-box binding protein by ultraviolet irradiation. *FEBS Lett.* **417**, 390–4.

[613] Yamanaka, K, Fang, L, & Inouye, M. (1998) The CspA family in Escherichia coli: multiple gene duplication for stress adaptation. *Mol. Microbiol.* **27**, 247–55.

[614] Etchegaray, J. P, Jones, P. G, & Inouye, M. (1996) Differential thermoregulation of two highly homologous cold-shock genes, cspA and cspB, of Escherichia coli. *Genes Cells* **1**, 171–8.

[615] Nakashima, K, Kanamaru, K, Mizuno, T, & Horikoshi, K. (1996) A novel member of the cspA family of genes that is induced by cold shock in Escherichia coli. *J. Bacteriol.* **178**, 2994–7.

[616] Wang, N, Yamanaka, K, & Inouye, M. (1999) CspI, the ninth member of the CspA family of Escherichia coli, is induced upon cold shock. *J. Bacteriol.* **181**, 1603–9.

[617] Goldstein, J, Pollitt, N. S, & Inouye, M. (1990) Major cold shock protein of Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 283–7.

[618] Xia, B, Ke, H, & Inouye, M. (2001) Acquirement of cold sensitivity by quadruple deletion of the cspA family and its suppression by PNPase S1 domain in Escherichia coli. *Mol. Microbiol.* **40**, 179–188.

[619] Yamanaka, K & Inouye, M. (1997) Growth-phase-dependent expression of cspD, encoding a member of the CspA family in Escherichia coli. *J. Bacteriol.* **179**, 5126–30.

[620] Phadtare, S & Inouye, M. (2001) Role of CspC and CspE in regulation of expression of RpoS and UspA, the stress response proteins in Escherichia coli. *J. Bacteriol.* **183**, 1205–1214.

[621] Jiang, W, Hou, Y, & Inouye, M. (1997) CspA, the Major Cold-shock Protein of Escherichia coli, Is an RNA Chaperone. *J. Biol. Chem.* **272**, 196–202.

[622] Phadtare, S & Inouye, M. (1999) Sequence-selective interactions with RNA by CspB, CspC and CspE, members of the CspA family of Escherichia coli. *Mol. Endocrinol.* **33**, 1004–1014.

[623] Hanna, M. M & Liu, K. (1998) Nascent RNA in transcription complexes interacts with CspE, a small protein in E. coli implicated in chromatin condensation. *J. Mol. Biol.* **282**, 227–39.

[624] Bae, W, Xia, B, Inouye, M, & Severinov, K. (2000) Escherichia coli CspA-family RNA chaperones are transcription antiterminators. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7784–9.

[625] Yamanaka, K, Zheng, W, Crooke, E, Wang, Y. H, & Inouye, M. (2001) CspD, a novel DNA replication inhibitor induced during the stationary phase in Escherichia coli. *Mol. Microbiol.* **39**, 1572–84.

[626] Chen, C. S, Korobkova, E, Chen, H, Zhu, J, Jian, X, Tao, S. C, He, C, & Zhu, H. (2008) A proteome chip approach reveals new DNA damage recognition activities in Escherichia coli. *Nat. Methods* **5**, 69–74.

[627] Hu, K. H, Liu, E, Dean, K, Gingras, M, DeGraff, W, & Trun, N. J. (1996) Overproduction of three genes leads to camphor resistance and chromosome condensation in Escherichia coli. *Genetics* **143**, 1521–32.

[628] Karlson, D & Imai, R. (2003) Conservation of the cold shock domain protein family in plants. *Plant Physiol.* **131**, 12–5.

[629] Karlson, D, Nakaminami, K, Toyomasu, T, & Imai, R. (2002) A cold-regulated nucleic acid-binding protein of winter wheat shares a domain with bacterial cold shock proteins. *J. Biol. Chem.* **277**, 35248–35256.

[630] Nakaminami, K, Karlson, D. T, & Imai, R. (2006) Functional conservation of cold shock domains in bacteria and higher plants. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10122–10127.

[631] Imai, R, Kim, M. H, Sasaki, K, Sato, S, & Sonoda, Y. (2013) *Plant and Microbe Adaptations to Cold in a Changing World.* (Springer New York).

[632] Juntawong, P, Sorenson, R, & Bailey-Serres, J. (2013) Cold shock protein1 chaperones mRNAs during translation in Arabidopsis thaliana. *Plant J.* **74**, 1016–1028.

[633] Kim, J. S, Park, S. J, Kwak, K. J, Kim, Y. O, Kim, J. Y, Song, J, Jang, B, Jung, C. H, & Kang, H. (2007) Cold shock domain proteins and glycine-rich RNA-binding proteins from Arabidopsis thaliana can promote the cold adaptation process in Escherichia coli. *Nucleic Acids Res.* **35**, 506–16.

[634] Triqueneaux, G, Velten, M, Franzon, P, Dautry, F, & Jacquemin-Sablon, H. (1999) RNA binding specificity of Unr, a protein with five cold shock domains. *Nucleic Acids Res.* **27**, 1926–34.

[635] Eliseeva, I. A, Kim, E. R, Guryanov, S. G, Ovchinnikov, L. P, & Lyabin, D. N. (2012) Y-box-binding protein 1 (YB-1) and its functions. *Biochemistry (Mosc)* **76**, 1402–1433.

[636] Mihailovich, M, Militti, C, Gabaldn, T, & Gebauer, F. (2010) Eukaryotic cold shock domain proteins: highly versatile regulators of gene expression. *BioEssays* **32**, 109–118.

[637] Hermann, A, Schmitt, S, & Jeltsch, A. (2003) The human Dnmt2 has residual DNA-(cytosine-C5) methyltransferase activity. *J. Biol. Chem.* **278**, 31717–31721.

[638] Lyko, F, Jurkowski, T. P, & Jeltsch, A. (2011) On the Evolutionary Origin of Eukaryotic DNA Methyltransferases and Dnmt2. *PLoS ONE* **6**, e28104.

[639] Pavlopoulou, A & Kossida, S. (2009) Phylogenetic analysis of the eukaryotic RNA (cytosine-5)-methyltransferases. *Genomics* **93**, 350–357.

[640] Reiter, N. J, Maher, L. J, & Butcher, S. E. (2008) DNA mimicry by a high-affinity anti-NF-κB RNA aptamer. *Nucleic Acids Res.* **36**, 1227–1236.

[641] Ray, D, Kazan, H, Cook, K. B, Weirauch, M. T, Najafabadi, H. S, Li, X, Gueroussov, S, Albu, M, Zheng, H, Yang, A, Na, H, Irimia, M, Matzat, L. H, Dale, R. K, Smith, S. A, Yarosh, C. A, Kelly, S. M, Nabet, B, Mecenas, D, Li, W, Laishram, R. S, Qiao, M, Lipshitz, H. D, Piano, F, Corbett, A. H, Carstens, R. P, Frey, B. J, Anderson, R. A, Lynch, K. W, Penalva, L. O. F, Lei, E. P, Fraser, A. G, Blencowe, B. J, Morris, Q. D, & Hughes, T. R. (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172–177.

[642] Nakagawa, S, Gisselbrecht, S. S, Rogers, J. M, Hartl, D. L, & Bulyk, M. L. (2013) DNA-binding specificity changes in the evolution of forkhead transcription factors. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 12349–12354.

[643] Tompa, P & Csermely, P. (2004) The role of structural disorder in the function of RNA and protein chaperones. *FASEB J.* **18**, 1169–1175.

[644] Volders, P. J, Helsens, K, Wang, X, Menten, B, Martens, L, Gevaert, K, Vandesompele, J, & Mestdagh, P. (2012) LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res.* **41**, D246–D251.

[645] Remnyi, A, Lins, K, Nissen, L. J, Reinbold, R, Schler, H. R, & Wilmanns, M. (2003) Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev.* **17**, 2048–2059.

[646] Rossini, G. P. (1984) RNA-containing nuclear binding sites for glucocorticoid-receptor complexes. *Biochem. Biophys. Res. Commun.* **123**, 78–83.

[647] Necsulea A, Soumillon M, W. M. L. A. D. T. Z. U. B. J. G. F. K. H. (2014) The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**, 635–640.

[648] Flicek, P, Amode, M. R, Barrell, D, Beal, K, Billis, K, Brent, S, Carvalho-Silva, D, Clapham, P, Coates, G, Fitzgerald, S, Gil, L, Giron, C. G, Gordon, L, Hourlier, T, Hunt, S, Johnson, N, Juettemann, T, Kahari, A. K, Keenan, S, Kulesha, E, Martin, F. J, Maurel, T, McLaren, W. M, Murphy, D. N, Nag, R, Overduin, B, Pignatelli, M, Pritchard, B, Pritchard, E, Riat, H. S, Ruffier, M, Sheppard, D, Taylor, K, Thormann, A, Trevanion, S. J, Vullo, A, Wilder, S. P, Wilson, M, Zadissa, A, Aken, B. L, Birney, E, Cunningham, F, Harrow, J, Herrero, J, Hubbard, T. J. P, Kinsella, R, Muffato, M, Parker, A, Spudich, G, Yates, A, Zerbino, D. R, & Searle, S. M. J. (2013) Ensembl 2014. *Nucleic Acids Res.* **42**, D749–D755.

[649] Galperin, M. Y & Koonin, E. V. (2010) From complete genome sequence to 'complete'understanding? *Trends Biotechnol.* **28**, 398–406.

[650] (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.

[651] Kapranov, P, Laurent, G. S, Raz, T, Ozsolak, F, Reynolds, C. P, Sorensen, P. H, Reaman, G, Milos, P, Arceci, R. J, Thompson, J. F, & Triche, T. J. (2010) The majority of total nuclear-encoded non-ribosomal RNA in a human cell is 'dark matter' un-annotated RNA. *BMC Biol.* **8**, 149–149.

[652] Khalil, A. M & Rinn, J. L. (2011) RNA-protein interactions in human health and disease. *Semin. Cell Dev. Biol.* **22**, 359–365.

[653] Guttman, M, Amit, I, Garber, M, French, C, Lin, M. F, Feldser, D, Huarte, M, Carey, O. Z. B. W, Cassady, J. P, Cabili, M. N, Jaenisch, R, Mikkelsen, T. S, Jacks, T, Hacohen, N, Bernstein, B. E, Kellis, M, Regev, A, Rinn, J. L, & Lander, E. S. (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227.

[654] Guttman, M, Donaghey, J, Carey, B. W, Garber, M, Grenier, J. K, Munson, G, Young, G, Lucas, A. B, Ach, R, Bruhn, L, Yang, X, Amit, I, Meissner, A, Regev, A, Rinn, J. L, Root, D. E, & Lander, E. S. (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**, 295–300.

[655] Loewer, S, Cabili, M. N, Guttman, M, Loh, Y.-H, Thomas, K, Park, I. H, Garber, M, Curran, M, Onder, T, Agarwal, S, Manos, P. D, Datta, S, Lander, E. S, Schlaeger, T. M, Daley, G. Q, & Rinn, J. L. (2010) Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat. Genet.* **42**, 1113–1117.

[656] Sharma, S, Findlay, G. M, Bukwal, H. S, Oberdoerffer, S, Baust, B, Lib, Z, Schmidt, V, Hogan, P. G, Sacks, D. B, & Rao, A. (2011) Dephosphorylation of the

nuclear factor of activated T cells (NFAT) transcription factor is regulated by an RNA-protein scaffold complex. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 11381–11386.

[657] Guttman, M, Garber, M, Levin, J. Z, Donaghey, J, Robinson, J, Adiconis, X, Fan, L, Koziol, M. J, Gnirke, A, Nusbaum, C, Rinn, J. L, Lander, E. S, & Regev, A. (2010) Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* **28**, 503–510.

[658] Ponjavic, J, Ponting, C. P, & Lunter, G. (2007) Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res.* **17**, 556–565.

[659] Ulitsky, I, Shkumatava, A, Jan, C. H, Sive, H, & Bartel, D. P. (2011) Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* **147**, 1537–1550.

[660] Tsai, M.-C, Manor, O, Wan, Y, Mosammaparast, N, Wang, J. K, Lan, F, Shi, Y, Segal, E, & Chang, H. Y. (2010) Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science* **392**, 689–693.

[661] Frankish, A, Mudge, J. M, Thomas, M, & Harrow, J. (2012) The importance of identifying alternative splicing in vertebrate genome annotation. *Database* **2012**.

[662] Higa-Nakamine, S, Suzuki, T, Uechi, T, Chakraborty, A, Nakajima, Y, Nakamura, M, Hirano, N, Suzuki, T, & Kenmochi, N. (2011) Loss of ribosomal RNA modification causes developmental defects in zebrafish. *Nucleic Acids Res.* **40**, 391–398.

[663] Gibb, E. A, Vucic, E. A, Enfield, K. S. S, Stewart, G. L, Lonergan, K. M, Kennett, J. Y, Becker-Santos, D. D, MacAulay, C. E, Lam, S, Brown, C. J, & Lam, W. L. (2011) Human cancer long non-coding RNA transcriptomes. *PLoS ONE* **6**, e25915–e25915.

[664] Schneider, C, King, R. M, & Philipson, L. (1988) Genes specifically expressed at growth arrest of mammalian cells. *Cell* **54**, 787–793.

[665] Mourtada-Maarabouni, M, Hedge, V. L, Kirkham, L, Farzaneh, F, & Williams, G. T. (2010) Growth arrest in human T-cells is controlled by the non-coding RNA growth-arrest-specific transcript 5 (GAS5). *J. Cell Sci.* **123**, 1181–1181.

[666] Mourtada-Maarabouni, M, Hasan, A. M, Farzaneh, F, & Williams, G. T. (2010) Inhibition of human T-cell proliferation by mammalian target of rapamycin (mTOR) antagonists requires noncoding RNA growth-arrest-specific transcript 5 (GAS5). *Mol. Pharmacol.* **78**, 19–28.

[667] Mourtada-Maarabouni, M, Pickard, M. R, Hedge, V. L, Farzaneh, F, & Williams, G. T. (2009) GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. *Oncogene* **28**, 195–208.

[668] Pickard, M. R, Mourtada-Maarabouni, M, & Williams, G. T. (2013) Long noncoding RNA GAS5 regulates apoptosis in prostate cancer cell lines. *Biochim. Biophys. Acta* **1832**, 1613–1623.

[669] Meijsing, S. H, Pufall, M. A, So, A. Y, Bates, D. L, Chen, L, & Yamamoto, K. R. (2009) DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science* **324**, 407–410.

[670] Battiste, J. L, Mao, H, Rao, N. S, Tan, R, Muhandiram, D. R, Kay, L. E, Frankel, A. D, & Williamson, J. R. (1996) Helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE RNA complex. *Science* **273**, 1547–1551.

[671] Weeks, K & Crothers, D. (1993) Major groove accessibility of RNA. *Science* **261**, 1574–1577.

[672] Watson, L. C, Kuchenbecker, K. M, Schiller, B. J, Gross, J. D, Pufall, M. A, & Yamamoto, K. R. (2013) The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals. *Nat. Struct. Mol. Biol.* **20**, 876–83.

[673] de Vries, S. J, van Dijk, M, & Bonvin, A. M. J. J. (2010) The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* **5**, 883–897.

[674] Strobel, S & Cech, T. (1995) Minor groove recognition of the conserved G.U pair at the Tetrahymena ribozyme reaction site. *Science* **267**, 675–679.

[675] Magali, F & Schimmel, P. (1997) Subtle atomic group discrimination in the RNA minor groove. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 11291–11294.

[676] Romanuik, T. L, Wang, G, Morozova, O, Delaney, A, Marra, M. A, & Sadar, M. D. (2010) LNCaP Atlas: gene expression associated with in vivo progression to castration-recurrent prostate cancer. *BMC Med. Genomics* **3**, 43.

[677] John, S, Sabo, P. J, Thurman, R. E, Sung, M. H, Biddie, S. C, Johnson, T. A, Hager, G. L, & Stamatoyannopoulos, J. A. (2011) Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat. Genet.* **43**, 264–268.

[678] Shu, Z & Bevilacqua, P. C. (1999) Isolation and characterization of thermodynamically stable and unstable RNA hairpins from a Triloop Combinatorial Library. *Biochemistry* **38**, 15369–15379.

[679] Hirose, T & Steitz, J. A. (2001) Position within the host intron is critical for efficient processing of box C/D snoRNAs in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12914–12919.

[680] Batzer, M. A & Deininger, P. L. (2002) Alu repeats and human genomic diversity. *Nat. Rev. Genet.* **3**, 370–379.

[681] Flicek, P, Amode, M. R, Barrell, D, Beal, K, Brent, S, Chen, Y, Clapham, P, Coates, G, Fairley, S, Fitzgerald, S, Gordon, L, Hendrix, M, Hourlier, T, Johnson, N, Kahari, A, Keefe, D, Keenan, S, Kinsella, R, Kokocinski, F, Kulesha, E, Larsson, P, Longden, I, McLaren, W, Overduin, B, Pritchard, B, Riat, H. S, Rios, D, Ritchie, G. R. S, Ruffier, M, Schuster, M, Sobral, D, Spudich, G, Tang, Y. A, Trevanion, S, Vandrovcova, J, Vilella, A. J, White, S, Wilder, S. P, Zadissa, A, Zamora, J, Aken, B. L, Birney, E, Cunningham, F, Dunham, I, Durbin, R, Fernandez-Suarez, X. M, Herrero, J, Hubbard, T. J. P, Parker, A, Proctor, G, Vogel, J, & Searle, S. M. J. (2011) Ensembl 2011. *Nucleic Acids Res.* **39**, D800–D806.

[682] Presman, D. M, Ogara, M. F, Stortz, M, Alvarez, L. D, Pooley, J. R, Schiltz, R. L, Grntved, L, Johnson, T. A, Mittelstadt, P. R, Ashwell, J. D, Ganesan, S, Burton, G, Levi, V, Hager, G. L, & Pecci, A. (2014) Live cell imaging unveils multiple domain requirements for in vivo dimerization of the glucocorticoid receptor. *PLoS Biol.* **12**, e1001813.

[683] Cote, G. J & Gagel, R. F. (1986) Dexamethasone differentially affects the levels of calcitonin and calcitonin gene-related peptide mRNAs expressed in a human medullary thyroid carcinoma cell line. *J. Biol. Chem.* **261**, 15524–15528.

[684] Kosaki, A & Webster, N. J. (1993) Effect of dexamethasone on the alternative splicing of the insulin receptor mRNA and insulin action in HepG2 hepatoma cells. *J. Biol. Chem.* **268**, 21990–21996.

[685] Lai, G. J & McCobb, D. P. (2002) Opposing actions of adrenal androgens and glucocorticoids on alternative splicing of Slo potassium channels in bovine chromaffin cells. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7722–7727.

[686] Webster, J. C & Cidlowski, J. A. (1999) Mechanisms of Glucocorticoid-receptor-mediated repression of gene expression. *Trends Endocrinol. Metab.* **10**, 396–402.

[687] Schorderet, P & Duboule, D. (2011) Structural and functional differences in the long non-coding RNA hotair in mouse and human. *PLoS Genet.* **7**, e1002071.

[688] Easton, L. E, Shibata, Y, & Lukavsky, P. J. (2010) Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography. *RNA* **16**, 647–653.

[689] Wahid, A. M, Coventry, V. K, & Conn, G. L. (2009) The PKR-binding domain of adenovirus VA RNAI exists as a mixture of two functionally non-equivalent structures. *Nucleic Acids Res.* **37**, 5830–5837.

[690] Emsley, P & Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132.

[691] Adams, P. D, Afonine, P. V, Bunkoczi, G, Chen, V. B, Davis, I. W, Echols, N, Headd, J. J, Hung, L. W, Kapral, G. J, Grosse-Kunstleve, R. W, McCoy, A. J, Moriarty, N. W, Oeffner, R, Read, R. J, Richardson, D. C, Richardson, J. S, Terwilliger, T. C, & Zwart, P. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–21.

[692] Williamson, M. P. (2013) Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Mag. Res. Sp.* **73**, 1–16.

[693] Johnson, B. A. (2004) in *Protein NMR Techniques.* (Springer), pp. 313–352.

[694] Case, D. A, III, T. E. C, Darden, T, Gohlke, H, Luo, R, Jr., K. M. M, Onufriev, A, Simmerling, C, Wang, B, & Woods, R. J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688.

[695] Jorgensen, W. L, Chandrasekhar, J, Madura, J. D, Impey, R. W, & Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Biol. Chem.* **79**, 926–935.

[696] Götz, A. W, Williamson, M. J, Xu, D, Poole, D, Grand, S. L, & Walker, R. C. (2012) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **8**, 1542–1555.

[697] Le Grand, S, Götz, A. W, & Walker, R. C. (2013) SPFP: Speed without compromise‘A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* **184**, 374–380.

[698] Shang, Y & Simmerling, C. (2012) Molecular dynamics applied in drug discovery: the case of HIV-1 protease. *Methods Mol. Biol.* **819**, 527–549.

[699] Allen, M. P & Tildesley, D. J. (1987). *Computer Simulation of Liquids.*

[700] Ryckaert, J. P, Ciccotti, G, & Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341.

[701] Reese, M, Eeckman, F, Kulp, D, & Haussler, D. (1997) Improved Splice Site Detection in Genie. *J. Comput. Biol.* **4**, 311–323.

[702] Tamura, K, Peterson, D, Peterson, N, Stecher, G, Nei, M, & Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739.

[703] Chrousos, G. P. (2009) Stress and disorders of the stress system. *Nat. Rev. Endocrinol.* **5**, 374–381.

[704] Pikulkaew, S, Benato, F, Celeghin, A, Zucal, C, Skobo, T, Colombo, L, & Dalla Valle, L. (2011) The knockdown of maternal glucocorticoid receptor mRNA alters embryo development in zebrafish. *Dev. Dyn.* **240**, 874–889.

[705] Rose, A. J, Vegiopoulos, A, & Herzig, S. (2010) Role of glucocorticoids and the glucocorticoid receptor in metabolism: insights from genetic manipulations. *J. Steroid Biochem.* **122**, 10–20.

[706] Baschant, U & Tuckermann, J. (2010) The role of the glucocorticoid receptor in inflammation and immunity. *J. Steroid Biochem.* **120**, 69–75.

[707] Langlais, D, Couture, C, Balsalobre, A, & Drouin, J. (2012) The Stat3/GR interaction code: predictive value of direct/indirect DNA recruitment for transcription outcome. *Mol. Cell* **47**, 38–49.

[708] Tuckermann, J. P, Reichardt, H. M, & Arribas, R. (1999) The DNA binding-independent function of the glucocorticoid receptor mediates repression of AP-1-dependent genes in skin. *J. Cell Biol.* **147**, 1365–1370.

[709] Schcke, H, Döcke, W. D, & Asadullah, K. (2002) Mechanisms involved in the side effects of glucocorticoids. *Pharmacol. Ther.* **96**, 23–43.

[710] Strähle, U, Klock, G, & Schütz, G. (1987) A DNA sequence of 15 base pairs is sufficient to mediate both glucocorticoid and progesterone induction of gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7871–7875.

[711] Reardon, C, Lechmann, M, Brustle, A, Gareau, M. G, Shuman, N, Philpott, D, Ziegler, S. F, & Mak, T. W. (2011) Thymic stromal lymphopoietin-induced expression of the endogenous inhibitory enzyme SLPI mediates recovery from colonic inflammation. *Immunity* **35**, 223–235.

[712] Rochman, Y, Kashyap, M, Robinson, G. W, Sakamoto, K, Gomez-Rodriguez, J, Wagner, K. U, & Leonard, W. J. (2010) Thymic stromal lymphopoietin-mediated STAT5 phosphorylation via kinases JAK1 and JAK2 reveals a key difference from IL-7-induced signaling. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 19455–19460.

[713] Siracusa, M. C, Saenz, S. A, Hill, D. A, Kim, B. S, Headley, M. B, Doering, T. A, Wherry, E. J, Jessup, H. K, Siegel, L. A, Kambayashi, T, Dudek, E. C, Kubo, M, Cianferoni, A, Spergel, J. M, Ziegler, S. F, Comeau, M. R, & Artis, D. (2011) TSLP promotes interleukin-3-independent basophil haematopoiesis and type 2 inflammation. *Nature* **477**, 229–233.

[714] Al-Shami, A, Spolski, R, Kelly, J, Keane-Myers, A, & Leonard, W. J. (2005) A role for TSLP in the development of inflammation in an asthma model. *J. Exp. Med.* **202**, 829–839.

[715] Koyama, K, Ozawa, T, Hatsushika, K, Ando, T, Takano, S, Wako, M, Suenaga, F, Ohnuma, Y, Ohba, T, Katoh, R, Sugiyama, H, Hamada, Y, Ogawa, H, Okumura, K, & Nakao, A. (2007) A possible role for TSLP in inflammatory arthritis. *Biochem. Biophys. Res. Commun.* **357**, 99–104.

[716] Soumelis, V, Reche, P. A, Kanzler, H, Yuan, W, Edward, G, Homey, B, Gilliet, M, Ho, S, Antonenko, S, Lauerma, A, Smith, K, Gorman, D, Zurawski, S, Abrams, J, Menon, S, McClanahan, T, de Waal-Malefyt Rd, R, Bazan, F, Kastelein, R. A, & Liu, Y. J. (2002) Human epithelial cells trigger dendritic cell mediated allergic inflammation by producing TSLP. *Nat. Immunol.* **3**, 673–680.

[717] Taylor, B. C, Zaph, C, Troy, A. E, Du, Y, Guild, K. J, Comeau, M. R, & Artis, D. (2009) TSLP regulates intestinal immunity and inflammation in mouse models of helminth infection and colitis. *J. Exp. Med.* **206**, 655–667.

[718] Yoo, J, Omori, M, Gyarmati, D, Zhou, B, Aye, T, Brewer, A, Comeau, M. R, Campbell, D. J, & Ziegler, S. F. (2005) Spontaneous atopic dermatitis in mice expressing an inducible thymic stromal lymphopoietin transgene specifically in the skin. *J. Exp. Med.* **202**, 541–549.

[719] Heck, S, Kullmann, M, Gast, A, Ponta, H, Rahmsdorf, H. J, Herrlich, P, & Cato, A. C. (1994) A distinct modulating domain in glucocorticoid receptor monomers in the repression of activity of the transcription factor AP-1. *EMBO J.* **13**, 4087–4095.

[720] Baniahmad, A, Steiner, C, Köhne, A, & Renkawitz, R. (1990) Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site. *Cell* **61**, 505–514.

[721] Tini, M, Otulakowski, G, Breitman, M. L, Tsui, L. C, & Giguere, V. (1993) 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.* **7**, 295–307.

[722] Krissinel, E & Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797.

[723] Heneghan, A. F, Connaghan-Jones, K. D, Miura, M. T, & Bain, D. L. (2007) Coactivator assembly at the promoter: efficient recruitment of SRC2 is coupled to cooperative DNA binding by the progesterone receptor. *Biochemistry* **46**, 11023–11032.

[724] Tsai, S. Y, Carlstedt-Duke, J, Weigel, N. L, Dahlman, K, Gustafsson, J. A, Tsai, M. J, & O'Malley, B. W. (1988) Molecular interactions of steroid hormone receptor with its enhancer element: evidence for receptor dimer formation. *Cell* **55**, 361–369.

[725] Robblee, J. P, Miura, M. T, & Bain, D. L. (2012) Glucocorticoid receptor-promoter interactions: energetic dissection suggests a framework for the specificity of steroid receptor-mediated gene regulation. *Biochemistry* **51**, 4463–4472.

[726] Brélivet, Y, Rochel, N, & Moras, D. (2012) Structural analysis of nuclear receptors: From isolated domains to integral proteins. *Mol. Cell. Endocrinol.* **348**, 466–473.

[727] Nwachukwu, J. C & Nettles, K. W. (2012) The nuclear receptor signalling scaffold: insights from full-length structures. *EMBO J.* **31**, 251–253.

[728] Surjit, M. (2011) Widespread negative response elements mediate direct repression by agonist-liganded glucocorticoid receptor. *Cell* **145**, 224–241.

[729] Liu, W, Wang, J, Yu, G, & Pearce, D. (1996) Steroid receptor transcriptional synergy is potentiated by disruption of the DNA-binding domain dimer interface. *Mol. Endocrinol.* **10**, 1399–1406.

[730] Rogatsky, I, Hittleman, A. B, Pearce, D, & Garabedian, M. J. (1999) Distinct glucocorticoid receptor transcriptional regulatory surfaces mediate the cytotoxic and cytostatic effects of glucocorticoids. *Mol. Cell. Biol.* **19**, 5036–5049.

[731] Irusen, E, Matthews, J. G, Takahashi, A, Barnes, P. J, Chung, K. F, & Adcock, I. M. (2002) p38 Mitogen-activated protein kinase-induced glucocorticoid receptor phosphorylation reduces its activity: role in steroid-insensitive asthma. *J. Allergy Clin. Immunol.* **109**, 649–657.

[732] McGowan, P. O, Sasaki, A, D'Alessio, A. C, Dymov, S, Labonte, B, Szyf, M, Turecki, G, & Meaney, M. J. (2009) Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nat. Neurosci.* **12**, 342–348.

[733] Okret, S, Poellinger, L, Dong, Y, & Gustafsson, J. A. (1986) Down-regulation of glucocorticoid receptor mRNA by glucocorticoid hormones and recognition by the receptor of a specific binding sequence within a receptor cDNA clone. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5899–5903.

[734] Oster, H, Damerow, S, Kiessling, S, Jakubcakova, V, Abraham, D, Tian, J, Hoffmann, M. W, & Eichele, G. (2006) The circadian rhythm of glucocorticoids is regulated by a gating mechanism residing in the adrenal cortical clock. *Cell Metab.* **4**, 163–173.

[735] Ingraham, H. A, Flynn, S. E, Voss, J. W, Albert, V. R, Kapiloff, M. S, Wilson, L, & Rosenfeld, M. G. (1990) The POU-specific domain of Pit-1 is essential for sequence-specific, high affinity DNA binding and DNA-dependent Pit-1-Pit-1 interactions. *Cell* **61**, 1021–1033.

[736] Scully, K. M, Jacobson, E. M, Jepsen, K, Lunyak, V, Viadiu, H, Carriere, C, Rose, D. W, Hooshmand, F, Aggarwal, A. K, & Rosenfeld, M. G. (2000) Allosteric effects of Pit-1 DNA sites on long-term repression in cell type specification. *Science* **290**, 1127–1131.

[737] Dewint, P, Gossye, V, De Bosscher, K, Vanden Berghe, W, Van Beneden, K, Deforce, D, Van Calenbergh, S, Muller-Ladner, U, Vander Cruyssen, B, Verbruggen, G, Haegeman, G, & Elewaut, D. (2008) A plant-derived ligand favoring monomeric glucocorticoid receptor conformation with impaired transactivation potential attenuates collagen-induced arthritis. *J. Immunol.* **180**, 2608–2615.

[738] Drouin, J, Sun, Y. L, Chamberland, M, Gauthier, Y, De Lean, A, Nemer, M, & Schmidt, T. J. (1993) Novel glucocorticoid receptor complex with DNA element of the hormone-repressed POMC gene. *EMBO J.* **12**, 145–156.

[739] Robertson, S, Allie-Reid, F, Vanden Berghe, W, Visser, K, Binder, A, Africander, D, Vismer, M, De Bosscher, K, Hapgood, J, Haegeman, G, & Louw, A. (2010) Abrogation of glucocorticoid receptor dimerization correlates with dissociated glucocorticoid behavior of compound A. *J. Biol. Chem.* **285**, 8061–8075.

[740] Falkenstein, E, Tillmann, H. C, Christ, M, Feuring, M, & Wehling, M. (2000) Multiple actions of steroid hormones–a focus on rapid, nongenomic effects. *Pharmacol. Rev.* **52**, 513–56.

[741] Tsai, M & O'Malley, B. W. (1994) Molecular mechanisms of action of steroid/thyroid receptor superfamily members. *Annu Rev Biochem* **63**, 451–486.

[742] Rastinejad, F, Huang, P, Chandra, V, & Khorasanizadeh, S. (2013) Understanding nuclear receptor form and function using structural biology. *J. Mol. Endocrinol.* **51**, T1–T21.

[743] Bridgham, J. T, Carroll, S. M, & Thornton, J. W. (2006) Evolution of hormone-receptor complexity by molecular exploitation. *Science* **312**, 97–101.

[744] Herman, J. P, Patel, P. D, Akil, H, & Watson, S. J. (1989) Localization and regulation of glucocorticoid and mineralocorticoid receptor messenger RNAs in the hippocampal formation of the rat. *Mol. Endocrinol.* **3**, 1886–94.

[745] Lombes, M, Oblin, M. E, Gasc, J. M, Baulieu, E. E, Farman, N, & Bonvalet, J. P. (1992) Immunohistochemical and biochemical evidence for a cardiovascular mineralocorticoid receptor. *Circ. Res.* **71**, 503–10.

[746] Marver, D, Stewart, J, Funder, J. W, Feldman, D, & Edelman, I. S. (1974) Renal aldosterone receptors: studies with (^{3}H)aldosterone and the anti-mineralocorticoid (^{3}H)spirolactone (SC-26304). *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1431–5.

[747] Van Eekelen, J. A, Jiang, W, De Kloet, E. R, & Bohn, M. C. (1988) Distribution of the mineralocorticoid and the glucocorticoid receptor mRNAs in the rat hippocampus. *J. Neurosci. Res.* **21**, 88–94.

[748] DeRijk, R. H, Wust, S, Meijer, O. C, Zennaro, M. C, Federenko, I. S, Hellhammer, D. H, Giacchetti, G, Vreugdenhil, E, Zitman, F. G, & de Kloet, E. R. (2006) A common polymorphism in the mineralocorticoid receptor modulates stress responsiveness. *J. Clin. Endocrinol. Metab.* **91**, 5083–9.

[749] Joels, M, Karst, H, DeRijk, R, & de Kloet, E. R. (2008) The coming out of the brain mineralocorticoid receptor. *Trends Neurosci* **31**, 1–7.

[750] Berger, S, Bleich, M, Schmid, W, Cole, T. J, Peters, J, Watanabe, H, Kriz, W, Warth, R, Greger, R, & Schutz, G. (1998) Mineralocorticoid receptor knockout mice: pathophysiology of Na+ metabolism. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9424–9.

[751] McGraw, A. P, McCurley, A, Preston, I. R, & Jaffe, I. Z. (2013) Mineralocorticoid receptors in vascular disease: connecting molecular pathways to clinical implications. *Curr Atheroscler Rep* **15**, 340.

[752] Ivanes, F, Susen, S, Mouquet, F, Pigny, P, Cuilleret, F, Sautiere, K, Collet, J. P, Beygui, F, Hennache, B, Ennezat, P. V, Juthier, F, Richard, F, Dallongeville, J, Hillaert, M. A, Doevendans, P. A, Jude, B, Bertrand, M, Montalescot, G, & Van Belle, E. (2012) Aldosterone, mortality, and acute ischaemic events in coronary artery disease patients outside the setting of acute myocardial infarction or heart failure. *Eur. Heart J.* **33**, 191–202.

[753] Pitt, B, Zannad, F, Remme, W. J, Cody, R, Castaigne, A, Perez, A, Palensky, J, & Wittes, J. (1999) The effect of spironolactone on morbidity and mortality in patients with severe heart failure. *N. Engl. J. Med.* **341**, 709–17.

[754] Ueda, K, Fujiki, K, Shirahige, K, Gomez-Sanchez, C. E, Fujita, T, Nangaku, M, & Nagase, M. (2014) Genome-wide analysis of murine renal distal convoluted tubular cells for the target genes of mineralocorticoid receptor. *Biochem. Biophys. Res. Commun.* **445**, 132–7.

[755] Barnes, P. J. (1998) Anti-inflammatory actions of glucocorticoids: molecular mechanisms. *Clin. Sci. (Lond)* **94**, 557–72.

[756] Chantong, B, Kratschmar, D. V, Nashev, L. G, Balazs, Z, & Odermatt, A. (2012) Mineralocorticoid and glucocorticoid receptors differentially regulate NF-B activity and pro-inflammatory cytokine production in murine BV-2 microglial cells. *J. Neuroinflammation* **9**, 260.

[757] Rickard, A. J, Morgan, J, Bienvenu, L. A, Fletcher, E. K, Cranston, G. A, Shen, J. Z, Reichelt, M. E, Delbridge, L. M, & Young, M. J. (2012) Cardiomyocyte mineralocorticoid receptors are essential for deoxycorticosterone/salt-mediated inflammation and cardiac fibrosis. *Hypertension* **60**, 1443–50.

[758] Young, M. J. (2008) Mechanisms of mineralocorticoid receptor-mediated cardiac fibrosis and vascular inflammation. *Curr. Opin. Nephrol. Hy.* **17**, 174–180.

[759] Caldenhoven, E, Liden, J, Wissink, S, Van de Stolpe, A, Raaijmakers, J, Koenderman, L, Okret, S, Gustafsson, J. A, & Van der Saag, P. T. (1995) Negative cross-talk between RelA and the glucocorticoid receptor: a possible mechanism for the antiinflammatory action of glucocorticoids. *Mol. Endocrinol.* **9**, 401–12.

[760] Caprio, M, Newfell, B. G, la Sala, A, Baur, W, Fabbri, A, Rosano, G, Mendelsohn, M. E, & Jaffe, I. Z. (2008) Functional mineralocorticoid receptors in human vascular endothelial cells regulate intercellular adhesion molecule-1 expression and promote leukocyte adhesion. *Circ. Res.* **102**, 1359–1367.

[761] Sartorato, P, Cluzeaud, F, Fagart, J, Viengchareun, S, Lombs, M, & Zennaro, M.-C. (2004) New naturally occurring missense mutations of the human mineralocorticoid receptor disclose important residues involved in dynamic interactions with deoxyribonucleic acid, intracellular trafficking, and ligand binding. *Mol. Endocrinol.* **18**, 2151–2165.

[762] Boyer, M, Poujol, N, Margeat, E, & Royer, C. A. (2000) Quantitative characterization of the interaction between purified human estrogen receptor alpha and DNA using fluorescence anisotropy. *Nucleic Acids Res.* **28**, 2494–502.

[763] Schena, M, Freedman, L. P, & Yamamoto, K. R. (1989) Mutations in the glucocorticoid receptor zinc finger region that distinguish interdigitated DNA binding and transcriptional enhancement activities. *Genes Dev.* **3**, 1590–1601.

[764] Thomas-Chollier, M, Watson, L. C, Cooper, S. B, Pufall, M. A, Liu, J. S, Borzym, K, Vingron, M, Yamamoto, K. R, & Meijsing, S. H. (2013) A naturally occurring insertion of a single amino acid rewires transcriptional regulation by glucocorticoid receptor isoforms. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17826–31.

[765] Kim, S, Brostromer, E, Xing, D, Jin, J, Chong, S, Ge, H, Wang, S, Gu, C, Yang, L, Gao, Y. Q, Su, X. d, Sun, Y, & Xie, X. S. (2013) Probing allostery through DNA. *Science* **339**, 816–819.

[766] Pujo, L, Fagart, J, Gary, F, Papadimitriou, D. T, Clas, A, Jeunematre, X, & Zennaro, M.-C. (2007) Mineralocorticoid receptor mutations are the principal cause of renal type 1 pseudohypoaldosteronism. *Hum. Mutat.* **28**, 33–40.

[767] Sartorato, P, Lapeyraque, A.-L, Armanini, D, Kuhnle, U, Khaldi, Y, Salomon, R, Abadie, V, Di Battista, E, Naselli, A, Racine, A, Bosio, M, Caprio, M, Poulet-Young, V, Chabrolle, J.-P, Niaudet, P, De Gennes, C, Lecornec, M.-H, Poisson, E, Fusco, A. M, Loli, P, Lombs, M, & Zennaro, M.-C. (2003) Different inactivating

mutations of the mineralocorticoid receptor in fourteen families affected by type I pseudohypoaldosteronism. *J. Clin. Endocrinol. Metab.* **88**, 2508–2517.

[768] Daniel, A. R, Hagan, C. R, & Lange, C. A. (2011) Progesterone receptor action: defining a role in breast cancer. *Expert Rev. Endocrinol. Metab.* **6**, 359–369.

[769] Liang, J & Shang, Y. (2013) Estrogen and Cancer. *Annu Rev Physiol* **75**, 225–240.

[770] Tindall, D. J & Lonergan, P. E. (2011) Androgen receptor signaling in prostate cancer development and progression. *J. Carcinog* **10**, 20.

[771] Arora, V. K, Schenkein, E, Murali, R, Subudhi, S, Wongvipat, J, Balbas, M, Shah, N, Cai, L, Efstathiou, E, Logothetis, C, Zheng, D, & Sawyers, C. (2013) Glucocorticoid receptor confers resistance to antiandrogens by bypassing androgen receptor blockade. *Cell* **155**, 1309–1322.

[772] Leo, J. C, Guo, C, Woon, C. T, Aw, S. E, & Lin, V. C. (2004) Glucocorticoid and mineralocorticoid cross-talk with progesterone receptor to induce focal adhesion and growth inhibition in breast cancer cells. *Endocrinology* **145**, 1314–21.

[773] Moschetta, A, Tiberio, L, Nascimbeni, R, Villanacci, V, Casella, C, Fra, A, Vezzoli, V, Furlan, L, Meyer, G, Parrinello, G, Baroni, M. D, Salerni, B, & Schiaffonati, L. (2013) The Decrease of Mineralcorticoid Receptor Drives Angiogenic Pathways in Colorectal Cancer. *PLoS ONE* **8**, e59410.

[774] Vilasco, M, Communal, L, Mourra, N, Courtin, A, Forgez, P, & Gompel, A. (2011) Glucocorticoid receptor and breast cancer. *Breast Cancer Res. Treat.* **130**, 1–10.

[775] Cerami, E, Gao, J, Dogrusoz, U, Gross, B. E, Sumer, S. O, Aksoy, B. A, Jacobsen, A, Byrne, C. J, Heuer, M. L, Larsson, E, Antipin, Y, Reva, B, Goldberg, A. P, Sander, C, & Schultz, N. (2012) The cBio cancer genomics portal: an open

platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404.

[776] Gao, J, Aksoy, B. A, Dogrusoz, U, Dresdner, G, Gross, B, Sumer, S. O, Sun, Y, Jacobsen, A, Sinha, R, Larsson, E, Cerami, E, Sander, C, & Schultz, N. (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1–pl1.

[777] Zilliacus, J, Carlstedt-Duke, J, Gustafsson, J. A, & Wright, A. P. (1994) Evolution of distinct DNA-binding specificities within the nuclear receptor family of transcription factors. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4175–9.

[778] Geller, D. S, Farhi, A, Pinkerton, N, Fradley, M, Moritz, M, Spitzer, A, Meinke, G, Tsai, F. T, Sigler, P. B, & Lifton, R. P. (2000) Activating mineralocorticoid receptor mutation in hypertension exacerbated by pregnancy. *Science* **289**, 119–123.

[779] Mune, T, Rogerson, F. M, Nikkil, H, Agarwal, A. K, & White, P. C. (1995) Human hypertension caused by mutations in the kidney isozyme of 11hydroxysteroid dehydrogenase. *Nat. Genet.* **10**, 394–399.

[780] Di Fabio, F, Alvarado, C, Majdan, A, Gologan, A, Voda, L, Mitmaker, E, Beitel, L. K, Gordon, P. H, & Trifiro, M. (2007) Underexpression of mineralocorticoid receptor in colorectal carcinomas and association with VEGFR-2 overexpression. *J. Gastrointest. Surg.* **11**, 1521–8.

[781] Jeong, Y, Xie, Y, Xiao, G, Behrens, C, Girard, L, Wistuba, I, Minna, J. D, & Mangelsdorf, D. J. (2010) Nuclear receptor expression defines a set of prognostic biomarkers for lung cancer. *PLoS Med.* **7**, e1000378.

[782] Walther, R. F, Atlas, E, Carrigan, A, Rouleau, Y, Edgecombe, A, Visentin, L, Lamprecht, C, Addicks, G. C, Hache, R. J. G, & Lefebvre, Y. A. (2005) A

serine/threonine-rich motif is one of three nuclear localization signals that determine unidirectional transport of the mineralocorticoid receptor to the nucleus. *J. Biol. Chem.* **280**, 17549–17561.

[783] Faresse, N. (2014) Post-translational modifications of the mineralocorticoid receptor: How to dress the receptor according to the circumstances? *J Steroid Biochem. Mol. Biol.* **143**, 334–342.

[784] Sherry, S. T, Ward, M. H, Kholodov, M, Baker, J, Phan, L, Smigielski, E. M, & Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–11.

[785] Liu, W, Wang, J, Sauter, N. K, & Pearce, D. (1995) Steroid receptor heterodimerization demonstrated in vitro and in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 12480–4.

[786] Fuse, H, Kitagawa, H, & Kato, S. (2000) Characterization of transactivational property and coactivator mediation of rat mineralocorticoid receptor activation function-1 (AF-1). *Mol. Endocrinol.* **14**, 889–99.

[787] Morohashi, K, Baba, T, & Tanaka, M. (2013) Steroid hormones and the development of reproductive organs. *Sex. Dev.* **7**, 61–79.

[788] Blaustein, J. (2012) Steroid Hormone Receptors: Long- and Short-term Integrators of the Internal Milieu and the External Environment. *Horm Metab Res* **44**, 563–568.

[789] Evans, R & Mangelsdorf, D. (2014) Nuclear Receptors, RXR, and the Big Bang. *Cell* **157**, 255–266.

[790] Overington, J. P, Al-Lazikani, B, & Hopkins, A. L. (2006) How many drug targets are there? *Nat. Rev. Drug Discov.* **5**, 993–996.

[791] Webster, N. J, Green, S, Jin, J. R, & Chambon, P. (1988) The hormone-binding domains of the estrogen and glucocorticoid receptors contain an inducible transcription activation function. *Cell* **54**, 199–207.

[792] Hollenberg, S. M, Giguere, V, Segui, P, & Evans, R. M. (1987) Colocalization of DNA-binding and transcriptional activation functions in the human glucocorticoid receptor. *Cell* **49**, 39?46.

[793] Schüle, R, Rangarajan, P, Kliewer, S, Ransone, L. J, Bolado, J, Yang, N, Verma, I. M, & Evans, R. M. (1990) Functional antagonism between oncoprotein c-Jun and the glucocorticoid receptor. *Cell* **62**, 1217–1226.

[794] Baumann, H, Paulsen, K, Kovács, H, Berglund, H, Wright, A. P, Gustafsson, J. A, & Härd, T. (1993) Refined solution structure of the glucocorticoid receptor DNA-binding domain. *Biochemistry* **32**, 13463–13471.

[795] Thornton, J. W. (2001) Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5671–5676.

[796] Aharoni, A, Gaidukov, L, Khersonsky, O, Gould, S. M, Roodveldt, C, & Tawfik, D. S. (2004) The evolvability of promiscuous protein functions. *Nat. Genet.* **37**, 73–6.

[797] Carroll, S. M, Ortlund, E. A, & Thornton, J. W. (2011) Mechanisms for the Evolution of a Derived Function in the Ancestral Glucocorticoid Receptor. *PLoS Genet.* **7**, e1002117.

[798] McKeown, A, Bridgham, J, Anderson, D, Murphy, M, Ortlund, E, & Thornton, J. (2014) Evolution of DNA Specificity in a Transcription Factor Family Produced a New Gene Regulatory Module. *Cell* **159**, 58–68.

[799] Beato, M. (1989) Gene regulation by steroid hormones. *Cell* **56**, 335–344.

[800] Breen, M. S, Kemena, C, Vlasov, P. K, Notredame, C, & Kondrashov, F. A. (2012) Epistasis as the primary factor in molecular evolution. *Nature* **490**, 535–538.

[801] Bloom, J. D & Arnold, F. H. (2009) In the light of directed evolution: Pathways of adaptive protein evolution. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 9995–10000.

[802] Soskine, M & Tawfik, D. S. (2010) Mutational effects and the evolution of new protein functions. *Nat. Rev. Genet.* **11**, 572–582.

[803] Eick, G. N, Colucci, J. K, Harms, M. J, Ortlund, E. A, & Thornton, J. W. (2012) Evolution of Minimal Specificity and Promiscuity in Steroid Hormone Receptors. *PLoS Genet.* **8**, e1003072.

[804] Romero, P. A & Arnold, F. H. (2009) Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10**, 866–876.

[805] Bridgham, J. T, Ortlund, E. A, & Thornton, J. W. (2009) An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* **461**, 515?519.

[806] Heck, S, Bender, K, Kullman, M, M., G, Herrlich, P, & Cato, A. C. (1997) Ikappa Balpha -independent downregulation of NF-kappa B activity by glucocorticoid receptor. *EMBO J.* **16**, 4698–4707.

[807] Sharma, S & Lichtenstein, A. (2008) Dexamethasone-induced apoptotic mechanisms in myeloma cells investigated by analysis of mutant glucocorticoid receptors. *Blood* **112**, 1338–1345.

[808] Tao, Y, Williams-Skipp, C, & Scheinman, R. I. (2000) Mapping of Glucocorticoid Receptor DNA Binding Domain Surfaces Contributing to Transrepression of NF-κB and Induction of Apoptosis. *J. Biol. Chem.* **276**, 2329–2332.

[809] Liberman, A. C, Antunica-Noguerol, M, Ferraz-de Paula, V, Palermo-Neto, J, Castro, C. N, Druker, J, Holsboer, F, Perone, M. J, Gerlo, S, De Bosscher, K, & et al. (2012) Compound A, a Dissociated Glucocorticoid Receptor Modulator, Inhibits T-bet (Th1) and Induces GATA-3 (Th2) Activity in Immune Cells. *PLoS ONE* **7**, e35155.

[810] McCoy, A. J, Grosse-Kunstleve, R. W, Adams, P. D, Winn, M. D, Storoni, L. C, & Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674.

[811] Pettersen, E. F, Goddard, T. D, Huang, C. C, Couch, G. S, Greenblatt, D. M, Meng, E. C, & Ferrin, T. E. (2004) UCSF Chimera:A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612.

[812] Joosten, R. P, Long, F, Murshudov, G. N, & Perrakis, A. (2014) The PDBREDO server for macromolecular structure model optimization. *IUCrJ* **1**, 213–220.

[813] Jorgensen, W. L, Chandrasekhar, J, Madura, J. D, Impey, R. W, & Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926.

[814] Wang, J, Wang, W, Kollman, P. A, & Case, D. A. (2006) Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **25**, 247–260.

[815] Peréz, A, Marchán, I, Svozil, D, Sponer, J, Cheatham, T. E, Laughton, C. A, & Orozco, M. (2007) Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ conformers. *Biophys. J.* **92**, 3817?3829.

[816] Wang, J, Cieplak, P, & Kollman, P. A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **21**, 1049–1074.

[817] Phillips, J. C, Braun, R, Wang, W, Gumbart, J, Tajkhorshid, E, Villa, E, Chipot, C, Skeel, R. D, Kalé, L, & Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802.

[818] Tuckerman, M, Berne, B. J, & Martyna, G. J. (1992) Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **97**, 1990.

[819] Forester, T. R & Smith, W. (1998) SHAKE, rattle, and roll: efficient constraint algorithms for linked rigid bodies. *J. Comput. Chem.* **19**, 102–111.

[820] Essmann, U, Perera, L, Berkowitz, M. L, Darden, T, Lee, H, & Pedersen, L. G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577.

[821] Sethi, A, Eargle, J, Black, A. A, & Luthey-Schulten, Z. (2009) Dynamical networks in tRNA:protein complexes. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6620–6625.

[822] Humphrey, W, Dalke, A, & Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graphics* **14**, 33–38.

[823] Newman, M. E. J. (2006) Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8577–8582.

[824] Newman, M. E. (2005) A measure of betweenness centrality based on random walks. *Social networks* **27**, 39–54.

[825] Floyd, R. W. (1962) Algorithm 97: shortest path. *Communications of the ACM* **5**, 345.

[826] Newton, R. (2000) Molecular mechanisms of glucocorticoid action: what is important? *Thorax* **55**, 603–613.

[827] Liberman, A. C, Refojo, D, Druker, J, Toscano, M, Rein, T, Holsboer, F, & Arzt, E. (2007) The activated glucocorticoid receptor inhibits the transcription factor T-bet by direct protein-protein interaction. *FASEB J.* **21**, 1177–1188.

[828] Bledsoe, R. K, Montana, V. G, Stanley, T. B, Delves, C. J, Apolito, C. J, McKee, D. D, Consler, T. G, Parks, D. J, Stewart, E. L, Willson, T. M, & et al. (2002) Crystal Structure of the Glucocorticoid Receptor Ligand Binding Domain Reveals a Novel Mode of Receptor Dimerization and Coactivator Recognition. *Cell* **110**, 93?105.

[829] Scheinman, R. I, Cogswell, P. C, Lofquist, A. K, & Baldwin, A. S. (1995) Role of Transcriptional Activation of IkappaBalpha in Mediation of Immunosuppression by Glucocorticoi ds. *Science* **270**, 283–286.

[830] Glass, C. K & Saijo, K. (2010) Nuclear receptor transrepression pathways that regulate inflammation in macrophages and T cells. *Nat. Rev. Immunol.* **10**, 365–376.

[831] Uhlenhaut, N, Barish, G, Yu, R, Downes, M, Karunasiri, M, Liddle, C, Schwalie, P, H?bner, N, & Evans, R. (2013) Insights into Negative Regulation by the Glucocorticoid Receptor from Genome-wide Profiling of Inflammatory Cistromes. *Mol. Cell* **49**, 158–171.

[832] Carson-Jurica, M. A, Schrader, W. T, & O'Malley, B. W. (1990) Steroid receptor family: structure and functions. *Endocr. Rev.* **11**, 201–220.

[833] Reichardt, H. M, Kaestner, K. H, Tuckermann, J, Kretz, O, Wessely, O, Bock, R, Gass, P, Schmid, W, Herrlich, P, Angel, P, et al. (1998) DNA binding of the glucocorticoid receptor is not essential for survival. *Cell* **93**, 531–541.

[834] Rao, N. A. S, McCalman, M. T, Moulos, P, Francoijs, K.-J, Chatziioannou, A, Kolisis, F. N, Alexis, M. N, Mitsiou, D. J, & Stunnenberg, H. G. (2011) Coactivation of GR and NFKB alters the repertoire of their binding sites and target genes. *Genome Res.* **21**, 1404–1416.

[835] Chen, F. E, Huang, D.-B, Chen, Y.-Q, & Ghosh, G. (1998) Crystal structure of p50/p65 heterodimer of transcription factor NF-κB bound to DNA. *Nature* **391**, 410–413.

[836] Huxford, T, Hoffmann, A, & Ghosh, G. (2011) in *NF-kB in Health and Disease*. (Springer), pp. 1–24.

[837] Gu, G, Wang, T, Yang, Y, Xu, X, & Wang, J. (2013) An improved SELEX-Seq strategy for characterizing DNA-binding specificity of transcription factor: NF-kappaB as an example. *PLoS ONE* **8**.

[838] Berkowitz, B, Huang, D.-B, Chen-Park, F. E, Sigler, P. B, & Ghosh, G. (2002) The x-ray crystal structure of the NF-κB p50· p65 heterodimer bound to the interferon β-κB site. *J. Biol. Chem.* **277**, 24694–24700.

[839] Gutierrez, O, Pipaon, C, Inohara, N, Fontalba, A, Ogura, Y, Prosper, F, Nunez, G, & Fernandez-Luna, J. L. (2002) Induction of Nod2 in Myelomonocytic and Intestinal Epithelial Cells via Nuclear Factor-κB Activation. *J. Biol. Chem.* **277**, 41701–41705.

[840] Lu, N. Z, Collins, J. B, Grissom, S. F, & Cidlowski, J. A. (2007) Selective Regulation of Bone Cell Apoptosis by Translational Isoforms of the Glucocorticoid Receptor. *Mol. Cell. Biol.* **27**, 7143?7160.

[841] Ueda, A, Okuda, K, Ohno, S, Shirai, A, Igarashi, T, Matsunaga, K, Fukushima, J, Kawamoto, S, Ishigatsubo, Y, & Okubo, T. (1994) NF-kappa B and Sp1 regulate transcription of the human monocyte chemoattractant protein-1 gene. *J. Immunol.* **153**, 2052–2063.

[842] Moriuchi, H, Moriuchi, M, & Fauci, A. S. (1997) Nuclear factor-kappa B potently up-regulates the promoter activity of RANTES, a chemokine that blocks HIV infection. *J. Immunol.* **158**, 3483–3491.

[843] Castro, M, Bloch, S. R, Jenkerson, M. V, DeMartino, S, Hamilos, D. L, Cochran, R. B, Zhang, X. E. L, Wang, H, Bradley, J. P, Schechtman, K. B, et al. (2004) Asthma exacerbations after glucocorticoid withdrawal reflects T cell recruitment to the airway. *Am. J. Respir. Crit. Care Med.* **169**, 842–849.

[844] Hinz, M, Löser, P, Mathas, S, Krappmann, D, Dörken, B, & Scheidereit, C. (2001) Constitutive NF-κB maintains high expression of a characteristic gene network, including CD40, CD86, and a set of antiapoptotic genes in Hodgkin/Reed-Sternberg cells. *Blood* **97**, 2798–2807.

[845] Piemonti, L, Monti, P, Allavena, P, Sironi, M, Soldini, L, Leone, B. E, Socci, C, & Di Carlo, V. (1999) Glucocorticoids affect human dendritic cell differentiation and maturation. *J. Immunol.* **162**, 6473–6481.

[846] Schubert, L. A, Cron, R. Q, Cleary, A. M, Brunner, M, Song, A, Lu, L.-S, Jullien, P, Krensky, A. M, & Lewis, D. B. (2002) AT cell-specific enhancer of the human CD40 ligand gene. *J. Biol. Chem.* **277**, 7386–7395.

[847] Jabara, H. H, Brodeur, S. R, & Geha, R. S. (2001) Glucocorticoids upregulate CD40 ligand expression and induce CD40L-dependent immunoglobulin isotype switching. *J. Clin. Invest.* **107**, 371.

[848] Tolnay, M, Vereshchagina, L. A, & Tsokos, G. C. (2002) NF-κB regulates the expression of the human complement receptor 2 gene. *J. Immunol.* **169**, 6236–6243.

[849] Galon, J, Franchimont, D, Hiroi, N, Frey, G, Boettner, A, Ehrhart-Bornstein, M, O'Shea, J. J, Chrousos, G. P, & Bornstein, S. R. (2002) Gene profiling reveals unknown enhancing and suppressive actions of glucocorticoids on immune cells. *FASEB J.* **16**, 61–71.

[850] Keates, A. C, Keates, S, Kwon, J. H, Arseneau, K. O, Law, D. J, Bai, L, Merchant, J. L, Wang, T. C, & Kelly, C. P. (2001) ZBP-89, Sp1, and nuclear factor-κB regulate epithelial neutrophil-activating peptide-78 gene expression in Caco-2 human colonic epithelial cells. *J. Biol. Chem.* **276**, 43713–43722.

[851] Ishmael, F. T, Fang, X, Galdiero, M. R, Atasoy, U, Rigby, W. F, Gorospe, M, Cheadle, C, & Stellato, C. (2008) Role of the RNA-binding protein tristetraprolin in glucocorticoid-mediated gene regulation. *J. Immunol.* **180**, 8342–8353.

[852] Tanaka, Y, Hayashi, M, Takagi, S, & Yoshie, O. (1996) Differential transactivation of the intercellular adhesion molecule 1 gene promoter by Tax1 and Tax2 of human T-cell leukemia viruses. *J. Virol.* **70**, 8508–8517.

[853] Huang, Y.-H, Wu, J. Y, Zhang, Y, & Wu, M. X. (2002) Synergistic and opposing regulation of the stress-responsive gene IEX-1 by p53, c-Myc, and multiple NF-kappaB/rel complexes. *Oncogene* **21**, 6819–6828.

[854] Hiscott, J, Alper, D, Cohen, L, Leblanc, J, Sportza, L, Wong, A, & Xanthoudakis, S. (1989) Induction of human interferon gene expression is associated with a nuclear factor that interacts with the NF-kappa B site of the human immunodeficiency virus enhancer. *J. Virol.* **63**, 2557–2566.

[855] Baudy, A. R, Reeves, E. K, Damsker, J. M, Heier, C, Garvin, L. M, Dillingham, B. C, McCall, J, Rayavarapu, S, Wang, Z, Vandermeulen, J. H, et al. (2012) delta-9, 11 Modification of Glucocorticoids Dissociates Nuclear Factor-κB Inhibitory Efficacy from Glucocorticoid Response Element-Associated Side Effects. *J. Pharmacol. Exp. Ther.* **343**, 225–232.

[856] Kunsch, C, Lang, R. K, Rosen, C. A, & Shannon, M. F. (1994) Synergistic transcriptional activation of the IL-8 gene by NF-kappa B p65 (RelA) and NF-IL-6. *J. Immunol.* **153**, 153–164.

[857] Lombardi, L, Ciana, P, Cappellini, C, Trecca, D, Guerrini, L, Migliazza, A, Maiolo, A. T, & Neri, A. (1995) Structural and functional characterization of the promoter regions of the NFKB2 gene. *Nucleic acids research* **23**, 2328–2336.

[858] Hinz, M, Lemke, P, Anagnostopoulos, I, Hacker, C, Krappmann, D, Mathas, S, Dörken, B, Zenke, M, Stein, H, & Scheidereit, C. (2002) Nuclear factor κB–dependent gene expression profiling of Hodgkin's disease tumor cells, pathogenetic significance, and link to constitutive signal transducer and activator of transcription 5a activity. *J. Exp. Med.* **196**, 605–617.

[859] Wang, Y, Dang, J, Wang, H, Allgayer, H, Murrell, G. A, & Boyd, D. (2000) Identification of a novel nuclear factor-κB sequence involved in expression of urokinase-type plasminogen activator receptor. *Eur. J. Biochem.* **267**, 3248–3254.

[860] Bren, G. D, Solan, N. J, Miyoshi, H, Pennington, K. N, Pobst, L. J, & Paya, C. V. (2001) Transcription of the RelB gene is regulated by NF-κB. *Oncogene* **20**, 7722–7733.

[861] Häupl, T, Yahyawi, M, Lübke, C, Ringe, J, Rohrlach, T, Burmester, G. R, Sittinger, M, & Kaps, C. (2007) Gene expression profiling of rheumatoid arthritis synovial cells treated with antirheumatic drugs. *J. Biomol. Screen.* **12**, 328–340.

[862] Mettouchi, A, Cabon, F, Montreau, N, Dejong, V, Vernier, P, Gherzi, R, Mercier, G, & Binétruy, B. (1997) The c-Jun-induced transformation process involves complex regulation of tenascin-C expression. *Mol. Cell. Biol.* **17**, 3202–3209.

[863] Grundberg, E, Adoue, V, Kwan, T, Ge, B, Duan, Q. L, Lam, K. C. L, Koka, V, Kindmark, A, Weiss, S. T, Tantisira, K, Mallmin, H, Raby, B. A, Nilsson, O, & Pastinen, T. (2011) Global analysis of the impact of environmental perturbation on cis-regulation of gene expression. *PLoS Genet.* **7**, e1001279.

[864] Nwachukwu, J. C, Srinivasan, S, Bruno, N. E, Parent, A. A, Hughes, T. S, Pollock, J. A, Gjyshi, O, Cavett, V, Nowak, J, Garcia-Ordonez, R. D, et al. (2014) Resveratrol modulates the inflammatory response via an estrogen receptor-signal integration network. *Elife* **3**, e02057.

[865] Beck, I. M, De Bosscher, K, & Haegeman, G. (2011) Glucocorticoid receptor mutants: man-made tools for functional research. *Trends Endocrinol. Metab.* **22**, 295–310.

[866] Scheinman, R. I, Gualberto, A, Jewell, C. M, Cidlowski, J. A, & Baldwin, A. (1995) Characterization of mechanisms involved in transrepression of NF-kappa B by activated glucocorticoid receptors. *Mol. Cell. Biol.* **15**, 943–953.

[867] Mukaida, N, Morita, M, Ishikawa, Y, Rice, N, Okamoto, S.-i, Kasahara, T, & Matsushima, K. (1994) Novel mechanism of glucocorticoid-mediated gene repression. Nuclear factor-kappa B is target for glucocorticoid-mediated interleukin 8 gene repression. *J. Biol. Chem.* **269**, 13289–13295.

[868] De Bosscher, K, Berghe, W. V, Beck, I. M, Van Molle, W, Hennuyer, N, Hapgood, J, Libert, C, Staels, B, Louw, A, & Haegeman, G. (2005) A fully dissociated compound of plant origin for inflammatory gene repression. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15827–15832.

[869] Liden, J, Rafter, I, Truss, M, Gustafsson, J.-Å, & Okret, S. (2000) Glucocorticoid effects on NF-κB binding in the transcription of the ICAM-1 gene. *Biochem. Biophys. Res. Commun.* **273**, 1008–1014.

[870] Brostjan, C, Anrather, J, Csizmadia, V, Stroka, D, Soares, M, Bach, F. H, & Winkler, H. (1996) Glucocorticoid-mediated repression of NFκB activity in endothelial cells does not involve induction of IκBα synthesis. *J. Biol. Chem.* **271**, 19612–19616.

[871] Nissen, R. M & Yamamoto, K. R. (2000) The glucocorticoid receptor inhibits NFκB by interfering with serine-2 phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev.* **14**, 2314–2329.

[872] Bladh, L.-G, Lidén, J, Dahlman-Wright, K, Reimers, M, Nilsson, S, & Okret, S. (2005) Identification of endogenous glucocorticoid repressed genes differentially regulated by a glucocorticoid receptor mutant able to separate between nuclear factor-κB and activator protein-1 repression. *Mol. Pharmacol.* **67**, 815–826.

[873] Ray, A, LaForge, K. S, & Sehgal, P. B. (1991) Repressor to activator switch by mutations in the first Zn finger of the glucocorticoid receptor: is direct DNA binding necessary? *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7086–7090.

[874] Oro, A. E, Hollenberg, S. M, & Evans, R. M. (1988) Transcriptional inhibition by a glucocorticoid receptor-β-galactosidase fusion protein. *Cell* **55**, 1109–1114.

[875] Wu, J, Li, Y, Dietz, J, & Lala, D. S. (2004) Repression of p65 transcriptional activation by the glucocorticoid receptor in the absence of receptor-coactivator interactions. *Mol. Endocrinol.* **18**, 53–62.

[876] De Bosscher, K, Beck, I. M, Dejager, L, Bougarne, N, Gaigneaux, A, Chateauvieux, S, Ratman, D, Bracke, M, Tavernier, J, Berghe, W. V, et al. (2014) Selective modulation of the glucocorticoid receptor can distinguish between transrepression of NF-κB and AP-1. *Cell. Mol. Life Sci.* **71**, 143–163.

[877] Vayssière, B. M, Dupont, S, Choquart, A, Petit, F, Garcia, T, Marchandeau, C, Gronemeyer, H, & Resche-Rigon, M. (1997) Synthetic Glucocorticoids That Dissociate Transactivation and AP-1 Transrepression Exhibit Antiinflammatory Activity in Vivo. *Mol. Endocrinol.* **11**, 1245?1255.

[878] Berghe, W. V, Francesconi, E, De Bosscher, K, Resche-Rigon, M, & Haegeman, G. (1999) Dissociated glucocorticoids with anti-inflammatory potential repress

interleukin-6 gene expression by a nuclear factor-κB-dependent mechanism. *Mol. Pharmacol.* **56**, 797–806.

[879] Schäcke, H, Schottelius, A, Döcke, W.-D, Strehlke, P, Jaroch, S, Schmees, N, Rehwinkel, H, Hennekes, H, & Asadullah, K. (2004) Dissociation of transactivation from transrepression by a selective glucocorticoid receptor agonist leads to separation of therapeutic effects from side effects. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 227–232.

[880] Govindan, M. V, Devic, M, Green, S, Gronemeyer, H, & Chambon, P. (1985) Cloning of the human glucocorticoid receptor cDNA. *Nucleic Acids Res.* **13**, 8293–8304.

[881] Hollenberg, S. M, Weinberger, C, Ong, E. S, Cerelli, G, Oro, A, Lebo, R, Brad Thompson, E, Rosenfeld, M. G, & Evans, R. M. (1985) Primary structure and expression of a functional human glucocorticoid receptor cDNA. *Nature* **318**, 635–641.

[882] Lim, H.-W, Uhlenhaut, N. H, Rauch, A, Weiner, J, Hübner, S, Hübner, N, Won, K.-J, Lazar, M. A, Tuckermann, J, & Steger, D. J. (2015) Genomic redistribution of GR monomers and dimers mediates transcriptional response to exogenous glucocorticoid in vivo. *Genome Res.* **25**, 836–44.

[883] Bain, D. L, Yang, Q, Connaghan, K. D, Robblee, J. P, Miura, M. T, Degala, G. D, Lambert, J. R, & Maluf, N. K. (2012) Glucocorticoid receptor–DNA interactions: binding energetics are the primary determinant of sequence-specific transcriptional activity. *J. Mol. Biol.* **422**, 18–32.

[884] Cassiday, L. A & Maher III, L. J. (2002) Having it both ways: transcription factors that bind DNA and RNA. *Nucleic Acids Res.* **30**, 4118–4126.