Studies on the human ALOX5 promoter: Analysis of promoter-interacting proteins by quantitative proteomics and evaluation of secondary DNA structures

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich Chemie, Biochemie, Pharmazie (14)
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von Katharina Schlag
aus Mainz

Frankfurt (2019)
(D30)

vom Fachbereich Chemie, Biochemie, Pharmazie (14) der
Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan:	Prof. Dr. Clemens Glaubitz
Gutachter:	Prof. Dr. Michael Karas
	Prof. Dr. Dieter Steinhilber
Datum der Disputation:	29.05.2020

CONTENTS

Index of Tables

Index of figures

ABBREVIATIONS

$1,25(OH)_2D_3$	1,25-dihydroxyvitamin D3, calcitriol
4-HCCA	α-cyano-4-hydroxycinnamic acid
5-HETE	5-hydroxyeicosanoic acid
5-HPETE	5-hydroxyperoxyeicosa-tetraenoic acid
5-LO	5-lipoxygenase
AA	arachidonic acid
ACN	acetonitril
AdC	5-azadeoxycytidine
AF4	AF4/FMR2 family member 1
AGC	automatic gain control
ambic	ammonium bicarbonate
AP-MS	affinity purification mass spectrometry
APP	amyloid precursor protein
APS	ammonium peroxodisulfate
AQUA	absolute quantification
ATP	adenosine triphosphate
BCA	bicinchoninic acid
Bcl-2	Apoptosis regulator Bcl-2
BLM	Bloom syndrome protein
bp	basepair
BRACO 19	N,N′-(9-(4-(Dimethylamino)-phenylamino)acridine-3,6-diyl)bis(3 (pyrrolidin-1-yl)-propanamide) hydrochloride
BSA	bovine serum albumin
CDN2B	cyclin-dependent kinase 4 inhibitor B
C/EBPs	CCAAT/enhancer-binding proteins
CD	circular dichroism
CE	cytosolic extract
ChIP	chromatin-immunoprecipitation
CLP	coactosin-like protein
$cPLA_2$	cytosolic phospholipase A_2
CRM	charged-residue model
CTR	control
Da	Dalton
DC voltage	direct current voltage
DMSO	dimethylsulfoxide
DNA	deoxyribonuleic acid
DTE	dithioerythritol
DTT	dithiothreitol
EDTA	ethylene diamine tetraacetic acid
Egr-1	early growth response protein 1
ELISA	enzyme-linked immunosorbent assay
emPAI	exponentially modified PAI
EMSA	electrophoretic mobility shift assay
ERK1/2	extracellular-signal-related kinases 1/2
ESI	electrospray ionization
FBS	fetal bovine serum
FDR	false discovery rate
FLAP	5-lipoxygenase activating protein
FRET	fluorescence resonance energy transfer
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GATA2	endothelial transcription factor GATA-2
GM-CSF	granulocyte/macrophage colony stimulating factor
GRAVY	grand average of hydropathicity
GTP/GDP	guanine triphosphate/ guanine diphosphate
HDAC	histone deacetylase
HEPES	N-2-hydroxyethylpiperazin-N'-2-ethansulfonic acid
HIF-1α	Hypoxia-inducible factor 1 alpha
HIV-1	human immunodeficiency virus 1
hnRNPs	heterogeneous nuclear ribonucleoproteins
HRP	horseradish peroxidase
hTERT	human telomerase reverse transcriptase
iBAQ	intensity based absolute quantification
iCAT	isotope-coded affinity tags
IL	interleukine
IMAC	immobilized metal affinity chromatography
iTRAQ	isobaric tag for relative and absolute quantification
KCl	potassium chloride
KLFs	Krüppel-like factors
LC	liquid chromatography
LFQ	label-free quantification
LPS	lipopolysaccharide

ABBREVIATIONS

LTs	leukotrienes
m/z ratio	mass to charge ratio
MALDI	matrix-assisted laser desorption ionization
MAZ	myc-associated zinc finger protein
MBDs	methyl-binding proteins
$MgCl_2$	magnesium chloride
miRNA	microRNA
MLL	histone-lysine N-methyltransferase 2A
MM6	Mono Mac 6
MS	mass spectrometry
NaCl	sodium chloride
NE	nuclear extract
NES	nuclear export signal
NF-κB	nuclear factor κ B
NHE	nuclease hypersensitivity element
NHS	N-hydroxysuccinimide
NLS	nuclear localization signal
NMR	nuclear magnetic resonance
NSAF	normalized spectral abundance factor
NSD1	nuclear receptor binding SET domain protein 1
p53	cellular tumor antigen p53
PAGE	polyacrylamide gel electrophoresis
PAI	protein abundance index
PARP-1	poly [ADP-ribose] polymerase 1
PBS	phosphate-buffered saline
PDGF A	platelet-derived growth factor subunit A
PDS	pyridostatin
PIPER	N,N'-bis-(2-(1-piperidino)ethyl)-3,4,9,10-perylene-tetracarboxylic acid diimide
PITX1	pituitary homeobox 1
PKA	protein kinase A
PMF	peptide mass fingerprint
Pol II	RNA polymerase II
PPI	protein-protein interaction
PQS	putative quadruplex sequences
PS	penicillin/streptomycin
PSAQ	protein standard absolute quantification
PSM	peptide-to-spectrum match
PTM	post-translational modification
PVDF	polyvinylidene difluoride
QconCAT	quantification concatamer technology
RF	radio frequency
RT-qPCR	reverse-transcriptase quantitative polymerase chain reaction
RUNX1	Runt-related transcription factor 1
RXR	retinoic acid receptor RXR-alpha
S/N ratio	signal to noise ratio
SBEs	SMAD binding elements
SCEs	sister chromatid exchanges
SDS	sodium dodecyl sulfate
SILAC	stable isotope labeling by amino acids in cell culture
SI_N	normalized spectral index
SL2	Schneider's drosophila cell line 2
SMAD	mothers against decapentaplegic homolog
Sp1	specificity protein 1
src	proto-oncogene tyrosine-protein kinase Src
SRM	selected/single reaction monitoring
TEAB	triethylammonium bicarbonate
TEMED	tetramethylethylenediamine
TF	transcription factor
TFA	trifluoroacetic acid
TF-AP_2	transcriptin factor AP_2
TGFβ	transforming growth factor beta
TMB	tetramethylbenzidine
TMPyP4	meso-tetra-(N-methyl-4-pyridyl)-porphyrine
TNFα	tumor necrosis factor α
TOF	time-of-flight mass analyzer
TsA	trichostatin A
UDG	uracil-DNA glycosylase
UTR	untranslated region
UV	ultraviolet
VDR	vitamin D_3 receptor
VDRE	vitamin D_3 response element
VEGF	vascular endothelial growth

	factor
WRN	Werner syndrome ATP-dependent helicase
WT	wildtype
WT1	Wilms tumor 1
XIC	extracted ion chromatogram
YY1	transcriptional repressor protein YY1

1. Introduction

1.1 5-Lipoxygenase

1.1.1 Physiological role of the enzyme

Arachidonate 5-lipoxygenase (5-LO) (EC 1.13.11.34) belongs to the family of lipoxygenases, a class of enzymes which catalyze the insertion of molecular oxygen into their substrates. 5-LO is one of the key components of arachidonic acid (AA) metabolism [1] and the main enzyme in the biosynthesis of leukotrienes, a group of lipid mediators involved in the formation of inflammatory and allergic reactions [2]. 5-LO protein expression is mostly limited to myeloid cells and therefore contributes to their function as major component of the immune system. Its substrate arachidonic acid is released out of membrane-associated phospholipids by cytosolic phospholipase A_2 ($cPLA_2$) and can then be further processed by 5-LO in a two-stage reaction [3]. The first step generates the intermediate 5-hydroxyperoxyeicosatetraenoic acid (5-HPETE), which is then converted to leukotriene A_4 (LTA_4). LTA_4 in turn can subsequently be either hydrolyzed to active leukotriene B_4 (LTB_4), or by conjugation with reduced glutathione be metabolized to active leukotriene C_4 (LTC_4) and its following products leukotriene D_4 (LTD_4) and E_4 (LTE_4) [1,4]. Alternatively, 5-HPETE can be converted to 5-hydroxy-eicosanoic acid (5-HETE) by reduction, which leads to the formation of 5-oxo-HETE. In conjugation with 15-LO, the enzyme is additionally involved in the production of anti-inflammatory lipoxins. Thus, the 5-LO pathway is not strictly limited to the biosynthesis of leukotrienes [5–7].

The biological effects of leukotrienes are mediated by G-protein-coupled receptors on the surface of leukocytes. In the case of LTB_4 they comprise for example the chemotaxis of eosinophils, monocytes or polymorphonuclear leukocytes to sites of active inflammation and contribute to the adhesion of neutrophils on endothelial cells [2,8,9]. Extracellularly released LTC_4 and its subsequent products LTD_4 and LTE_4 (together cysteinyl leukotrienes, cysLTs), which are formed by successive degradation of the respective amino acids, exhibit vaso- and bronchoconstrictive properties and are able to increase vascular permeability [1,2,10]. CysLTs furthermore contribute to the proliferation of epithelial cells and promote alveolar secretion [11,12]. They are mostly related to acute hypersensitivity reactions [4]. As mentioned, biosynthesis of leukotrienes is induced by activated $cPLA_2$ and its subsequent co-translocation with 5-LO to the nuclear envelope [13,14]. At the nuclear envelope $cPLA_2$ releases free arachidonic acid deriving from phospholipids, which in turn is presented by membrane bound 5-Lipoxygenase activating protein (FLAP) for further 5-LO dependent metabolism [3, 10].

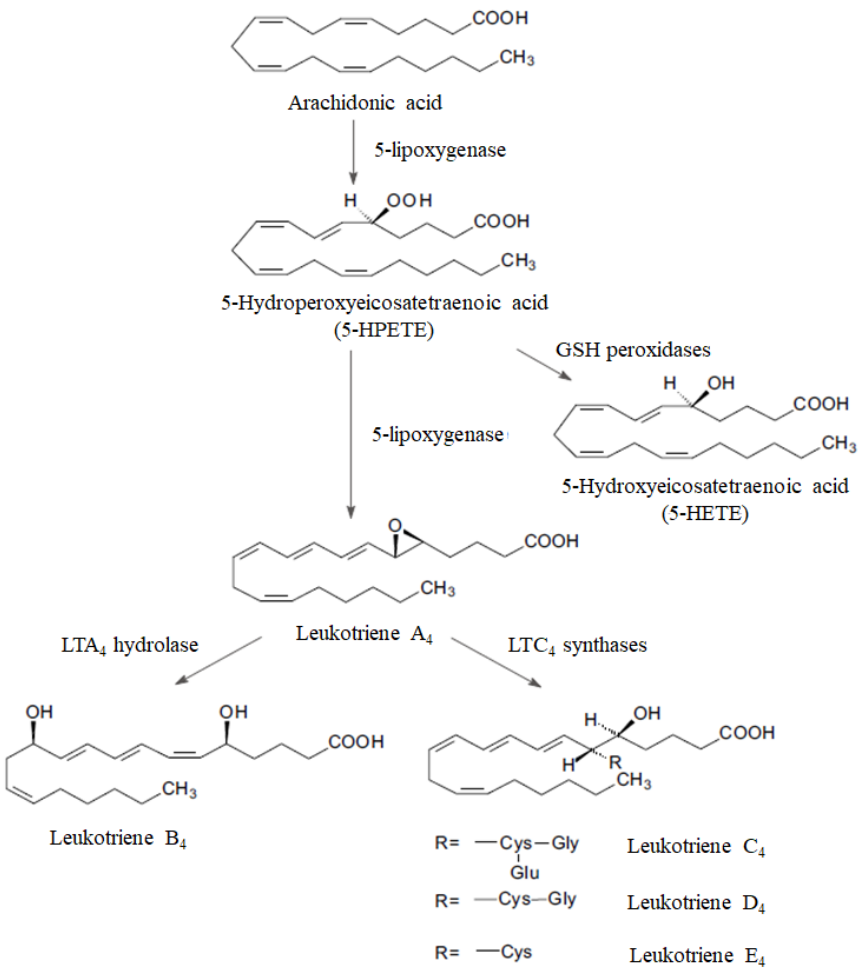


Figure 1: The leukotriene pathway, adapted from [15].

1.1.2 Pathophysiological role of the enzyme

Based on the physiological effects of leukotrienes, their role in allergic and inflammatory reactions, which comprise local effects and the influx of cells of the immune system into the respective tissue, is quite obvious. Increased production of cysLTs can hence be found in clinical diseases like asthma or allergic rhinitis and together with other mediators of inflammatory reactions they modulate complex processes [15]. Their effects are mediated by different cysLT-receptors ($cysLTR_1$ and $cysLTR_2$) and inhibition of these exhibits a possible option of therapeutical intervention in the treatment of asthma bronchiale. Hence, the $cysLTR_1$ antagonist Montelukast alleviates symptoms of asthmatic disorders like increased mucus production, alveolar hyperreactivity and constriction [12].

Besides, increased levels of metabolites of the leukotriene pathway and respective receptors can be found in atherosclerotic lesions of the vasculature. The infiltration of monocytes and their conversion to macrophages present the major part of leukotriene producing cells. The

transformation of the latter into foam cells thereby promotes the aggregation of atherosclerotic plaques, which after rupturing eventually can lead to infarcts of affected tissues. As cysLT-receptors are expressed on the smooth musculature of coronary arteries, their activation contributes to pathological arterial muscle hyperplasia [12]. Additionally, leukotrienes are involved in various other inflammatory diseases, including atopic dermatitis [16] or rheumatoid arthritis [17].

5-LO furthermore takes part in aging processes in the human brain and thus is connected to neurodegenerative diseases [18]. A contribution in the formation of β-amyloid could be shown, which plays a crucial role in the pathophysiology of Alzheimer's disease. 5-LO stimulates the transcription of γ-secretase, a protein complex responsible for the cleavage of amyloid precursor protein (APP). High enzyme levels resulted in higher β-amyloid concentrations, whereas 5-LO inhibition led to reduced formation of β-amyloid, respectively.

Higher levels of 5-lipoxygenase are also found in various tumor tissues. For example, increased proliferation of prostate and esophagus cancer cells is associated with increased levels of metabolites of the leukotriene pathway (in case of prostate cancer cells 5-HETE) in the respective tissues. By inhibiting 5-LO, apoptosis of these tumor cells could be induced [19, 20]. Similar results were obtained for colon and pancreas cancer cells, whose proliferation was diminished by 5-LO inhibition [21, 22]. Taken together, 5-LO is a major target for therapeutic intervention in both inflammatory and cancer diseases.

1.1.3 Expression profile in humans and subcellular localization

The physiological 5-LO protein expression is limited to cells of the immune, as well as the central nervous system (see section 1.1.2). It is mostly present in leukocytes of the myeloid lineage, including polymorphonuclear leukocytes (neutrophils, eosinophils), monocytes/macrophages and thereof arising foam cells of atherosclerotic plaques, mast cells, dendritic cells and to a much lower level in B-lymphocytes [23]. The enzyme's expression status in monocytes and granulocytes depends on the respective status of cellular differentiation [10]. The induction of cell differentiation by the agents TGFβ/1,25$(OH)_2D_3$, DMSO or retinoic acid leads to an upregulation of 5-LO mRNA- as well as protein levels, which also applies for the tissue infiltration of monocytes and their consecutive differentiation into macrophages [23]. The highest promotion of 5-LO protein expression could be observed for the combined addition of TGFβ/1,25$(OH)_2D_3$ [24]. Furthermore, 5-LO expression can be induced by glucocorticoids or the granulocyte/macrophage colony stimulating factor (GM-CSF) in monocytes or granulocytes, respectively [23].

5-lipoxygenase is a mobile enzyme, present either in the cytosol or the nucleus. The subcellular localization depends on the particular cell type. In resting peritoneal macrophages, neutrophils, eosinophils and monocytes, 5-LO is mostly found in the cytoplasm, whereas alveolar macrophages, mast cells and Langerhans cells of the skin display both nuclear and cytosolic enzyme. Import into, or export out of the nucleus is accomplished by existing nuclear localization signal (NLS) and nuclear export signal (NES) sequences in the 5-LO structure [10, 15].

1.1.4 Regulation of enzyme activity

Enzymatic activity of the enzyme is regulated on protein level by complex mechanisms (figure 2). Due to the non-heme iron present in the 5-LO, the cytosolic redox state contributes fundamentally to an activating or inhibiting milieu. For active catalysis the ferric state (Fe^{3+}) of the non-heme iron is essential, hence a low level of peroxides is crucial for enzyme functionality [10]. Accordingly, enzymatic activity is also altered by the presence of reactive oxygen species.

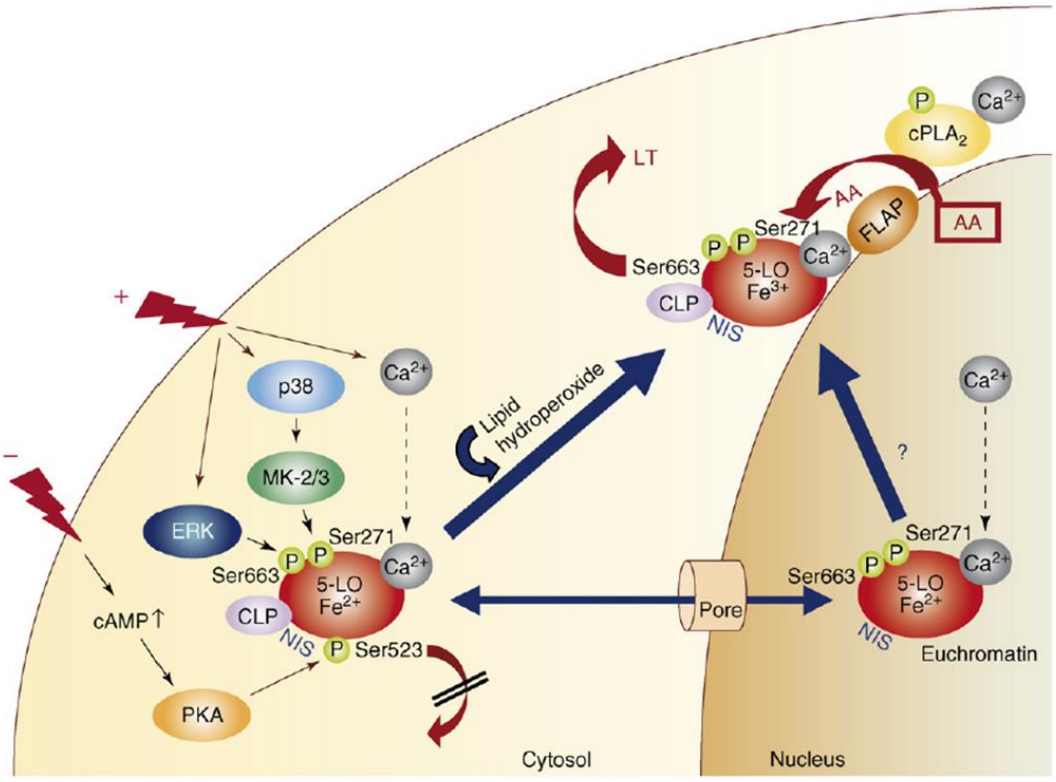


Figure 2: Regulation of 5-lipoxygenase activation on protein level [23].

Activity of the 5-LO is dependent on Ca^{2+}, concentrations thereof in the mikromolar range presumably support its binding to phosphatidylcholine or coactosin-like protein (CLP), which

in turn facilitates activation and the enzyme's translocation to the nuclear membrane [23]. CLP shows stabilizing effects on the enzyme and thus supports its calcium-dependent activation [15, 25]. Consistently, treatment with calcium ionophore induces the translocation of the 5-LO to the nuclear envelope [4].

ATP presents another mediator for enzyme activation. In contrast to other lipoxygenases, 5-LO is the only one which is activated by ATP [26]. Presumably, ATP likewise does have stabilizing effects on the enzyme [23], maximum activity however is achieved only in combination with Ca^{2+} [27].

The presence of FLAP is crucial for the biosynthesis of leukotrienes. FLAP is expressed accordingly in 5-LO positive cells; without its existence metabolism of AA is generally not taking place [4, 23, 28]. FLAP itself is an integral membrane protein, which associates with 5-LO at the nuclear membrane and facilitates the presentation of membrane-released free AA for its respective catalysis for leukotriene biosynthesis.

Lastly, different studies could prove that enzyme activation or inhibition also depends on its phosphorylation status (see fig. 2). 5-lipoxygenase is a substrate for varying kinases, including protein kinase A (PKA) [29], extracellular-signal-related kinases (ERK1/2) or p38 mitogen-activated protein kinase [30, 31], although phosphorylation of Ser663 by Erk2 remains controversial, since mass spectrometry-based studies did not confirm its phosphorylation state [32]. Normally, activation of the respective tyrosine kinases, triggered by varying stress responses, leads to increased 5-LO activity [6]. However, phosphorylations can also exhibit inhibitory effects on 5-LO activity, as shown by phosphorylation of Ser523, which represses enzyme activity and impedes translocation to the nuclear membrane [23, 29]. Furthermore, phosphorylation status is jointly responsible for the subcellular localization of the 5-LO [15].

1.1.5 Regulation of 5-LO expression

1.1.5.1 ALOX5 gene organization and promoter region

The 5-lipoxygenase is encoded by the ALOX5 gene, which is located on chromosome 10q11.21. It comprises over 82 kilobasepairs (kbp) and is divided into 14 exons and 13 introns of different length. Exons 1-7 encode for the amino-containing half of the protein and span over 65 kbp, whereas exons 8-14 encode for the carboxylic half of the enzyme including its active site and only span about 6 kbp [33].

The 5'-flanking region of the gene contains both positive and negative regulatory elements and binding sites for a variety of transcription factors (TFs) and was first characterized in the late 1980s by Hoshiko *et al* [34]. They described the promoter region to cover almost 6 kbp,

around 800 bp of them designated to represent the core promoter 5'-upstream of the major transcription initiation site. A noticeable feature in its sequence is the lack of classical TATA- or CCAAT-boxes, providing characteristics of housekeeping genes [35]. Accordingly, the GC-content of about 80% is quite high, including eight GC-boxes, which generate binding sites for the ubiquitously present transcription factor Sp1. Further GC-boxes can additionally be found in Intron A [33]. Out of these eight GC-boxes of the sequence GGGCGG, five are arranged in tandem close to the transcription initiation site. Putative binding sites for further transcription factors, including c-Myb, NF-κB and TF-AP2 are present in the upstream sequence of the promoter (figure 3) [34].

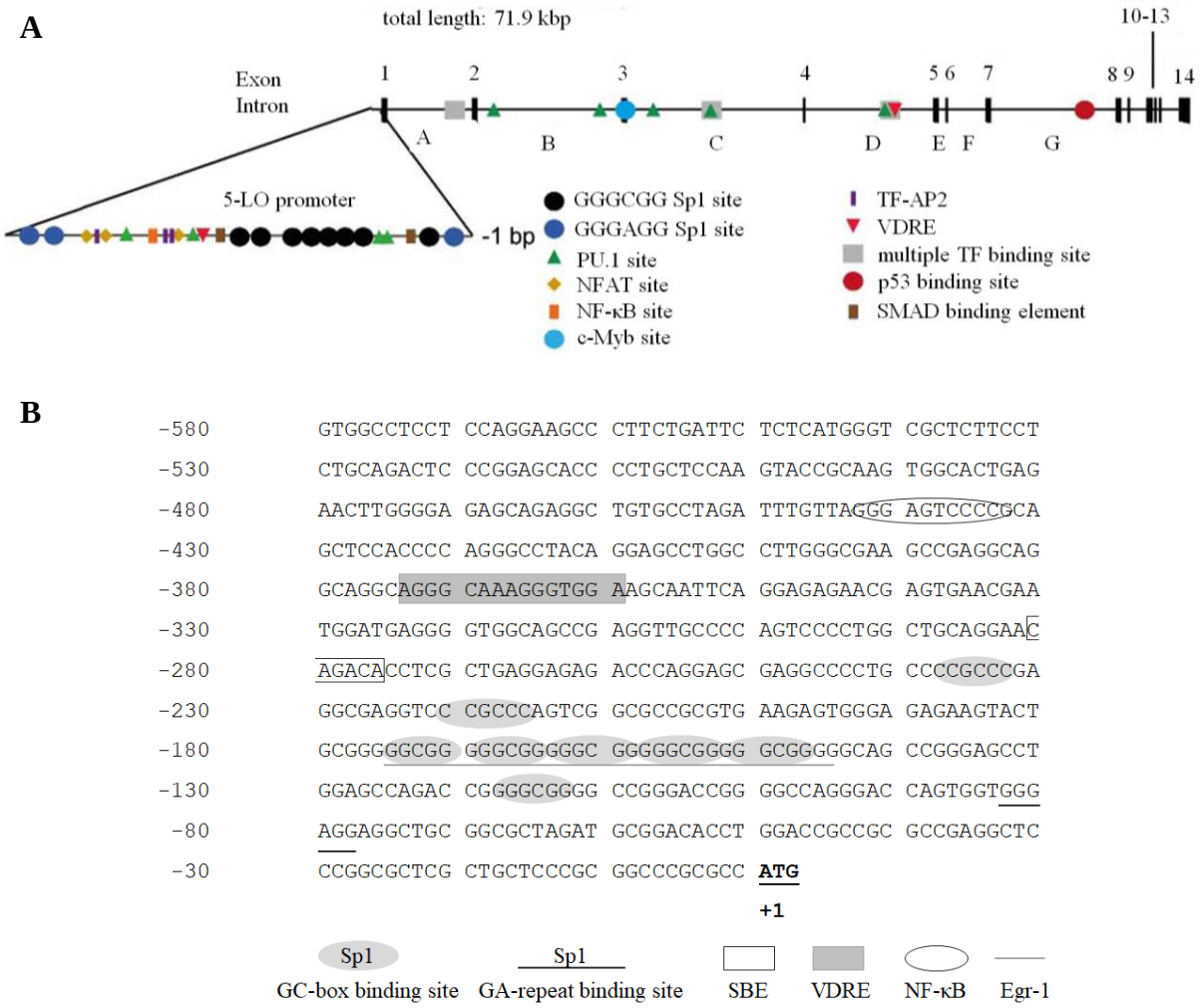


Figure 3: ALOX5 gene organization [36] *(A) and overview of the proximal promoter sequence (B). Transcription factor binding sites are defined in relation to the translational start site (ATG).*

Transcription factor Sp1 was shown to bind to the tandem repeats of GC-boxes, thereby inducing promoter activity [34, 37]. Sp1 represents one of nine members of transcription

factors in the Sp-family, further including Sp2 to Sp9, which in turn can be classified as subfamily of the overall Sp/KLF (Krüppel-like factor) transcription factor family [38]. These TFs share a common DNA-binding domain, which contains three zinc fingers of the C_2H_2-type in close proximity to the C-terminus and a glutamine rich stretch neighboring serine/threonine sequences [39, 40]. Activity or binding properties of Sp1 can be influenced by posttranslational modifications, for the most part phosphorylation and glycosylation [41, 42]. Sp1 is an activator of transcription in various promoters and contributes to transcriptional regulation in a large number of myeloid genes [39, 43]. Activation of promoter activity and enhanced transcription by RNA polymerase II is mediated by Sp1 alone or regulated in conjunction with the other members of the family, most prominently Sp3 [39, 40, 44].

The tandem GC-box provides additional binding sites for inducible transcription factor Egr-1 that overlap with those for Sp1. Diminished 5-LO expression after the addition of GC-box blocking substances, as well as constant levels of both factors before and after cell differentiation point to the involvement of both transcription factors in basal promoter activity [45]. Deletion of one or two, or addition of one GC-box in the tandem repeats are naturally occurring mutations of the 5-lipoxygenase promoter sequence [46]. Using reporter gene constructs with theses mutations, an induction of promoter activity by Sp1 and Egr-1 could be shown in SL2 cells, whereas in HeLa cells contradictory results were obtained and activity was repressed [37, 46]. Due to the inconsistent results on 5-LO expression in the different cell lines, basal promoter activity is suggested to depend on interplay of both transcription factors, whose effects however are cell-specific. Evidence for *in vivo* relevance of these mutations was provided by studies dealing with a distinct response to treatment with 5-LO inhibitors in line with asthma therapeutic intervention [47], as well as the increased susceptibility to develop atherosclerosis for patients carrying a mutation [48].

1.1.5.2 Epigenetic factors

The limitation of 5-LO expression to certain cell types and its regulation by inducers of differentiation posed the question, whether its cell specificity in part depends on DNA-based mechanisms that are responsible for the enzyme's presence in mostly immunocompetent cells. In this regard, analysis of HL-60TB and monocytic U937, which both express FLAP, but are 5-LO negative, revealed a correlation between methylation of the CpG islands in the promoter region and repression of 5-lipoxygenase expression [49]. Methylation of promoter sequences is a general mechanism for gene silencing [50]. Treatment of both cell lines with the demethylating agent 5-azadeoxycytidine (AdC) leads to the induction of 5-LO mRNA

expression, as well as enzyme activity which correlates to those of untreated wildtype HL-60 cells. The latter revealed an unmethylated core promoter, whereas HL-60TB and U937 exhibited highly methylated CpG islands. Reporter gene assays of completely methylated ALOX5 promoter constructs resulted in inhibition of promoter activity. Cell treatment with a combination of TGFβ/1,25$(OH)_2D_3$ and AdC provided synergistic effects on mRNA-transcription level, however no direct impact on promoter activity, allowing for the suggestion that the combination of TGFβ/1,25$(OH)_2D_3$ exhibited posttranscriptional influence on 5-LO expression [49]. Furthermore, an essential regulatory role on transcription efficiency could be attributed to the GC-box following the tandem GC-boxes. Methylation of the former in contribution with methyl-binding proteins (MBDs) impedes binding of Sp1, thereby suppressing promoter activity [51]. Thus, repressed 5-LO expression in non-myeloid cells is probably due to the respective DNA methylation status [23, 49].

Not only methylation status plays a role in the regulation of 5-LO transcription, but also histone acetylation is involved in promoter regulation, the latter however succumbing the effects of methylation [52]. Treatment of the monocytic cell line Mono Mac 6 (MM6) with HDAC inhibitor trichostatin A (TsA) both induces 5-LO expression and increases the occurrence of H3K4me3 as marker of active promoters. Furthermore, TsA-mediated induction of ALOX5 promoter activity is independent of the existence of the tandem GC-boxes (within the reporter constructs used in the study) or the addition of differentiating agents TGFβ/1,25$(OH)_2D_3$. TsA further induces basal promoter activity due to a promoted formation of transcription complex Sp1/Sp3/polymerase II on both of the GC-boxes following the tandem repeats [53]. Raised occurrence of H3K4me3 after TsA treatment requires increased recruitment of methyltransferases, such as MLL (mixed lineage leukemia), which in turn leads to higher levels of 5-LO mRNA expression. According to that, the proto-oncogenic fusion protein MLL-AF4 could be related to constitutive active promoter activity, which is independent of HDAC inhibitors, but is able to be antagonized by present MLL itself. These results suggest on the one hand an uncontrolled 5-LO expression in leukemia affected by a MLL translocation. On the other hand the respective leukemia could profit from the treatment of class I HDAC inhibitors as therapeutic intervention [54].

1.1.5.3 Regulation by TGFβ/1,25(OH)$_2$D$_3$ and downstream transcription factors

As already mentioned in section 1.1.3, induction of 5-lipoxygenase expression after differentiation of myeloid cells initiated by differentiating agents exhibits highest effects on the synthesis of mRNA- and protein levels by adding TGFβ/1,25(OH)$_2$D$_3$ [24]. Treatment of MM6 cells with TGFβ/1,25(OH)$_2$D$_3$ did not show any impact on the half-life of present mRNA, however altered amounts of 5-LO RNA were detected after the examination of different isolated exon constructs, which each comprised one of the 14 exons, giving evidence for posttranscriptional regulation of both differentiating agents. These effects could be attributed to enhanced maturation of mRNA as well as stimulation of transcript elongation [55]. In fact, binding of the heterodimer VDR/RXR to its respective response elements located in the promoter region was shown, which however does not lead to TGFβ/1,25(OH)$_2$D$_3$-dependent induction of promoter activity. This supports the assumption of regulatory elements present outside of the promoter sequence, which could be responsible for this effect [56]. In this regard, further studies gave proof for promoter-independent induction of reporter gene constructs containing the 5-LO coding sequence or introns J-M by TGFβ/1,25(OH)$_2$D$_3$ [57, 58]. Four functional response elements for TGFβ effectors SMAD 3/4 are located in intron M, further SMAD binding elements (SBEs) can be found in the coding sequence. Deletion of exons 10-14 results in the most profound repression of SMAD response [57]. Additionally, functional vitamin D$_3$ response elements (VDREs) are located in intron M and exons 10 and 12, which have been shown to bind VDR and extenuate its response after being mutated or deleted. The TGFβ/1,25(OH)$_2$D$_3$ effects on the distal part of the 5-LO gene furthermore induce both acetylation of histone H4, a marker for open chromatin structures, and increased presence of elongation markers, including H3K36me3 and H4K20me1, contributing to regulation by transcript elongation [58]. Further VDREs are present in the 5-LO gene, one of them located in intron 4, which is involved in transcript elongation as well. It is considered to be one of the most responsive response elements throughout the human genome [59]. Additional influence on VDR-dependent transcript elongation is mediated by AF4 and its oncogenic fusion protein AF4-MLL, which is equally affected by MLL translocation as the already mentioned MLL-AF4 (see section 1.1.5.2) [60]. According to the study, heterodimer VDR/RXR interacts with AF4 as regulator of elongation and the fusion protein AF4-MLL, which exceeds the elongation effects of AF4 alone. Inhibiting elongation by using class I HDAC inhibitors further contributed to a decrease of 5-LO expression and activity.

Further studies based on reporter gene construct activity provided evidence for the participation of SMAD proteins on transcript initiation [60, 61]. The core promoter contains two SBEs, which bind SMAD 3/4 and induce MLL-dependent Sp1-mediated promoter activity. SMAD 3/4 are furthermore shown to have alleviating effects on the activity of the fusion protein MLL-AF4 [61].

1.1.5.4 Other transcription factors in ALOX5 regulation

Besides the already mentioned transcription factors, ChIP-seq analysis from different laboratories provided data on additional regulatory elements located in the promoter sequence, as well as in different introns of the 5-LO gene. These sequences contain binding sites for transcription factors that are involved in myeloid differentiation or regulation of stem cell functions, including C/EBPα, GATA2 or RUNX1 among others [36].

Intron G for example contains a functional consensus sequence for tumor suppressor p53, which acts as transcriptional enhancer on 5-lipoxygenase expression. p53-mediated transcription can be induced by treatment with cytotoxic substances actinomycin D and etoposide and knockdown of p53 leads to repressed enzyme expression, respectively. Simultaneously, a 5-LO-mediated repression of p53-controlled transcriptional activity was confirmed as well [62].

1.2 DNA secondary structures

1.2.1 General features

Since the introduction of the established model of the DNA-double helix, which is formed by Watson-Crick base pairing [63], further possible secondary nucleic acid structures were investigated. Although most part of cellular DNA in fact is existent in form of the canonical double helix, termed B-form DNA, alternative secondary structures could be identified in the following years, whose functions are not completely elucidated as yet. According to different studies, the fraction of protein coding DNA of the human genome comprises around 2%, relating to an estimated number of genes ranging from 20000 to a maximum of 25000 [64]. For the remaining 98% there is still limited knowledge about its functions or regulatory effects on the human organism. Currently, alternative non-B-form DNA secondary structures comprise for example left handed Z-form DNA (Z-DNA), cruciforms, G-quadruplex DNA or i-motif DNA [65]. All of these secondary structures possibly contribute to transcriptional regulation and their role needs to be investigated profoundly.

Guanine-rich DNA sequences tend to form quadruplexes as secondary structures, also termed G4- or tetraplex-DNA [65]. They are arranged from planar guanine tetrads developing from four guanines that are stabilized by Hoogsteen-like hydrogen bonds. Multiple of these tetrads in turn stack upon each other to form a quadruplex, which usually is being stabilized by monovalent cations. These quadruplexes can exhibit different orientation of strands or conformations [66].

In general, the combination of one, two or four strands is possible for quadruplex formation, which are named mono-(intra), bi-(inter) or tetra-(linear)-molecular, respectively [67]. Depending on strand orientation, altering conformations are built that can be divided up into intramolecular basket, propeller or chair type (figure 4B-D), intermolecular basket or hairpin type or parallel-strand type [67, 68]. The parallelism of the strands can further be divided into parallel and anti-parallel [66, 69–71] (figure 4D-F). DNA sequences are capable of developing different topologies, depending on the length of the quadruplex involving DNA stretch or the disposable cations. An additional classification of syn- or anti-conformation can be attributed to the respective glycosylic conformation [66, 70, 71].

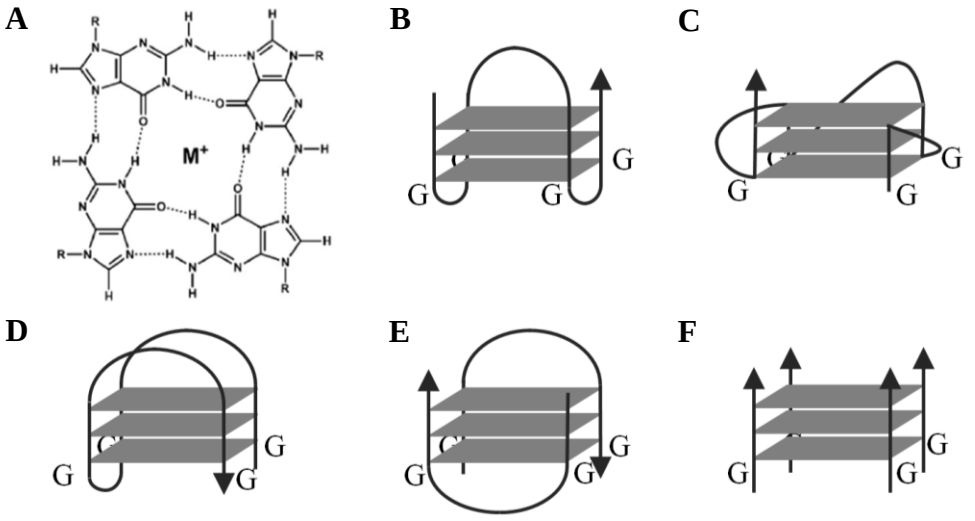


Figure 4: Selected G-quadruplex conformations as secondary DNA structures. (A) Schematic representation of a planar G-tetrad, taken from [65] *and exemplary structures forming (B) basket type, (C) propeller type, (D) chair type G-quadruplexes with single-strand-antiparallel, (E) two-strands-antiparallel and (F) four strands parallel orientation.*

Formation and stability of quadruplex structures are essentially influenced by bound cations. Generally, monovalent cations, including K^+, Na^+ or NH_4^+ are favored over divalent ones; optimal stability is achieved with K^+ [66]. However, various studies exist, which prove the

formation of quadruplex DNA being possible after the addition of alternative mono- or divalent cations [68]. For the *in vitro* investigation of present G4-DNA the suitability of cations has to be considered, since for example the exchange of a K^+-stabilized quadruplex into Na^+ buffer conditions can alter the topology of the quadruplex structure.

In addition to the appearance of G4-DNA RNA quadruplexes have been described; hence the formation thereof is a possible secondary structure of nucleic acids in general.

1.2.2 Biological relevance of G4-DNA

1.2.2.1 Telomeric quadruplexes

Over the years the *in vivo* existence of quadruplex structures was controversially discussed, as early investigations only dealt with their *in vitro* formation. However, they seemed likely to be involved in transcriptional regulation provided that they actually did exist in the human genome. With the discovery of G4-DNA in telomere ends, quadruplexes initially became a possible target for therapeutic intervention [72–74]. Telomere DNA in 3'-overhangs of chromosomes consist of single stranded repetitive guanine-rich sequences of the type d(TTAGGG), which normally contain 100-200 bp and contribute to maintaining chromosomal integrity. In this regard, telomere length is crucial for the regulation of cell cycle and cellular aging and is controlled by telomerase, an enzyme adding sequence repeats onto the respective telomere ends. Usually, telomerase activity is low in somatic cells resulting in a gradual degradation of telomeres. Their shortening eventually leads to cell senescence. An increased enzyme activity counteracts these procedures and yields virtually immortal cells. A variety of tumor cells are affected by elevated telomerase activity, so that telomerase inhibitors are considered possible cancer therapeutics [70, 75]. Since quadruplex structures in telomere ends inhibit enzyme activity, quadruplex stabilizing agents provide a means for potential indirect inhibition of telomerase [75–77]. Some of these indirect inhibitors are currently established agents for giving evidence of G4-DNA *in vitro* and *in vivo* and shall be further elucidated in 1.2.2.5.

1.2.2.2 Quadruplexes in promoter regions

After the discovery of quadruplex structures in telomere ends studies were expanded to guanine-rich sequences in human genes, which resulted in the detection of G4-DNA in promoter regions of certain genes. By now, fundamental knowledge was provided by extensive studies on proto-oncogenes including c-MYC [78], BCL-2 [79], k-RAS [80],

c-MYB [81], h-RAS [82], VEGF [83] or hTERT [84]. An overview of quadruplex structures in human promoter regions is given in table 1.

Investigations on the proto-oncogene c-MYC provide most insights into the regulatory role of G4-DNA on transcription in this context so far. Transcription factor c-Myc is primarily involved in both activation and repression of multiple growth-related genes and further contributes to the regulation of telomerase. It is widely overexpressed in different types of tumor tissues and is estimated to be involved in about 20% of total human cancers [85]. Accordingly, c-Myc possesses a critical role in cancer development [86]. The c-Myc gene transcription is regulated by dual promoters (P1 and P2); 85 to 90% of transcriptional activity is controlled by the nuclease hypersensitivity element (NHE) III_1 of promoter P1, which contains a stretch of 27 bp that is able to form quadruplex structures [78, 87]. Stabilization of these quadruplexes by interacting agents or telomerase inhibitors leads to the suppression of transcription and therefore results in impeded c-Myc expression [78]. This provides targets for downregulation of the latter and consecutive inhibition of proliferation of affected cancer cells.

Similar results could be obtained for the promoter region of the BCL-2 (B-cell lymphoma 2) gene. Its protein product is involved in the regulation of cell survival and displays a potent inhibitor of apoptosis [88]. Overexpression of BCL-2 is being found in various tumor tissues and contributes to carcinogenesis in breast, B-cells, or colorectal tissue [89]. Transcriptional regulation resides with a GC-rich element around -1400 bp upstream of the transcription start site, which tends to form quadruplex structures [90]. Mutations thereof which disrupt these quadruplexes resulted in increased transcriptional activity, whereas stabilization by quadruplex interacting agents diminished promoter activity [91, 92].

The proto-oncogenes of the RAS-family encode for GTP/GDP-binding proteins ($p21^{RAS}$) and similarly contain a GC-rich promoter region while lacking TATA boxes [93]. Their respective protein products are involved in cell proliferation and participate in signal transduction pathways. A protein dysfunction caused by mutations of the corresponding gene eventuates in constitutively active $p21^{RAS}$, thereby promoting growth of different kinds of tumors [82]. Both k-RAS and h-RAS are capable of forming quadruplex structures in their promoter region, which regulate transcription [80, 82]. Studies on the murine k-RAS promoter sequence revealed the presence of GA-repeats, which were essential for G4-formation [94]. Similar characteristics were shown for the c-MYB promoter that contains four GGA repeats 17 bp upstream of the transcription initiation site, which control promoter activity. Deletion of one or two of these repeats led to enhanced activity, suggesting that formed quadruplexes

exhibit a suppressing effect [81]. C-MYB likewise represents a proto-oncogene whose protein product is involved in cell proliferation [95] and is overexpressed in acute leukemia and small cell lung carcinoma [96, 97].

Since most of the G4-sequences so far identified belong to promoter regions of proto-oncogenes, efforts are made to specifically target quadruplex structures or their known interacting proteins to control their dysregulation and provide new possibilities for therapeutic intervention in cancer treatment.

gene	function of quadruplex	topology	reference
c-MYC	gene repression	intramolecular parallel	[78]
VEGF	gene repression	intramolecular parallel	[83, 98]
c-MYB	gene repression	intramolecular parallel	[81]
k-RAS	gene repression	intramolecular parallel	[80, 94]
h-RAS	gene repression	intramolecular antiparallel	[82]
insulin	gene activation	Intra-, intermolecular parallel/antiparallel	[99–101]
HIF-1α	gene repression	intramolecular parallel	[102, 103]
c-KIT	gene repression	intramolecular parallel, mixed topology	[104, 105]
BCL-2	gene repression	intramolecular parallel/antiparallel	[79, 92]
PDGF-A	gene repression	intramolecular parallel	[106]
hTERT	presumably gene repression	intramolecular hybrid (3+1), parallel	[84, 107]
muscle related genes	gene activation	?	[108]
SRC	gene repression	intramolecular parallel	[109]
HIV-1	gene repression	intramolecular antiparallel	[110, 111]
PARP-1	?	hybrid (3+1)	[112]
WT1	gene repression	intramolecular parallel	[113]

Table 1: DNA quadruplex structures found in human promoter regions and their respective topology and effect on the target gene

1.2.2.3 Quadruplex interacting proteins

Various proteins could be identified by now that interact with G4-DNA in a promoter region and thereby may influence transcriptional regulation. These include, among others, topoisomerase I which is able to both bind to preformed inter- and intramolecular quadruplexes and induce their formation, respectively [114]. Another class of G4-interacting proteins comprises the heterogeneous nuclear ribonucleoproteins (hnRNPs), RNA-binding proteins taking on different functions in transcriptional regulation. hnRNP D consists of two nucleic acid binding domains that recognize the already mentioned telomere repeat d(TTAGGG) and further unwind preformed single-strand quadruplexes thereof [115]. The same applies to hnRNP A1, which additionally stimulates telomerase activity [116]. Both proteins thereby contribute fundamentally to telomere maintenance [117]. The quadruplex structure of the k-RAS promoter was further shown to be unwound by hnRNP A1, whose function hence is not limited to telomere interaction, but also contributes to G4-mediated transcriptional regulation of the k-RAS promoter [118, 119]. Moreover, a B-cell specific isoform of hnRNP D is capable of heterodimerizing with nucleolin to regulate c-MYC transcription after binding to a DNA sequence upstream of the NHE III_1 element (see section 1.2.2.2) [120, 121]. Nucleolin is an utterly abundant nucleolar protein involved in cell proliferation and the assembly of ribosomes [71]. Its highly specific affinity and selectivity for binding quadruplexes is many times higher than for its regular RNA substrates [121]. Nucleolin was proven to induce the formation of quadruplex structures in the c-MYC NHE III_1 element and bind to them *in vivo*, and due to stabilizing effects thereon subsequently downregulate c-MYC transcription. Accordingly, overexpression of nucleolin resulted in a reduction of Sp1-mediated c-MYC expression [121]. In contrast, an activating effect on NHE III_1 traces back to hnRNP K, which binds to the C-rich complementary single strand DNA sequence, implicating that an interplay of the aforementioned ribonucleoproteins is responsible for the complex regulation of c-MYC expression. Similar results were obtained for studies dealing with the VEGF gene. Its promoter region is mainly activated by double-strand binding transcription factor Sp1, however a synergism of nucleolin and hnRNP K likewise activate transcription Sp1-independently [122].

According to genome wide studies, there exist over 700 000 DNA sequences that have potential for forming quadruplexes [123]. These putative quadruplex sequences (PQS) frequently emerge in promoter regions around 100 bp upstream of the transcription start site, in close proximity to or overlapping with consensus sequences of certain transcription factors [124–126]. Interestingly, a high correlation between Sp1-binding GC boxes of the sequence

GGGCGG and PQS could be ascertained [125]. In this context, an association of Sp1 and other regulating zinc finger proteins recognizing the same or a similar binding site and the presence of quadruplex motifs seems quite obvious. This could be shown in different organisms for a limited number of zinc finger TFs, mostly containing the C_2H_2-motif (e.g. Sp1, Egr-1, AP-2, MAZ) [126].

Studies provided evidence for zinc finger protein MAZ (myc-associated zinc finger protein, also termed transcription factor Pur-1 or SAF-1) acting as transcription factor, which is involved in the specific recognition of G4-DNA. MAZ binds GC/GA-rich sequences and has dual role as transcriptional activator or repressor. It is additionally capable of recognizing both double-stranded and G4-DNA [81], whereas its binding to the latter was first observed for the human insulin promoter [127]. MAZ gained further importance as a transcriptional regulator for interacting with the already mentioned GGA repeats of the c-MYB promoter. In this context, MAZ-mediated downregulation of c-MYB promoter activity was shown that was suggested to be relying on secondary recruitment of other transcriptionally active proteins (figure 5) [81]. More recent studies identified an additional MAZ binding site in the c-MYB promoter through which, on the other hand, the protein activates transcription in T-lymphocytes during their exit from quiescence. Thus, MAZ supposedly has dual role in the activation or repression of the gene by binding to either one of the binding sites [128].

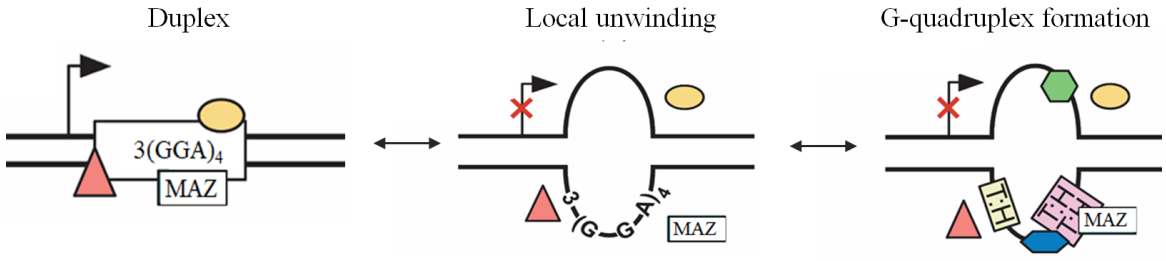


Figure 5: Role of transcription factor MAZ in G-quadruplex-mediated c-MYB gene transcription. Duplex DNA is in equilibrium with open and quadruplex DNA structures. The open form is regulated by transcriptional activator and repressor proteins; MAZ potentially acts as repressor is this state. G-quadruplex formation via local unwinding goes along with exchange of formerly bound associative proteins to be replaced by G4-interacting proteins. MAZ is proposed to recognize and stabilize G-quadruplex structures, thereby repressing c-MYB expression, figure modified from [81].

In contrast, an opposite function on the activity of murine k-RAS promoter was attributed to MAZ. In this case, the protein binds to quadruplexes present in the promoter region, thereby

stabilizing their structure and inducing expression. MAZ-silencing correlated with a loss of gene transcription of about 40% [94].

Further studies on transcription factors Sp1 and MAZ provided evidence for cooperative effects of both proteins on transcriptional regulation [129, 130]. Interestingly, proof for synergistic functions of Sp1 and MAZ on G4-DNA regulated transcription could also be shown for the regulation of the h-RAS promoter (figure 6). The formation of two quadruplexes therein, which are located upstream of the transcription start site are recognized by both TFs and exhibit repressing effects on promoter activity. Supposedly, MAZ is capable of unwinding these quadruplex structures to synergistically stimulate Sp1-driven transcription, which confirms an interplay of both transcription factors [82].

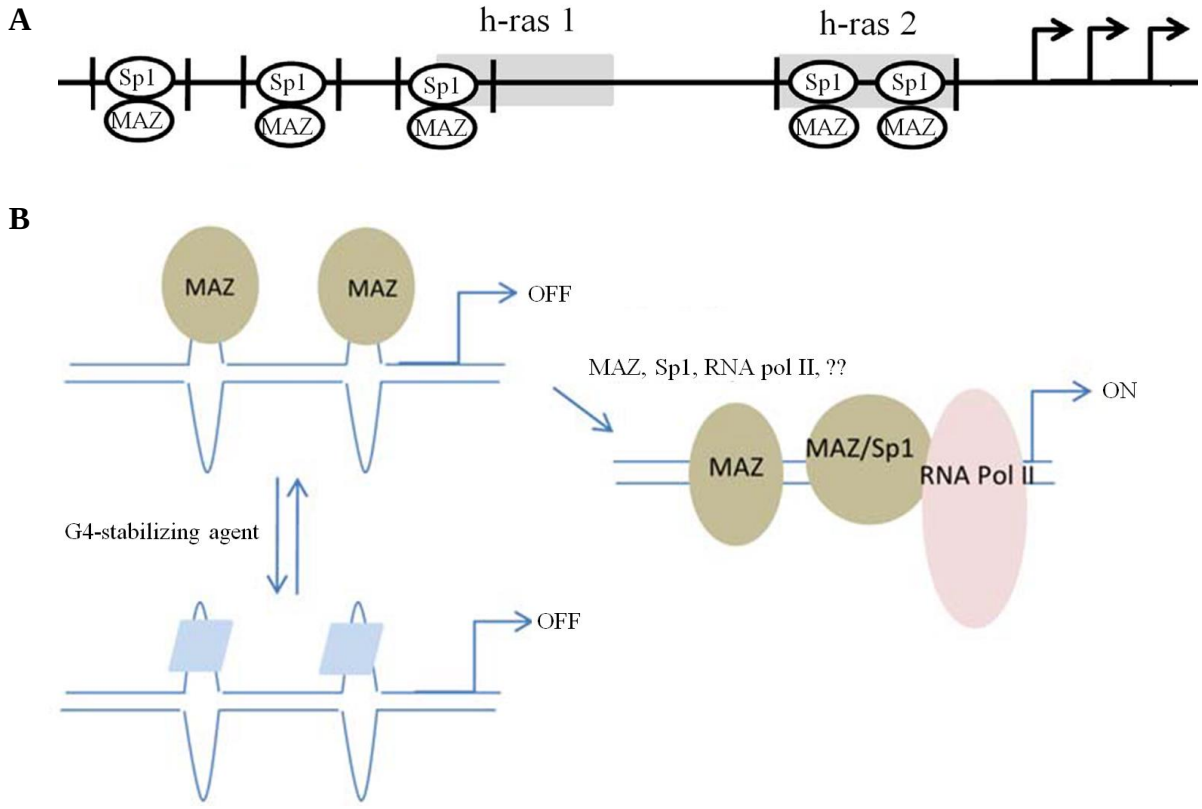


Figure 6: Role of transcription factor MAZ in G-quadruplex-mediated h-RAS gene transcription. (A) The h-RAS promoter sequence exhibits two distinct DNA regions for the formation of G-quadruplex structures (h-ras 1 and h-ras 2), which are recognized by transcription factors Sp1 and MAZ. (B) MAZ is proposed to bind to and unwind existing G4-DNA in the h-RAS promoter. A synergism of MAZ and Sp1 then leads to recruitment of RNA polymerase II in order to activate Sp1-driven gene transcription. Stabilization of G-quadruplex structures by G4-stabilizing agents represses h-RAS expression. Figure taken from [82].

1.2.2.4 Quadruplex unwinding helicases

Another class of specifically interacting proteins comprises G4-unwinding helicases. Bloom's syndrome protein (BLM) and Werner's syndrome protein (WRN) both are representatives of quadruplex unwinding proteins that belong to the family of ATP-dependent RecQ-helicases, which generally contribute to maintaining chromosomal integrity. Bloom's syndrome is characterized by genetic instability, resulting in growth retardation, immunodeficiency and a high incidence of developing cancer. Affected solid tumors or leukemias are emerging due to chromatid breaks or chromosomal aberrations and a high level of sister chromatid exchanges (SCEs). The BLM protein displays sequence homology to WRN helicase. Werner's syndrome in turn is characterized by chromosomal dysfunctions, resulting mostly in premature aging and respective earlier incident of death. BLM is able to unwind both duplex- and quadruplex-DNA in 3'-5' direction, however exhibits higher affinity for the latter [131]. WRN additionally possesses 3'-5' exonuclease function [132]. Its ability of functional unwinding of G4-DNA is limited to certain quadruplex structures [133]. Both helicases contribute to telomere maintenance [134] and are inhibited by quadruplex stabilizing agents and telomerase inhibitors [135].

Another ATP-dependent helicase which does not belong to the family of RecQ-helicases is the protein product of the gene DHX36, also named RHAU. DHX36 binds to RNA-, as well as DNA-quadruplexes with high specificity, with preference for the latter. It provides the major contribution to G4-resolving activity in HeLa cells [136, 137]. Supposedly, DHX36 is overexpressed in tumor cells, thereby co-regulating the expression of different proto-oncogenes [138]. Previous studies revealed modulating effects of the helicase on the expression of transcription factors YY1 [138] and PITX1 [139].

Although the mentioned helicases exhibit varying affinities and selectivity for their substrates, a common mechanism of action of unwinding G4-DNA was proposed, which is based on repetitive successive resolving cycles [140].

1.2.2.5 Quadruplex stabilizing agents

As mentioned, the development of new G4-interacting agents for use in anti-cancer therapy is experiencing brisk interest. A variety of substances have become established quadruplex-stabilizing agents in current studies. In general, DNA secondary structures can be targeted in different ways, depending on the kind of ligand-interaction taking place, including stacking effects on top of the quadruplex, which is due to π-π-interactions of the agents chromophore, binding between grooves or non-helical loops [70, 141]. Classes of small molecules of

common use comprise acridine analogues (BRACO19), cationic porphyrins (TMPyP4) and substituted perylenes (PIPER).

One of the most established interacting agents is represented by meso-tetra-(N-methyl-4-pyridyl)-porphyrine (TMPyP4), which belongs to the cationic porphyrin analogues. TMPyP4 is commonly used for detecting quadruplex structures by spectroscopic methods or giving evidence for transcriptional regulation by stabilization thereof. Furthermore, TMPyP4 is an inhibitor of telomerase [142, 143] and exhibits either G4-end-stacking or external interactions, depending on the quadruplexes' topology [144, 145], with preference of binding parallel over antiparallel ones [145]. In comparison, its isomer TMPyP2 (figure 7) displays weaker or no ability to bind to or stabilize G4-DNA [106]. Since TMPyP4 also shows concentration-dependent affinity to duplex DNA, a complete G4-selectivity cannot always be granted [106]. However, evidence was given for the repression of transcriptional activity of both the PDGF-A and c-MYC genes due to interactions of TMPyP4 and quadruplex structures in the respective promoter regions [106, 143]. TMPyP4 additionally was shown to unwind RNA-quadruplexes [146].

Further synthetic small molecules to target quadruplex DNA comprise pyridostatin (PDS) and its derivatives [147] which likewise interact by terminal G-quartet stacking [148]. PDS competes with telomere-associated proteins for their quadruplex binding. Stabilization thereof subsequently induces telomeric dysfunction and presumably contributes to DNA damage and growth arrest in different cell lines [149]. Pyridostatin exhibits high selectivity over double-stranded DNA [147]. Its G4-stabilizing impact is not limited to telomeres and was successfully used for targeting quadruplex structures of several proto-oncogenes. Tumor suppressing effects could, among others, be demonstrated for the gene products of SRC [109] or BCL-2 [92].

One of the most potent G4-stabilizing agents is represented by natural available macrocyclic telomestatin. Analogous to PDS it exhibits high selectivity for binding quadruplex structures compared to duplex DNA, with high affinity for intramolecular quadruplexes [70]. Telomestatin additionally facilitates quadruplex formation, even in media without stabilizing monovalent cations [69] and is a potent specific inhibitor of telomerase [150].

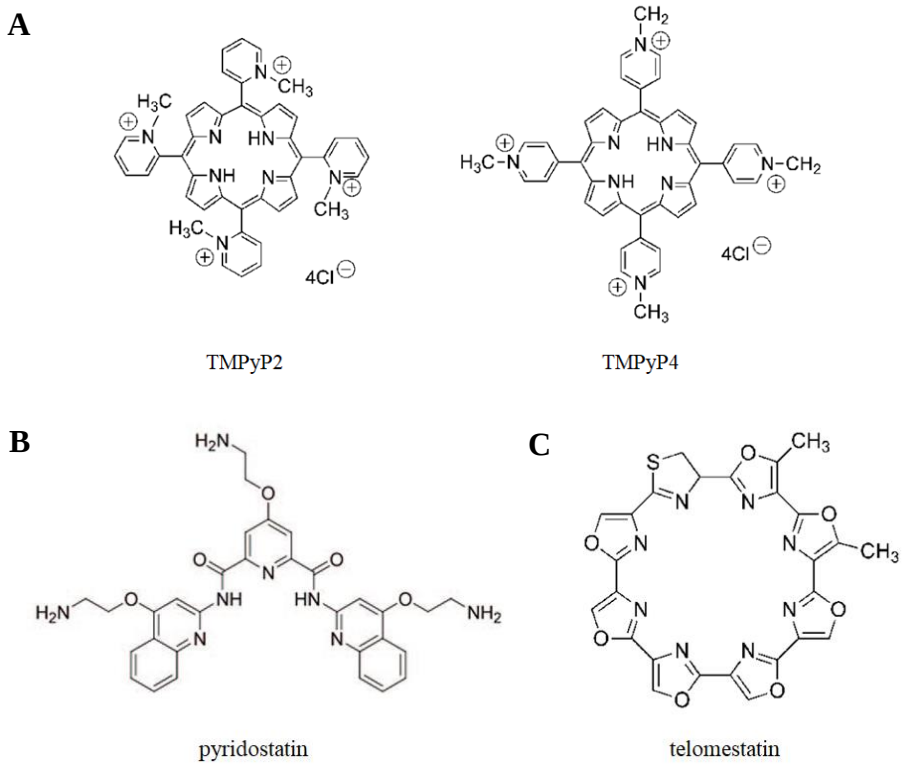


Figure 7: Overview of selected G-quadruplex-stabilizing agents. (A) Cationic porphyrin analoga TMPyP2 and its isomer TMPyP4 [106], *(B) pyridostatin* [147] *and (C) macrocyclic telomestatin* [106].

1.2.3 Methods for detection

Widespread used methods for the *in vitro* detection of present DNA quadruplex structures comprise most notably spectroscopic methods and gel shift assays.

Interactions between TMPyP4 and quadruplexes can be monitored photometrically to indicate the presence of G4-DNA, as well as the potential way of interaction or binding stoichiometry between quadruplex and the porphyrin. UV-VIS spectroscopy of stabilized quadruplexes thereby generate a noticeable bathochromic and hypochromic shift on the absorbance of the Soret band of the solely present ligand, which serve as indicators of the binding extent [144]. Because of their simplicity, circular dichroism (CD) studies are commonly used to gather structural information on quadruplexes alone or ligand-bound. CD allows to distinguish between parallel and antiparallel topology. Generally, parallel tetraplex DNA displays a positive band at 270nm and a negative band at 240nm, whereas antiparallel quadruplexes show a positive and a negative band at 295nm and 260nm, respectively [70, 151].

Gel-based methods mostly include native gel electrophoresis or electrophoretic mobility shift assays (EMSAs), which can be either used for proving quadruplex existence or providing

information about G4-DNA-ligand interactions, respectively. Due to its compactness and smaller surface, quadruplex DNA migrates quicker through the gel, compared to denatured DNA of the same size and mass, allowing for drawing conclusions of G4-formation. Additional incubation with stabilizing agents in turn results in band shift and decelerated migration of the complex when compared to free quadruplex DNA [66, 69].

The development of a quadruplex-sensitive antibody (BG4 antibody) recently enabled the detection of G4-DNA by enzyme-linked immunosorbent assay (ELISA) [152]. The antibody holds nanomolar affinities for quadruplex topologies of any kind, while not displaying any binding capacity for single- or double-stranded DNA. The use of fluorescence-coupled secondary antibodies against BG4 further allowed the visualization of genomic G4-DNA in human cells.

More recently, the detection of quadruplex DNA became available for mass spectrometry (MS). For detecting nucleic acids by MS the measurement in negative ion mode is generally approved. To provide buffer conditions that are compatible with MS but still maintain cation-stabilized quadruplex structures, conventional Na^+ and K^+ containing buffers are replaced byones containing NH_4^+. The use of ammonium acetate buffers became established for this purpose. Usually, ammonium stabilized quadruplexes are detectable with their number of G-tetrads (n) in association with the respective number of bound NH_4^+-ions (n-1) in different charge states. G4-ligand interactions can be easily comprehended by their mass shift. The detection with ion mobility-MS can additionally help to provide information on conformation or topology of the quadruplex structure [68, 153–155].

The in 1.2.2.5 already mentioned effects of quadruplex stabilizing agents on the expression of various gene products obviously also provides a means of detecting G4-DNA and their respective impact on biological level.

Further methods comprise for example nuclear magnetic resonance (NMR) spectroscopy or fluorescence resonance energy transfer (Förster resonance energy transfer, FRET) [66].

1.3 Mass spectrometry

1.3.1 Proteomics

The successful completion of the human genome project in the beginning of the 2000s [156] delivered a full sequencing of all human genes and hence ongoing motivation for deciphering the entirety of RNA, proteins or metabolic pathways present in the organism, which can be termed transcriptome, proteome or metabolome, respectively. The elucidation of the proteome therefore comprises the thorough knowledge of existing proteins, their subcellular appearance

and interactions, the understanding of their appropriate functions and the discrimination between physiological and pathophysiological conditions. The dynamic, highly complex nature of the proteome exacerbates the complete understanding of proteins present in a biological system, since their biosynthesis and interactions are regulated in various ways and protein activities are controlled by diverse modifications. In this context, proteomics represents a powerful tool for providing consistent new insights of functions, post-translational modifications (PTMs) or protein-protein interactions.

In the following, explanations are given for relevant terminologies or general methods that are valid and used commonly in mass spectrometric applications.

Proteomic workflows: top-down, bottom-up

The classical workflows of proteomic approaches can basically be divided into ‘top-down’ or ‘bottom-up’ [157–159], the latter also named shotgun proteomics and being the more popular method for the identification and characterization of proteins. Top-down approaches introduce intact proteins into the mass spectrometer; bottom-up analysis requires the digestion of proteins into peptides prior to MS measurement. Proteins are then identified based on their present peptides. Top-down analysis demands less sample preparation and generally generates less complex mass spectra, resulting in an easier data analysis. Intact mass measurement however requires mass analyzers that are not limited to a certain m/z-range. Due to generated mass shifts, top-down approaches distinguish between unmodified and modified proteins that help to elucidate possible isoforms or PTMs. Subsequent fragmentation of a protein of interest further supports the identification of different combined modifications. However, the need for low complex samples limits the applicability for top-down analysis, thereby excluding high throughput methodologies [160]. These can be obtained by using bottom-up approaches [161]. Even highly complex protein mixtures can be investigated after proteolytic digestion and subsequent peptide measurement and database search. The constant improvements in high resolution MS techniques currently enable the identification of whole cell lysates in a single run by coupling LC-ESI-MS/MS.

Sample preparation

Sample preparation is an essential step in customizing biological material for consecutive mass spectrometry and depends on the respective issue of investigation. Most of the time, the enrichment of proteins of interest via different methods is pursued, so that they will not be dominated by highly abundant proteins afterwards. Therefore, it is advisable to subject biological samples to subcellular fractionation to reduce the complexity of starting material [162]. For further reduction of complexity of cell compartments on protein level, methods

including gel electrophoresis for protein separation, affinity chromatography based on metal affinity (IMAC), immune-affinity (via antibodies) or interaction affinity are available [163]. The latter represents a widely used method for studying protein-protein-, DNA-protein- or RNA-protein-interactions by means of bait-prey affinity pulldowns.

The adjustment of buffer conditions during sample preparation to resemble physiological buffer systems often generates the necessity for purifying processed peptide samples to ensure compatibility with mass spectrometry. Detergents or salts are a major source of ion suppression in MS measurements and demand change of buffers to contain volatile components [164]. Protein precipitation or the use of C18-material for purification on peptide level provide methods for changing buffer system and desalting samples prior to MS.

For proteolytic digestion several proteases are available with different specificities concerning their respective cleavage sites. The most commonly used protease is trypsin, as it provides high stability in various conditions and displays adequate substrate specificity [165, 166]. Trypsin cleaves at C-terminal ends of lysine and arginine residues and generates peptides that generally feature suitable masses for being accessible by MS. Other proteases include ArgC, GluC or elastase, the latter cleaving more unspecifically after several amino acids. For methodical studies or the improvement of digestion efficiencies, a combination of proteases is advised.

Identification of proteins: Peptide mass fingerprint and tandem-MS

Shotgun approaches provide two possibilities of identifying proteins deriving from biological samples, based on the number of MS-analyses performed. MS-scans of proteolytic peptides yield their respective masses and characteristic patterns belonging to correlating proteins, by which means they are identified by database search. For significant identification based on these peptide mass fingerprints (PMFs), only a small number of peptides per protein often suffice, however this method is essentially limited to isolated proteins and lacks suitability for complex samples [167–169].

Due to fragmentation of amino acid sequences, tandem MS-analysis (MS/MS, MS^2) of certain precursor ions enables a more discriminating technology of identification [170]. Proteins can be identified out of highly complex samples based on their peptide-related tandem MS measurements. The repetitive mass analysis is not limited and often is termed MS^n depending on the cycle of measurements carried out. Routinely, a combination of LC-ESI-MS/MS is performed for the analysis of biological cell samples, to ensure high identification rates of present proteins [161, 163].

1.3.2 DNA-protein interactions in proteomics

Reducing the complexity of cell samples by using isolated cellular compartments instead of whole cell lysates facilitates proteomic accessibility of present protein components, since their detection will be more sensitive than without fractionation. Studying the nuclear proteome gathers knowledge about components involved in the organization and regulation of genetic information. The MS-based investigation of transcription factors or chromatin-remodeling proteins and their complexes helps to enlarge the understanding of gene expression and subsequent effects on processes including cell cycle, metabolism or differentiation of different cell types [171–173].

Transcription factor proteomics still delivers new information about highly complex regulatory mechanisms of activation and inhibition of promoter regions of different genes based on DNA-protein interactions. The prevalent problem of their detection arises from their low abundance that aggravates analytical analysis, ranging from estimated 0.01-0.001% of total cellular protein content [174]. Therefore, enrichment techniques for successful detection of transcription factors are inevitable. The most important and most used methods for detecting DNA-protein interactions in proteomics and beyond are summarized in table 2.

method	functional principle	reference
in vitro		
EMSA	radioactive/fluorescence-labeled DNA-protein complex is retarded in non-denaturing PAGE compared to free DNA	[175–177]
EMSA-supershift	addition of specific antibody to detect one of the protein components in the complex, generates additional shift	[175, 178]
DNAse footprinting	radioactive-labeled DNA-protein complex is treated with DNAse; sequence bound to TF is protected from cleavage	[172, 175]
DNA affinity chromatography (DNA pulldown)	DNA sequence is bound to solid support; serves as bait to enrich and isolate pray protein-complexes, MS compatible	[172–174, 179]
in vivo		
chromatin-immunoprecipi-tation (ChIP)	protein components are crosslinked to DNA, DNA-protein complex is enriched by using a specific antibody, MS compatible	[175, 180]

Table 2: Overview of major techniques used for the identification of DNA-protein interactions

Affinity purification (AP) is based on capturing prey proteins with the aid of a specific bait, which can be a protein, antibody or a specific nucleic acid sequence. The currently used MS-coupled methods for the detection of DNA-protein interactions are mostly restrained to DNA-pulldowns and chromatin-immunoprecipitation. For DNA pulldowns, the oligonucleotide sequence of interest is immobilized on a solid support and incubated with a fractionated cell lysate. Immobilization can be conducted prior (direct method) or after (indirect method) incubation with cell extracts. For the successful discrimination of enriched versus non-enriched prey proteins, a suitable control is essential, which can either be a mutated or scrambled version of the DNA-sequence. Thereby enriched proteins are further subjected to mass spectrometry for identification [181]. The choice of the control sequence depends on the aim of the study. The investigation of single nucleotide polymorphisms (SNPs) within the context of genome-wide association studies (GWAS) for example, demands punctual mutations of the respective transcription factor consensus sequences [182, 183]. More general approaches for a basic screening of DNA-binding partners of certain sequences revert to scrambled sequences for a respective control.

Commonly used solid supports are provided by agarose or sepharose beads, as well as streptavidin-coated magnetic particles [172]. During experimental planning, non-specific binding of highly abundant proteins to the matrix has to be considered, which aggravates data interpretation and can lead to false positives when performing qualitative analysis [174, 184, 185]. Detection of nonspecific interactors can contribute to an amount of up to 95% of identified proteins. Depending on the selected matrix surface, different distributions of protein classes of bead contaminants are observed, designated as ‘beadome’ and have to be taken into account [186]. To minimize nonspecific binding, DNA-competitors including sheared salmon or herring sperm DNA or poly-dIdC are often added during incubation. Alternatively, concatamerization of the selected sequence enlarges the number of present consensus sequences, while reducing free DNA ends that are mostly responsible for nonspecific binders [182, 184]. Additionally, harsh washing conditions after protein binding shall attempt to wash off unspecific interactors due to their low binding affinity. However, high salt concentration buffers or denaturing conditions are able to disrupt transient complexes and are incompatible with mass spectrometry, so that further purification is inevitable, which in turn goes along with protein loss [187].

Usually, subsequent 1D- or 2D gel electrophoresis is carried out for the identification of in-gel digested protein complexes afterwards. Alternatively, direct on-bead digestion and consecutive LC-MS/MS is performed, which demands quantitative proteomic data analysis.

Recent development of methods used in quantitative interaction studies and their respective bioinformatics analysis that is based on differences in protein abundances, helped to overcome the mentioned difficulties of qualitative DNA-pulldowns. The problem of nonspecific binding was solved by the absence of enrichment of background binders in both wildtype (WT) and control (CTR) experiment and resulting equal abundance, so that specific and nonspecific interactors are securely discriminated [187–192]. Accordingly, specifically interacting proteins exhibit significant comparative enrichment in this setting and are subsequently identified by statistical analysis. Relinquishing DNA concatamerization and harsh washing buffers is beneficial for retaining conditions as physiological as possible to maintain transient protein complexes as well [190, 192]. Direct on-bead digestion and consecutive LC-MS/MS measurement of the isolated DNA-protein complex currently provides established methods for DNA-pulldowns. The development of high resolution mass spectrometers further extended the limit of detection of low abundant proteins. Coupled LC-ESI-MS/MS systems allow for identifying whole cell lysates in one run, thereby superseding the classical 2D gel electrophoresis [173, 174, 181, 193, 194].

1.3.3 Quantitative proteomics

Due to the already mentioned difficulties of portraying dynamic protein interactions, a paradigm shift was taking place in the beginning of the 2000s by replacing qualitative MS-methods by quantitative ones. The focus on qualitative protein analytics for the identification of protein complexes by applying subtractive wildtype and control experiments was superseded by comparative analysis of protein intensities in the respective experiments. Quantitative proteomics provides the benefit of not only relying on mere presence or absence of interacting proteins in different samples, but rather to draw on differential enrichment through distinct abundances for their identification [188, 195]. Background binders yield a ratio of 1:1 when comparing their intensities in WT and control approach, since they are present in same amounts, specific interactors account for ratios that distinguish thereof. Thereby, a successful discrimination of nonspecific and specific interactors can be achieved [189, 191, 193, 196]. The shift from qualitative to quantitative proteomics currently enables fundamental high-throughput shotgun proteomics for elucidating a variability of biological issues. Different labeling strategies, as well as label-free methods have evolved to facilitate identification of DNA-protein or mostly protein-protein interactions.

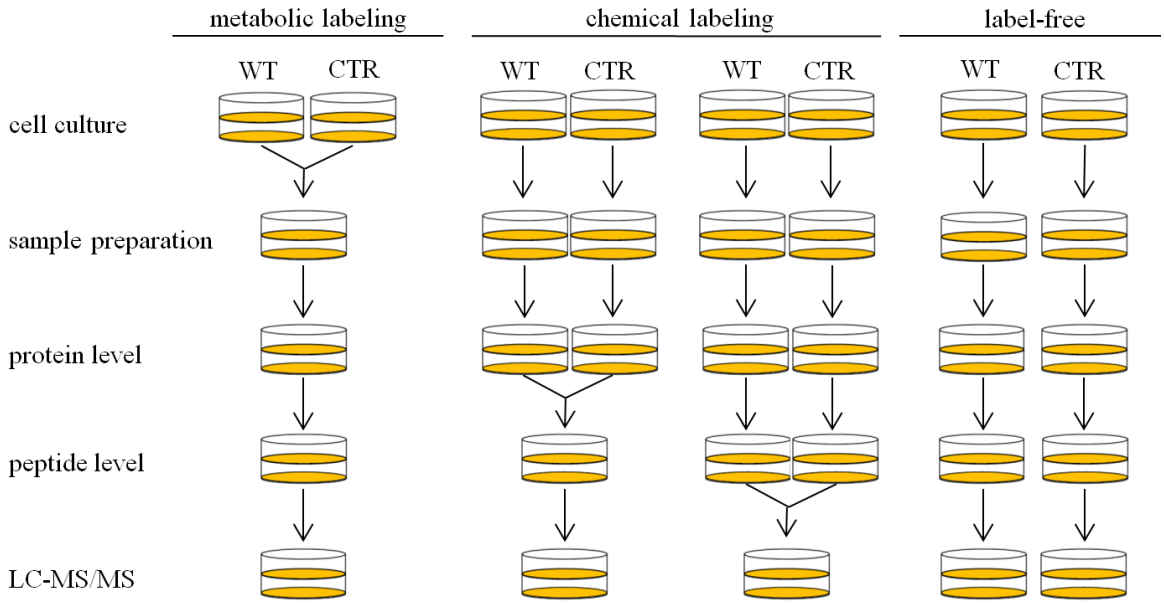


Figure 8: Overview of different quantification strategies in quantitative proteomics. Methods can be distinguished by their respective type of labeling reaction or their point in time of combining wildtype and control sample.

Label-based quantification can essentially be separated into metabolic or chemical labeling and, in correlation to the point of time of label incorporation, into labeling on cell cultural, protein or peptide level, after proteolytic digestion (figure 8). Label-free techniques provide comparative results without combining samples prior to analysis, which leads to lower quantification accuracy and precision than label-based methods, since an introduced bias in the following downstream sample preparation cannot be adjusted afterwards [189, 197, 198]. The choice of quantification strategy and its respective advantages and disadvantages depends among others on the used input material, number of samples prepared and sample complexity. Apart from high costs for labeling-kits or agents, quantitative proteomic methods afford cost-effective identification of interactome complexes, since the number of LC-MS/MS runs can be held at minimum by combining WT and control samples compared to qualitative proteomic samples deriving from gel-based approaches [187]. Within one MS run, up to several thousand proteins can be identified [199]. With the development of commercially or freely available bioinformatics software for the analysis of complex datasets deriving from interaction studies, the application of quantitative proteomics strategies established itself for the identification of differential protein expression or binding patterns in varying fields, including human biology or clinics [197], microbiology [200] and genomics, as mentioned by

present SNPs [189]. An overview of the different most commonly used quantification techniques and their respective major characteristics is given in table 3.

	metabolic labeling (e.g. SILAC)	chemical TMT, iTRAQ	labeling dimethyl labeling	label-free
pro	-sample combination on cell cultural level -high precision and accuracy	-multiplex approaches -no increase in spectrum complexity, as quantification on MS^2 level	simple and cost-effective	-no need/costs for extra media/reagents -simple method
con	-limited application on cell material -time-consuming costly method	expensive reagents	deuterium-induced differences in retention times	high variability due to separate sample preparation

Table 3: Overview of quantification strategies commonly used in mass spectrometry

1.3.3.1 Label-based quantification methods

SILAC (stable isotope labeling by amino acids in cell culture) [201] is based on the concept of incorporating ^{13}C or $^{13}C/^{15}N$ isotope variants of essential amino acids into cell culture conditions, thus belongs to *in vivo* labeling strategies. The addition of heavy arginine or lysine facilitates tryptic digestion afterwards, thereby ensuring quantitative accessibility of most peptides [187, 200]. Two experimental groups of interest are cultured in media containing either light or heavy labeled amino acids, respectively. Labels are incorporated on cell cultural level, completion being achieved after estimated five cell doublings [201]. Both sample conditions are combined on cell cultural level and further processed as one, providing minimal variability of downstream preparation thereby resulting in highest precision and accuracy of quantification of all methods available [187, 188, 197]. Quantification of purified proteins occurs on MS^1 level by measuring relative peptide intensities, which arise during MS runs of each sample and their respective mass shift [187, 188]. Since heavy labeled amino acids are incorporated to 100% into the cell system investigated, no bias in labeling efficiencies is generated [201], furthermore a complete theoretical sequence coverage is conceivable. One of the main drawbacks of quantification by SILAC is its limitation to cell culture material [187], although its application could be extended to tissue samples with the

implementation of Super-SILAC [202, 203]. Currently, SILAC still is considered to be gold standard of quantification methods for its accuracy and reproducibility [196, 204] and is widely used in fields of differential protein expression profiles [205–207], posttranslational modifications (particularly phosphoproteomics) [208–210], histone modifications [211, 212], lysine acetylation [213, 214] or interaction studies for protein-protein interactions (PPIs) [215–217] or DNA-protein interactions [218, 219], as part of AP-MS studies.

ICAT (isotope-coded affinity tags) [220] bears on chemical *in vitro* labeling on protein level. The basic ICAT reagent is composed of three elements: an affinity tag, generally biotin, a deuterated or ^{13}C labeled linker element and a reactive group with specificity toward sulfhydryl residues [199]. Cysteine residues of protein samples of two cell conditions are either labeled lightly or heavily, combined and subsequently processed by proteolysis. The resulting cysteine-tagged peptides are further isolated by avidin-affinity chromatography and analyzed by mass spectrometry. Peptide pairs of the same sequence of both conditions generate a mass shift and are quantified based on comparative signal intensities. Since shared peptides are chemically equivalent, there is no bias to labeling efficiencies [220]. Peptides are likewise quantified on MS^1 level. ICAT exclusively identifies cysteine-containing proteins, which has to be taken into consideration prior to analysis, as on the one hand sample complexity is reduced, but on the other hand proteins that only contain a small amount of cysteines are likely to be missed [200, 221]. Thus, ICAT only provides limited proteome sequence coverage [200]. Furthermore, altering fragmentation behavior due to the label has to be taken into account, as well as primary isotope effects deriving from the use of deuterium, possibly resulting in differential retention times during LC separation [193, 222]. Therefore, the application of ICAT for quantification is mainly restricted to differential protein expression profiles and protein complexes [193, 195]. However, cysteine reactivity proved to be beneficial for studying redox reactions, thus ICAT currently is the method of choice for the analysis of the redox proteome in different organisms [223–227].

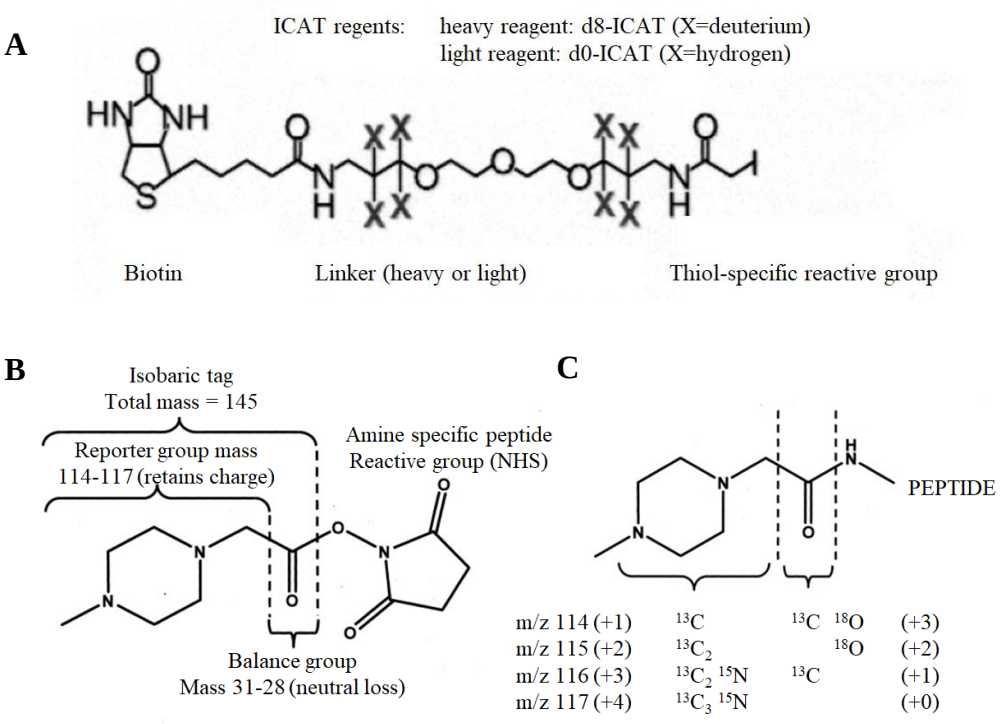


Figure 9: ICAT and iTRAQ labeling reagents (A) the ICAT reagent [220] *and (B) the iTRAQ reagent. (C) The mass of the isobaric tag is kept constant by varying different combinations of used isotopes* [228].

In contrast to isotope labeling strategies, iTRAQ (isobaric tag for relative and absolute quantification) is based on introducing isobars on peptide level [228]. iTRAQ reagents consist of three elements as well, including an amino-reactive group (NHS-ester), a mass balance group (carbonyl function) and a reporter group, which introduces the isobaric tag. The total mass of reporter- and balance group is kept constant with the combined use of ^{13}C, ^{15}N and ^{18}O isotope variants, thereby generating differently labeled peptides that exhibit the same m/z-ratio in MS^1 level, but after fragmentation on MS^2 level and neutral loss of the mass balance group result in shifted m/z-ratios of the reporter ion. Quantification of labeled amino-termini or present lysine residues therefore takes place on MS^2 level [228–230]. Since different iTRAQ reagents are available, currently up to eight samples can be compared to each other (multiplex approach). Limitation is given through the increase in complexity which is due to labeling on peptide level as well as enhanced signal complexity in MS/MS spectra, which impairs data analysis at some point [229] and lowers peptide quantification quality [231]. Furthermore, depending on the chosen MS/MS method peptides are quantified only partially, with bias to high abundant proteins [232]. The accuracy and precision of quantification of the technique is inferior to labeling methods on protein or cell cultural level and was reported to

show a trend towards underestimation of peptide ratios [233, 234] as well as to be prone to interferences during MS-analyses due to co-eluting precursor ions [234]. iTRAQ is used in the investigation of differential protein expression profiles [235] plasma proteome [231] and plant proteome research [236].

The development of dimethyl labeling [237] extended the labeling strategies on peptide level and because of its simplicity and efficiency established itself as a commonly used alternative to metabolic labeling. Amino-termini and lysine residues of proteolytically treated protein samples are reductively dimethylated after the addition of formaldehyde and cyanoborohydrid. Depending on the used isotope variant of the dimethylating reagent, light, intermediate and heavy labels can be introduced that generate mass shifts of 28, 32 or 36 dalton, respectively [238, 239]. The combination of up to three different conditions allows for quantification on MS^1 level (Figure 10). The successful identification rate by peptide-to-spectrum matches (PSMs) and the respective quantification efficiencies are comparable to those of the SILAC technology [240], however dimethyl labeling provides less precision [241]. The labeling reaction itself does not generate interfering side products and proceeds almost completely quantitative, with one exception being the monomethylation in case of an N-terminal proline [239, 242]. Similarly to the ICAT technique, the use of deuterated formaldehyde can lead to retention time shifts during LC-run, which however were reported to have minor effects on quantification accuracy [237, 239] when quantification is performed based on extracted ion chromatograms (XICs) rather than single scans. Due to its predominant benefits in practicability and quantitative efficiency, dimethyl labeling experiences widespread application for studying protein expression patterns [243, 244], PTMs [245–247], glycoproteomics [248, 249] or interactomics, notably AP-MS studies [183, 213, 250].

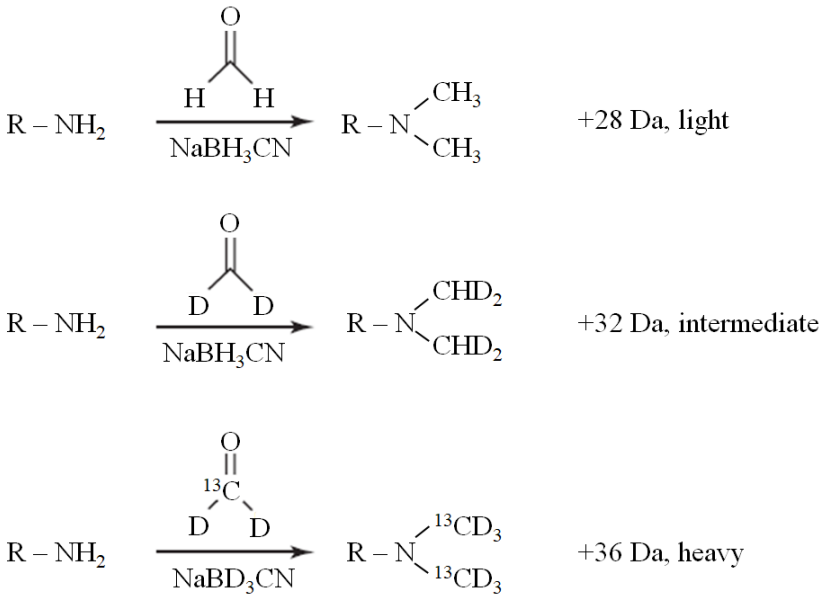


Figure 10: Schematic overview of dimethyl labeling based reactions [238].

1.3.3.2 Label-free quantification (LFQ)

The use of label-based strategies often is associated with extra costs and limitation of sample conditions being compared to each other. In this context implementation of alternative quantification techniques without the need for additional reagents and no restriction to the number of comparative analyses was highly interesting. Development of suitable bioinformatics software for computer-assisted statistical analysis of peptide abundances in separate LC-MS/MS runs widened the application of LFQ-techniques. As the sample of interest and its control are not combined in this approach, quantification accuracy is inferior to label-based techniques, therefore generally demanding higher numbers of replicates. However, label-free quantification currently still provides one of the most used methods. Quantification itself is either based on spectral counting or comparative ion intensity measurements, the latter being the established method of choice.

Spectral counting is based on the assumption that the amount of a certain protein correlates to its number of recorded MS/MS spectra [251]. Since longer proteins potentially generate more peptides than shorter ones and to further diminish variation of quantification within replicates, normalization factors were needed and introduced in the normalized spectral abundance factor (NSAF) [252]. Another consolidation of several spectral counting quantifications into one calculation is given in the normalized spectral index (SI_N). It considers spectral counts, the number of unique peptides to its corresponding proteins and fragment ion intensities, thereby improving fluctuating variability of replicates and robustness

of quantification compared to other abundance factors [253]. Spectral counting can provide reliable quantification results, assuming a suitable normalization.

Intensity-based quantification is based on the proportionality between signal intensity and concentration of a respective protein that is given at same retention time in separate LC-runs [251] and is equivalent to superior to spectral counting in terms of accuracy of quantification and reproducibility between replicates [192, 254–257]. Reproducibility corresponds to that achieved by metabolic labeling [198]. To minimize differences occurring in retention times that would lead to impaired quantification accuracy, retention time alignment should be performed. Generally, the three most intense peptide signals of each protein are included to determine its quantity [192]. For differential protein expression analysis in AP-MS experiments, outlier determination during data processing mostly is performed via t-test-based statistics. Therefore, a minimum of three replicates is required, which are compared to the same number of control experiments, eventually resulting in the identification of significant interactors. After a statistical step called missing data imputation, t-test statistics provide the advantage of quantifying even proteins of lowest abundance [192]. For imputation of missing data, randomized values are imputed according to the present protein abundance distribution of a sample, resembling values in a defined range of detected low abundant quantities. This is beneficial for including proteins into statistical analysis, whose signal intensities may be present below the limit of detection in some of the replicates [256, 258].

Label-free quantification methods share the need for bioinformatics software for respective data processing that must provide tools for effective peak picking, data alignment or statistical analysis of the respective raw data. Currently, both commercial and freely available software does exist. Examples for intensity-based quantification comprise commercially available SIEVE (Thermo Fisher), which applies chromatographic alignment algorithms and is used for differential protein expression analysis or Progenesis (Nonlinear Dynamics), which including additional multiple statistical tools [251]. Freely available software comprise for example SuperHirn [259] for protein profiling in complex samples or MaxQuant [260], which is widely used for label-based or label-free analyses in general.

As already mentioned, label-free quantification techniques provide less quantification accuracy than label-based methods, however the number of total identified proteins is generally higher in label-free approaches [257, 258, 261]. Thus, these methods provide the highest proteome coverage along with widest dynamic range of protein identification among all quantification strategies [198, 262]. Label-free quantification by now offers reliable,

comparable use to label-based quantification in studying human-derived samples in clinics [197], protein interaction networks via AP-MS-studies and protein expression patterns [251].

1.3.3.3 Absolute quantification

Although relative quantification plays a more prominent role in MS-based proteomics, absolute quantification shall be presented shortly in the following. Mostly, absolute quantification is carried out using internal standards, which correspond to an isotope variant of the peptide or protein of interest to maintain the same chemical characteristics. These proteotypic peptides are spiked into the sample prior to LC-MS/MS analysis. Subsequent quantification is based on the comparison of ion intensities of the respective peptide and the added standard, optimally measured in selected reaction monitoring (SRM) mode. Synthetic peptides used for this purpose comprise AQUA-(absolute quantification)-peptides. AQUA-peptides generally correspond to tryptic peptides of up to 15 amino acids, which should not contain reactive groups including tryptophan and cysteine residues [263]. By using a combination of non-phosphorylated/phosphorylated AQUA-peptides, a possibility is given to study phosphorylation states of proteins, to elucidate activities of enzymes or pathways of signal transduction [263, 264]. To extend the possible number of quantified peptides in one approach, QconCAT-(quantification concatamer)-technology was established [263, 265]. Here, artificially labeled proteotypic peptide sequences of up to 50 proteins to be quantified are concatamerized and synthesized. Prior to tryptic digestion, QconCAT is added to release signal peptides in a ratio of 1:1 during proteolysis. Subsequently, generated signal peptides can be compared to the non-labeled proteins to be quantified. This method provides excellent predictions of present protein stoichiometries [263, 265]. Both methods share the disadvantage of not capturing processing steps on protein level or prior to protease digestion and consequently can affect digestion efficiencies. To solve this problem, protein standard absolute quantification (PSAQ) was introduced [266], which employs a labeled version of a target protein that is added directly in the beginning of sample processing. Although this strategy comes along with high costs, it is suitable for the quantification of target proteins present in a certain medium, including biomarkers or therapeutic proteins [263, 267].

For label-free absolute quantification both spectral counting-based and intensity-based strategies are available. Factors for estimating absolute amounts of proteins present in a complex sample are provided by the protein abundance index (PAI) or the exponentially modified PAI (emPAI) [268, 269]. It is calculated by dividing the number of detected peptides of a protein by its respective number of theoretically possible tryptic peptides [251].

Intensity-based iBAQ-(intensity based absolute quantification)-values are given by the sum of all intensities of tryptic peptides of a protein, divided by the number of all theoretically ascertainable peptides for the respective protein and provides an estimation of the absolute amount of protein in a sample [270]. iBAQ quantification was successfully used for determining ratios of proteins within an isolated protein complex by AP-MS with comparable quantification accuracy to labeled peptides [271]. A comparative study of some of the available methods displayed superiority of intensity-based quantification over spectral counting ones, in terms of quantification accuracy and reproducibility, as well as proteome coverage [272].

1.3.4 Ionization methods

For the ionization of non-volatile, labile compounds for suitable application in proteomic research, two major soft ionization techniques are available, both developed in the late 80's, termed MALDI (matrix-assisted laser desorption ionization) [273–275] and ESI [276] (electrospray ionization).

MALDI represents a pulsed ionization technique, taking place in the high vacuum of an ionization source and makes use of a co-crystallization of analyte molecules together with a small organic compound that constitutes the matrix. Both components are combined on a sample target prior to ionization and co-crystallize in a thin layer. By irradiating with wavelengths in the range of the absorption maximum of the matrix deriving from an UV-laser, this co-crystal desorbs into the gas phase, while ionizing both analyte and matrix molecules. The utilized matrix supports ionization of analyte molecules by proton-transfer reactions. The ionization process demands a threshold irradiance, which is mostly depending on the chosen matrix. Energy deposition is controlled by the matrix molecules, as these absorb to a higher extent than the analyte. Thereby the latter is being protected from exceeding energetic exposure, which otherwise would lead to its fragmentation [274, 275, 277]. Thus, an excess of matrix is needed to keep energy absorption by the analyte low and prevent a critically high threshold irradiance [274]. Furthermore, by using excess amounts of matrix analyte molecules are sufficiently separated from each other for the ionization process [277]. Typical ratios of matrix-sample mixture range from 1000:1 to 10000:1 [278]. The matrix crystallization and ionization efficiencies co-depend on the sample, which should be considered while choosing a suitable matrix. Successful homogenous crystallization of the analyte in the matrix compound is crucial for sufficient ionization capacity [275, 279]. After

ionization, the generated ions are then introduced into the mass analyzer by applying an accelerating voltage.

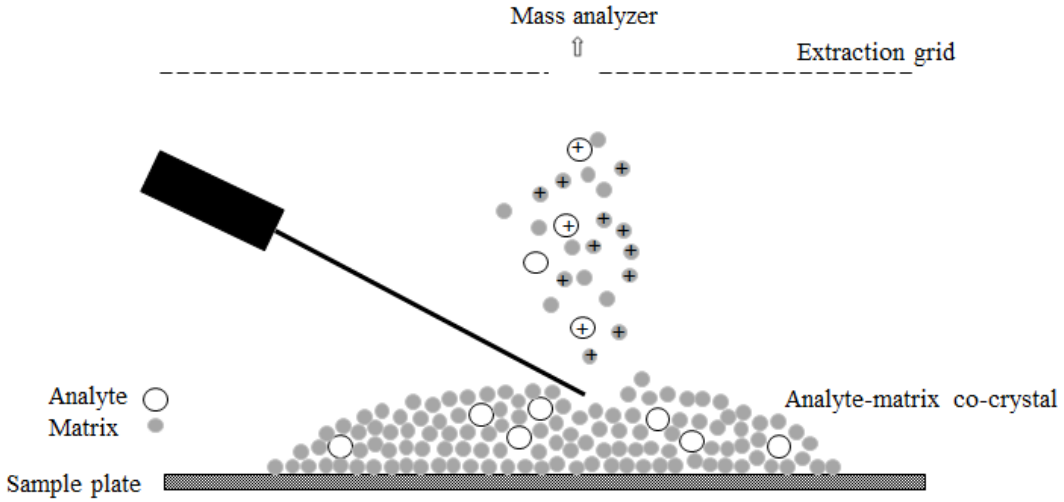


Figure 11: MALDI process.

The MALDI process mostly generates singly charged analyte ions, the exact mechanism of ionization still lacking detailed insight. An extensive explanation is provided by the lucky survivor model [277]. It postulates matrix-molecule clusters to have tendency of performing neutralizing reactions after laser irradiation and desorption. More highly charged analyte molecules display higher possibility of losing their charge than singly charged ones. Since the neutralizing reaction represents the prevalent process, singly charged ions exhibit highest probability of 'surviving'.

ESI employs the use of electrostatic fields to desorb ions from an aerosol continuously. This process is taking place in atmospheric pressure. The dissolved analyte is positioned into the ionization chamber with the help of a steel needle, consisting of the needle itself surrounded by a cylindrical electrode that is purged by an inert gas (mostly tempered nitrogen to evaporate the solvent). On both components a voltage is applied. The potential that is applied at the end of the needle leads to the charging of the emerging liquid that contains the sample. Due to the existent electrical field a shift of charges is induced in the liquid, eventually resulting in the formation of the Taylor cone [280], out of which a current of liquid sample is vaporized. Present Coulomb forces promote the dispersion thereof into finely charged droplets, containing the analyte molecule. These droplets further migrate towards the skimmer electrode and a transition of differentially pumped stages to the mass analyzer, which is operated under vacuum conditions. While migrating, the solvent of the droplets is partly

evaporated, resulting in reduction of their respective sizes and consecutive increase in charge until they reach the Rayleigh limit; this equilibrium of Coulomb repulsion and surface tension leads to Coulomb fission to generate smaller daughter droplets that in their turn further pass through the same process several times. This repetitive reduction of droplet sizes eventually leads to the desorption of ions into the bath gas present [276, 281].

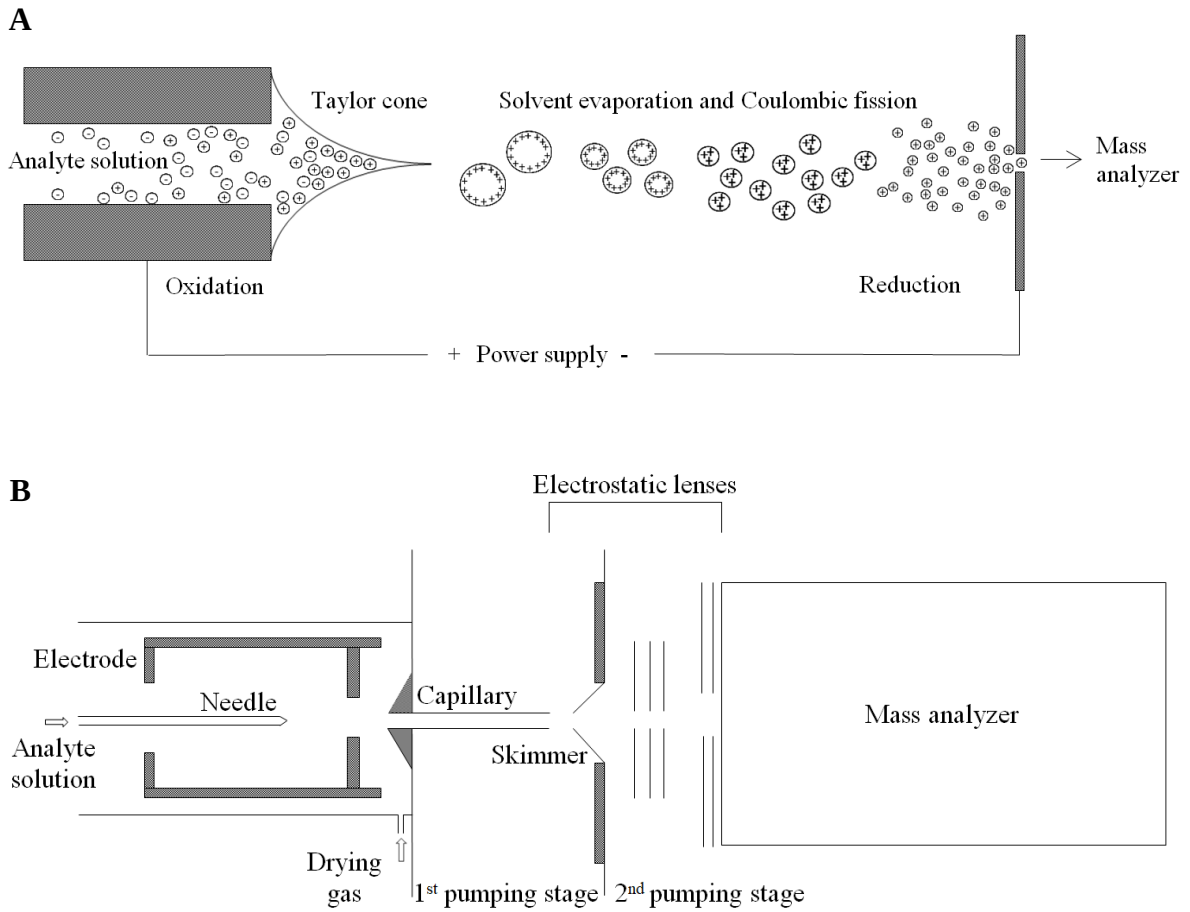


Figure 12: (A) ESI process. (B) Schematic simplified representation of an electrospray mass spectrometer adapted from [276].

Different theories do exist for the release of ions out of the droplets. The charged-residue model (CRM) resorts to the described process of consecutive droplet size reduction. For sufficiently diluted analyte molecules, the repetitive size reduction results in single molecules present in these droplets. Solvent residue evaporates to release the analyte, which retains the charge, so that free ions are generated in the gas phase. Thus, the CRM is proposed to be suitable for large or globular analytes [281, 282]. The ion evaporation model is another model, which postulates the desorption of small analyte droplets by means of a sufficiently high field strength before reaching the Rayleigh limit. Droplets of altering sizes that contain

single ions or ion clusters are able to desorb out of the liquid and release charged ions after complete evaporation [281, 283, 284].

The ESI process generates multiply charged molecules; charge distribution depends among other facts on the size of the analyte, its content of basic amino acids and its tertiary structure [276, 284, 285]. Protein conformations often reduce the maximum of possible charge states, as compared to disrupted conformations due to broken disulfide bridges, which enable higher charge states [281, 285]. The major advantage of multiple charges is based on the mass spectrometric accessibility of large biomolecules and their respective m/z-ratios, which possibly would not be detectable in mass analyzers of a limited mass range [276, 285, 286]. Another great benefit is given by well-established coupling to LC-MS/MS systems, which provide access for highly complex biological samples [284].

1.3.5 Mass analyzers

There are currently several mass analyzers for mass spectrometric measurements that differ in characteristics including mass range, ion transmission or resolution achievable and are more or less compatible with the mentioned ionization techniques. The time-of-flight analyzer (TOF) is mostly coupled to MALDI ionization, as both rely on pulsed functionality. After the ionization process, ions are accelerated by a voltage being applied to enter the TOF analyzer and are separated in the field free drift tube of the analyzer itself, according to their differences in m/z-ratios and velocities, resulting in time differences being detected [287, 288]. Operated in the linear ion mode, TOF analyzers only display low resolution. On the one hand, analyte molecular ions of the same mass exhibit a spatial spread during ionization process, thereby being located in differing starting positions of acceleration [288]. On the other hand, the same ions yield a spread of initial velocities from the ablation process [288, 289]. To compensate for these effects, delayed ion extraction, as well as reflector TOF analyzers are used. Both techniques increase resolution considerably [288–290]. By using delayed ion extraction, generated ions are not accelerated directly, but after a certain interval [288], which generally amounts to about 100 nanoseconds. The time delay results in energetic focusing and enables time-adjusted detection of ions of the same m/z-ratio and a correction of the initial energetic distribution [290, 291]. The use of reflectron analyzers both increase resolution and sensitivity [289]. These consist of consecutively arranged ring electrodes located at the end of the drift tube, which likewise generate a homogenous electrical field, serving basically as ion mirror. The first segment of the reflectron builds a retarding, the second segment the reflecting field [292]. Ions of higher kinetic energy enter deeper into the

reflecting field than those of lower kinetic energy, thereby correcting for time differences, so that ions are detected simultaneously. TOF analyzers exhibit simple setups and provide robust and fast mass spectra [289]. Since the mass range theoretically is not limited [287, 289], (however finally still limited by the detectability of ions in secondary-ion-multipliers), they are suitable for the detection of masses up to 250000 m/z [293] including for example polymers or large biomolecules.

The linear quadrupole is one of the most widespread mass analyzers in analytical research. It is composed of four metal rods arranged in parallel, which serve to generate a two dimensional quadrupole field. After being accelerated by typically low (100V) voltage drops, ions enter the analyzer in axial direction [294]. Mass selection is achieved by applying a combination of rf- and dc-voltage on opposing rods, so that only certain m/z-ratios exhibit stable trajectories along the length of the quadrupole [294, 295]. Undesired ions are eliminated by colliding with the rods. Coupling of a quadrupole segment with additional ones enables the possibility of performing tandem MS, characterized by high ion transmission between the segments, hence high sensitivity [296]. Therefore, three quadrupoles are arranged in tandem, two of them serving as mass filters. Q1 selects precursor ions, whereas Q3 filters m/z-ratios of the preceding fragmentation reaction, Q2 thereby serving as collision chamber operated with inert collision gas [295–297]. The so-called SRM or MRM mode is often used for screening approaches searching for known analytes. It provides ultimate sensitivity, since in this case rf- and dc- voltages both in Q1 and Q3 are adjusted to the analyte and the transmission of one single (SRM) or multiple (MRM) fragment ions is optimized thereby [296].

Compared to other mass analyzers quadrupoles display a medium resolution and a limitation of the accessible mass range to about m/z 4000. The optimally hyperbolically shaped rods are often formed cylindrically for convenience [298].

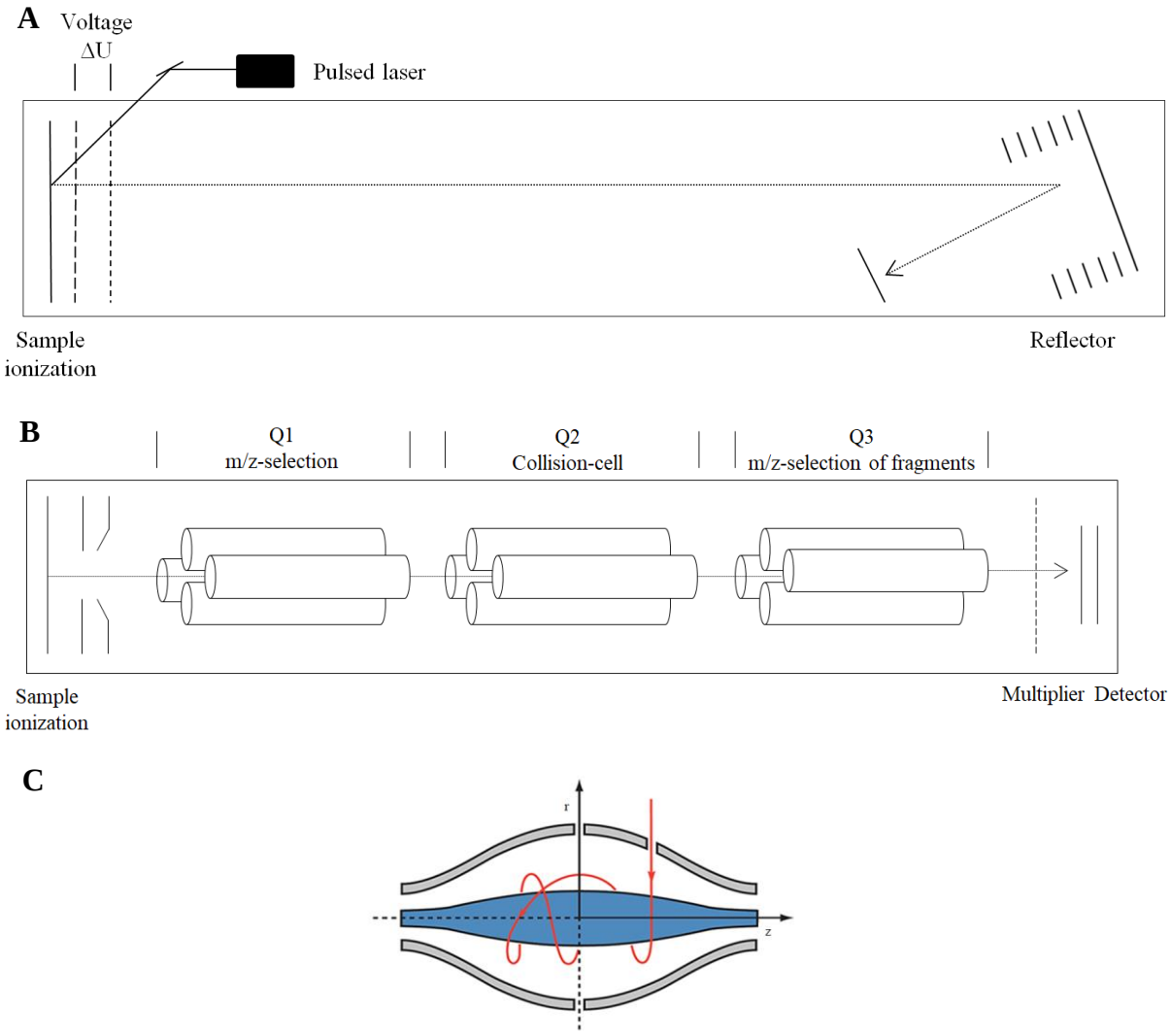


Figure 13: Schematic representation of common mass analyzers. (A) Reflector time-of-flight, (B)Triple quadrupole mass spectrometer and (C) Orbitrap [299].

The most recent type of mass analyzers is the Orbitrap, based on the early work of the development of the Kingdon-trap [300]. Its functionality rests on the oscillation of ions on stable trajectories surrounding a spindle-shaped electrode. Depending on the frequency with which ions are orbiting, their respective m/z-ratio can be determined. The spindle-shaped electrode is encircled by a barrel-shaped electrode composed of two halves, through which ions enter the analyzer. These are forced to rotate around the inner electrode, due to a defined present electrical field. Further oscillation takes place in axial field direction that can be transformed into a frequency spectrum by Fourier-transformation (FT) [301, 302]. The axial frequencies are inversely proportional to the square root of their m/z-ratios of the respective ions and independent of their energy or spatial spread, resulting in critical advantages compared to other analyzers, including high resolution, high mass accuracy and a wider mass

range [301, 303]. The coupling of the Orbitrap to different ionization techniques demands the use of hybrid devices to ensure directed ion transport, mostly carried out by linear quadrupoles or quadrupole-ion traps as preceding elements [304, 305]. After ions are generated, they are transferred to the linear quadrupole, which gives the possibility of recording MS^n spectra [302]. Ions are then further transferred to the C-trap, consisting of a C-shaped RF-quadrupole that accumulates the ions to eventually eject them in a pulsed manner into the Orbitrap. By entering the C-trap, ions loose energy due to mild collisions with collision gas without fragmenting. The C-trap ensures pooling of collisionally-cooled ions so that they enter the Orbitrap tangentially [304]. Furthermore, it uncouples the Orbitrap from preceding analyzers, enabling the coupling to virtually every analyzer and fragmentation technique available [306].

2. Aim of the study

Transcription of the human ALOX5 gene is tightly regulated by various factors that contribute to the differential expression pattern and activity of 5-lipoxygenase in the body, whereas the enzyme is found mostly in cells of the immune system. Since the 5-LO pathway is involved in the pathogenesis of distinct diseases, there is constant striving for finding novel intervening targets including inhibition of the enzyme itself, but also a potential modulation on genetic level. In this respect, several efforts have already been made to obtain a deep understanding of how 5-LO transcription is essentially controlled and to elucidate its regulatory mechanisms, such as the enzyme's regulation by DNA methylation and histone acetylation [49, 52], its inducible expression after TGFβ/1,25$(OH)_2D_3$ treatment in myeloid cell lines and its respective effects on transcript elongation and initiation [24, 57–60]. Additionally, a couple of transcription factors were identified to be responsible for basal promoter activity [45, 53]; however the results obtained so far arise from hypothesis-driven *in vitro* studies. The identification and characterization of further DNA-binding proteins that contribute to transcriptional regulation is still lacking, although the ALOX5 promoter features several consensus sequences for a variety of transcription factors.

Accordingly, the present study aimed at an unbiased overall identification of proximal ALOX5 promoter interacting proteins, which might give further insight into gene regulatory mechanisms of 5-LO expression. For this purpose, a DNA affinity enrichment coupled to label-free quantitative proteomics was chosen as the central tool, since it provides the required sensitivity and accuracy of identifying potential DNA interactors out of complex samples with high significance. Different myeloid and B-lymphocytic cell lines were used to detect possible cell specific distinctions in identified protein patterns. In this regard, cells of the myeloid lineage were utilized in undifferentiated and differentiated state, to elucidate differentiation-dependent changes in protein composition.

In a second part, the putative existence of secondary DNA structures in the proximal ALOX5 promoter sequence was to be determined, since proteins identified by quantitative proteomics revealed several proteins involved in the interaction with G4-DNA. As the 5-LO promoter exhibits a high GC-content of 80% and contains five GC-boxes arranged in tandem [34], it represents a potential target for developing G-quadruplex structures. However, up to now no formation of G4-DNA could be confirmed. According to that, different assays for its *in vitro* detection, as well as *in cellulo* treatments with G4-stabilizing agents were carried out in order to investigate its potential of forming these secondary DNA structures.

3. Materials and methods

3.1 Materials

3.1.1 Cell culture and cell lines

Dulbecco's phosphate-buffered saline (DPBS 1x or 10x)	Thermo Fisher Scientific, Braunschweig, Germany
L-glutamine	
Penicillin-Streptomycin (PS) 100x	
RPMI 1640 medium w/glut	
D-MEM medium	
non-essential amino acids	
human insulin	
non-essential amino acids	
sodium pyruvate	
trypan blue	
1,25-dihydroxyvitamin D3	Cayman Chemical Company, Ann Arbor, MI
TGFβ	PeproTech GmbH, Hamburg, Germany
fetal bovine serum (FBS)	Capricorn Scientific GmbH, Ebsdorfergrund, Germany
HL-60	Deutsche Sammlung von Mikroorganismen und Zellkulturen DSMZ, Braunschweig, Germany
THP-1	
MM6	
Rec-1	
BL-41	
U937	
Hek293T	

3.1.2 Oligonucleotides

120-mer WT, CTR	Sigma-Aldrich, Steinheim, Germany IBA Life Sciences GmbH, Göttingen, Germany
80-mer WT, CTR	Eurofins Scientific, Ebersberg, Germany
5x(GC)	
c-myc CTR	
CTR ELISA	
CTR spectroscopy	
primer for RT-qPCR	

3.1.3 Subcellular fractionation

HEPES >99.5% p.a.,	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
glycerol, >99.5% p.a., water-free	
EDTA x 2Na $2H_2O$	
Tris	
glycine	
$MgCl_2$ x $6H_2O$	Merck KGaA, Darmstadt, Germany
KCl, p.a. >99%	Fluka Chemie AG, Buchs, Switzerland

NaCl p.a. >99.5%	
cOmplete Mini protease inhibitor cocktail tablets, EDTA-free	Roche Diagnostics GmbH, Mannheim, Germany
Igepal CA-630 (NP-40 equivalent)	Sigma-Aldrich, Steinheim, Germany
Triton X-100	
Pierce™ BCA Protein Assay Kit	Thermo Scientific, Rockford, IL
Pierce™ Coomassie (Bradford) Protein Assay Kit	

3.1.4 Electrophoresis and immunoblotting

Gene Ruler 1kb, 100bp	Thermo Fisher Scientific, Braunschweig, Germany
6x orange loading dye	
ultrapure agarose	Invitrogen, Carlsbad, CA
ethidium bromide	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Rotiphorese® Gel 30 (37.5:1)	
butanole	
SDS, ultrapure, for electrophoresis	
Roti®-PAGE 10%	
ammonium peroxodisulfate (APS)	Merck KGaA, Darmstadt, Germany
TEMED	
Tween® 20, cell culture tested	Sigma-Aldrich, Steinheim, Germany
bromophenol blue	
InstantBlue Protein Stain	Biozol Diagnostica Vertrieb GmbH, Eching, Germany
Trans-Blot® Turbo Transfer Pack, format: mini, midi, 0.2µm PVDF membranes	Bio-Rad Laboratories GmbH, Rüdigheim, Germany
Precision Plus Protein All Blue Standards	
Odyssey Blocking buffer	Licor Biosciences, Bad Homburg, Germany

3.1.5 Antibodies

anti-Sp1-antibody (1C6), mouse, monoclonal	Santa Cruz Biotechnology, Heidelberg, Germany
anti-RXRα-antibody (D-20), rabbit, polyclonal	
anti-VDR-antibody (C-20), rabbit, polyclonal	
anti-β-actin-antibody (C4), mouse, monoclonal	
anti-Sp1-antibody (PA5-29165), rabbit, polyclonal	Invitrogen, Carlsbad, CA
anti-Egr-1-antibody (MA5-15009), rabbit, monoclonal	Thermo Scientific, Rockford, IL
anti-WT1-antibody (AF5729), goat, polyclonal	R&D Systems, Wiesbaden, Germany
anti-AP2α-antibody, rabbit, polyclonal	Cell Signaling Technology, Frankfurt, Germany
anti-RXRα-antibody, rabbit, polyclonal	
anti-Histone H3-antibody (9715), rabbit, polyclonal	
anti-5-lipoxygenase-antibody, rabbit, monoclonal	
anti-5-lipoxygenase-antibody, clone 6A12, mouse, monoclonal	in-house-made, provided by working group Prof. Dr. Steinhilber
anti-DNA G-quadruplex structures, clone BG4, monoclonal	Merck KGaA, Darmstadt, Germany

anti-DDDDK tag antibody (HRP) (ab1238), goat, polyclonal	Abcam, Cambridge, UK
secondary antibodies donkey-anti-rabbit/goat/mouse IRDye® 680/800	Licor Biosciences, Bad Homburg, Germany

3.1.6 Affinity purification and MS sample preparation

Lightshift poly-dIdC	Thermo Fisher Scientific, Braunschweig, Germany
dithiothreitol (DTT)	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Tris-HCl	
trifluoroacetic acid (TFA), >99.9%, for peptide synthesis	
dithioerythritol (DTE)	Merck KGaA, Darmstadt, Germany
SeraMag™ SpeedBeads Blocked Streptavidin Particles	GE Lifesciences, Freiburg, Germany
ammonium bicarbonate	Sigma-Aldrich, Steinheim, Germany
iodoacetamide	
trypsin from porcine pancreas, proteomics grade	
triethylammonium bicarbonate (TEAB)	
formaldehyde solution, for molecular biology 36.5%-38% in H_2O	
formaldehyde-d_2, ~20 wt. % in D_2O, 98 atom % D	
deuterium oxide, 99.9 atom % D	
ammonia solution, puriss. p. a. plus	
sodium cyanoborohydride	
Pierce™ C-18 spin columns	Thermo Scientific, Rockford, IL
herring sperm DNA	Invitrogen, Carlsbad, CA
α-cyano-4-hydroxycinnamic acid	Bruker Daltonic GmbH, Bremen, Germany

3.1.7 ELISA

Streptavidin high binding capacity coated 96-well plates	Thermo Scientific, Rockford, IL
TMB substrate solution	
bovine serum albumin (BSA)	Sigma-Aldrich, Steinheim, Germany
potassium phosphate, dibasic	
sulfuric acid 95%	VWR Chemicals, Darmstadt, Germany

3.1.8 RNA extraction

DNAse I, RNAse free	Thermo Fisher Scientific, Braunschweig, Germany
high capacity RNA-to-c DNA kit	
powerUp SYBR green master mix	
TRIzol reagent	
ethanol, absolute, analytical reagent grade	
water, BioScience grade, nuclease-free, DEPC-treated	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
chloroform	Sigma-Aldrich, Steinheim, Germany
sodium acetate	
isopropanol	

3.1.9 Other reagents and devices

TMPyP4	Merck KGaA, Darmstadt, Germany
pyridostatin (PDS)	Cayman Chemical Company, Ann Arbor, MI
recombinant human Wilms Tumor protein (ab82233) (+KTS)	Abcam, Cambridge, UK
recombinant human Sp1 GC-box binding protein (SRP2030)	Sigma-Aldrich, Steinheim, Germany
acetonitrile for HPLC	VWR Chemicals, Darmstadt, Germany
acetone	
methanol, gradient grade for HPLC	
Tecan infinite m200®	Tecan Group, Männedorf, Switzerland
Tecan Spark	
Q Exactive Orbitrap™	Thermo Fisher Scientific
Trans-Blot Turbo® Transfer System	Bio-Rad Laboratories GmbH, Rüdigheim, Germany
TC20™ Automated Cell Counter	
Odyssey Infrared Imaging Systems	Licor Biosciences, Bad Homburg, Germany
Nanodrop 2000	Thermo Fisher Scientific, Braunschweig, Germany
StepOne™	Applied Biosystems, Darmstadt, Germany
Hamilton Microlab® STAR ML System	Hamilton Germany GmbH, Höchst, Germany
4800 MALDI TOF/TOF mass analyzer	AB Sciex Germany GmbH, Darmstadt, Germany
Jasco-810 spectrophotometer	Jasco, Pfungstadt, Germany

3.2 Methods

3.2.1 Cell culture

3.2.1.1 Growth and treatment conditions

The monocytic cell lines HL-60, THP-1, U937 and the B-lymphocytic cell lines Rec-1, BL-41 were cultured in RPMI 1640 medium, supplemented with 10% FBS, 1% PS and 2 mM glutamine at 5% CO_2, 37 °C. Monocytic cell lines were split every 3 to 4 days and seeded out at 0.35×10^6/ml, B-lymphocytes every 2 to 3 days at 0.5×10^6/ml, respectively. MM6 were cultured in RPMI 1640 medium supplemented with 10% FBS, 1% PS, non-essential amino acids, 10 µg/mL insulin, 1 mM oxaloacetate and 1 mM sodium pyruvate under the same conditions as for the other monocytes. Hek293T cells were cultured in D-MEM supplemented with 10% FBS and split every 3 to 4 days at 0.5×10^6/ml.

Differentiation of monocytic cell lines HL-60, THP-1 and MM6 was induced by adding 50 nM 1α,25-dihydroxyvitamin D3 and 1ng/mL TGFβ at a cell density of 0.3×10^6/ml. Afterwards cells were kept at 6% CO_2, 37 °C for three days. For comparative pulldown experiments, undifferentiated cells were seeded at 0.4×10^6/mL and grown at 5% CO_2, 37 °C for three days. Cells were harvested afterwards to proceed with subcellular fractionation.

For blotting experiments of whole cell lysates, all mentioned cell lines were likewise harvested after three days. Therefore, cells were centrifuged at 1200 × g for 5 minutes, washed once with PBS and again centrifuged at 1200 × g for 5 minutes. Excess PBS was removed to leave a dry cell pellet.

For treatment of HL-60 myelocytes with PDS and TMPyP4 cells were simultaneously differentiated with 50 nM 1α,25-dihydroxyvitamin D3 and 1ng/mL TGFβ at a cell density of 0.2×10^6/ml, incubated with either 10 µM PDS or 50 µM TMPyP4 at 6% CO_2, 37 °C and harvested after 24h or 48h as described.

3.2.1.2 Determination of cell viability and cytotoxicity

Cell viability and cytotoxicity was determined using trypan blue stain. 20 µl of cell suspension was mixed with 20 µl trypan blue solution. For routine determination for subsequent splitting of the cell culture, cells were counted automatically by using a cell counter. For splitting myeloid cell lines for consecutive differentiation, as well as the determination of cytotoxicity after treatment with PDS and TMPyP4, cells were counted manually with the aid of a light microscope.

3.2.1.3 Subcellular fractionation

Subcellular fractionation to obtain the nuclear extract was performed as described in Kloet *et al.* [307]. Briefly, harvested cell pellets were resuspended in 5 volumes of a buffer containing 10 mM HEPES-KOH, pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl and left to swell for 10 minutes. After centrifugation at 400 × g for 5 minutes, cells were resuspended in 2 volumes of the same buffer, supplemented with protease inhibitors and 0.15% NP-40. They were transferred to a Dounce homogenizer to lyse the outer membrane with 40 strokes of pestle B. Lysates were centrifuged at 3200 × g for 15 minutes and cytosolic fraction was removed. Nuclear pellet was washed once with 500µl cold PBS and centrifuged again for 5 minutes at 3200 × g. Nuclei were resuspended in 300-500 µl of a hypotonic buffer consisting of 20 mM HEPES-KOH, pH 7.9, 420 mM NaCl, 20% v/v glycerol, 2 mM $MgCl_2$, 0.2 mM EDTA, protease inhibitors, 0.1% NP-40 and 0.5 mM DTT. These were incubated for one hour with rotation to be centrifuged again at 18000 × g for 15 minutes. The supernatant, which contained nuclear protein was either stored at -80 °C or used directly. All steps were performed at 4 °C.

3.2.1.4 Preparation of whole cell lysates

For immunoblotting experiments with whole cell lysates 15×10^6 cells were harvested and resuspended in100 µl of a buffer consisting of 50 mM Tris-HCl, pH 7.5, 0.5% v/v triton x100, 200 mM NaCl, 10% v/v glycerol, 1 mM DTT and protease inhibitors. Suspension was frozen at -80 °C and thawed on ice, repeating the steps two additional times. After thawing, cells were centrifuged at 10000 × g for 10 minutes, 4 °C and supernatant, which contained the soluble protein fraction, was stored at -80 °C for further experiments.

3.2.1.5 BCA assay

To determine protein concentrations of whole cell lysates a BCA assay was performed according to the manufacturer's protocol (Thermo Fisher Scientific). 10 µl of each BSA standard or unknown sample replicate was pipetted into a microplate in triplicate and incubated with 200 µl BCA working reagent for 30 minutes at 37 °C. After cooling of the plate, absorbance was read at 562 nm with the Tecan infinite m200® plate reader and concentration of protein samples was determined via calibration curve.

3.2.1.6 Bradford assay

For the determination of protein concentrations of cytosolic and nuclear fraction after cellular fractionation, Bradford assay was performed according to the manufacturer's protocol (Thermo Fisher Scientific). 5 µl of each BSA standard or unknown sample or a dilution thereof was incubated with 250 µl Coomassie reagent in a microplate. The plate was incubated at room temperature for 10 minutes and absorbance was read at 595 nm with the Tecan infinite m200® plate reader. Protein concentration was determined via calibration curve after correction for respective dilution factors.

3.2.2 Oligonucleotides

Single-stranded oligonucleotides were purchased in their respective forward and reverse strands for all experiments. All forward strands were biotinylated at the 5'-end. Same amounts of forward and reverse strand were mixed and diluted 1:10 to yield a concentration of 10 pmol/µl with annealing buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 1 mM EDTA). Annealing was performed by heating the mixture to 95 °C for 5 minutes and letting them cool down gradually to room temperature overnight. Successful annealing to duplex DNA was verified by 3% agarose gel electrophoresis.

For quadruplex experiments single-stranded DNA sequences were annealed in a buffer containing 10 mM Tris-HCl, pH 7.4 and 100 mM KCl under same conditions.

3.2.3 Electrophoresis and immunoblotting

3.2.3.1 Agarose gel electrophoresis

3% agarose gel electrophoresis was carried out to verify annealing of oligonucleotides. Gels were cast with 60 ml of agarose/ethidium bromide mixture and run at 75 V for 1 hour. Gels were afterwards visualized by irradiation with UV-light.

3.2.3.2 SDS-PAGE and Coomassie staining

SDS-PAGE was performed with either pre- or self-cast 10% Laemmli-gels in Bio-Rad electrophoresis chambers suitable for mini gels. Self-cast gels consisted of both separating (6.2 ml H_2O, 4.9 ml polyacrylamide solution Rotiphorese® Gel 30, 3.75 ml separation buffer 4×, 150 µl APS 10%, 15 µl TEMED) and stacking (6.1 ml H_2O, 1.3 ml polyacrylamide solution Rotiphorese® Gel 30, 2.5 ml stacking buffer 4×, 100 µl APS 10%, 10 µl TEMED) gel. Separation buffer (4×) contained 1.5 M Tris-HCl, pH 8.8 and 14 mM SDS, stacking buffer (4×) 0.5 M Tris-HCl, pH 6.8 and 14 mM SDS. Separation gel was cast and overlain

with butanol to remove any bubbles and left to polymerize prior to casting the stacking gel. Samples were mixed with Laemmli-buffer (2× 65.8 mM Tris-HCl, pH 6.8, 26.3% glycerol v/v, 2.1% SDS, 0.01% bromophenol blue) to which 5 mg/ml DTT was added prior to electrophoresis. Sample-buffer mixture was incubated for 5 minutes at 95 °C and loaded onto the gels to perform electrophoresis at a constant 75 V for 75 minutes. Proteins were visualized on Odyssey® Infrared Imaging System after staining with colloidal Coomassie stain InstantBlue™ overnight.

3.2.3.3 Immunoblotting

After SDS-PAGE, gels were blotted semi-dryly by using the TransBlot Turbo® and corresponding membrane sizes in mini or midi format. Therefore, pre-run gels were transferred onto PVDF membranes, which were enveloped with buffer-soaked filter papers. Proteins were blotted by using default blotting protocols for one or two gels. Membranes were than washed in PBS for 10 minutes with slight shaking, blocked with Odyssey® Blocking buffer for 1 hour at room temperature and washed again three times for 5 minutes to remove excess blocking buffer. Blots were incubated with 10 ml primary antibody solution consisting of equal amounts of blocking buffer and PBS, supplemented with 0.1% Tween 20 and the diluted primary antibody overnight at 4 °C. After removing the blots from the antibody solution, they were washed four times for 5 minutes with PBS, supplemented with 0.1% Tween 20 and another 5 minutes with PBS, to remove residual detergent. Incubation of membranes in 5 ml secondary antibody solution (1:1 blocking buffer: PBS, 0.1% Tween 20, 0.02% SDS, secondary antibody 1:10000) took place at room temperature for 45 minutes in the dark. Following incubation, blots were washed as before for five times with PBS+0.1% Tween 20/PBS. Visualization of proteins was performed using the Odyssey® Infrared Imaging System, quantification if needed by densitometry, with β-actin serving as loading control.

3.2.4 Mass spectrometry and data analysis

3.2.4.1 In-gel tryptic digestion

In-gel digestion was performed according to optimized and customized digestion protocols for the Hamilton Microlab® STAR ML System. After protein separation by SDS-PAGE and visualization of bands by Coomassie staining, gel slices of 1 mm thickness were excised from the lanes, which were further cut into 1mm cubes for the respective band and transferred into a 96-well plate. Gel pieces were washed three times with water and three times with 1:1 ACN:ambic (25 mM) and dehydrated twice with 100% ACN. They were further reduced by

adding 10 mM DTT and alkylated by adding 55 mM IAA. Another washing step with twice ammonium bicarbonate was performed. Gel pieces were dehydrated with 100% ACN, reswelled with ambic and again dehydrated with 100% ACN twice. Then, trypsin was added and proteins were digested overnight. Reaction was stopped by adding 0.3% TFA. Peptides were extracted by an interchanging combination of 100% ACN and 1:1 ACN:TFA 0.3%, transferred to another 96-well plate and dried by vacuum centrifugation.

3.2.4.2 MALDI-MS/MS

All in-gel digested proteins were subjected to MALDI-TOF-MS/MS analysis. Vacuum-centrifugation dried peptides were reconstituted by adding 5 µl of 1:1 ACN:TFA 0.1%. 0.5 µl thereof were mixed with 0.5 µl 4-HCCA matrix, spotted onto a sample target and measured in positive ion reflector mode. Further settings were set as follows: mass range 800-4000 m/z, 1000 total shots/spectrum for MS, 1800 for MS/MS for acquisition, fragmentation of 15 precursors at maximum, beginning with the strongest precursor first and a minimum S/N (signal-to-noise ratio) of 20 for their respective selection. Peaks were extracted at a peak density of a maximum of 5 peaks per 200 Da with a minimum S/N of 20 for both MS and MS/MS. Raw data was processed by Mascot MS/MS ion search against human SwissProt database. Trypsin was set as protease allowing 1 missed cleavage. Carbamidomethyl (C) was set as fixed, oxidation (M) and deamination (NQ) as variable modification. Search was performed with 70 ppm peptide tolerance and 0.4 Da MS/MS tolerance.

3.2.4.3 DNA pulldown

DNA pulldowns were carried out according to Hubner *et al.* [183], with modifications. After quantification of protein content of nuclear cell extracts, 300-400 µg thereof were incubated with 10 pmol of either WT or CTR ds-DNA together with 200-400 µl protein binding buffer (50 mM HEPES-KOH, pH 8.0, 150 mM NaCl, 0.1% NP-40, 1 mM DTT, protease inhibitors) and 10 µg poly dI-dC as competitor for nonspecific DNA-binding proteins. After 1.5 hours, 100 µl SeraMag^{TM} SpeedBeads Blocked Streptavidin Particles were added to capture formed DNA-protein complexes for another 0.5 hour under shaking (1000 rpm). Prior to incubation, magnetic beads were washed with 100 µl conditioning buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 2 M NaCl) and 100 µl binding buffer. Supernatant was removed and discarded. Enriched DNA-binding proteins were washed three times with equal volumes of protein binding buffer and three times with wash buffer (100 mM ambic buffer, 150 mM NaCl) to dispose of any remaining detergent. After the very first washing step, the

Eppendorf tube was changed, to discard of proteins that might have absorbed to the tube walls. All of the aforementioned steps were carried out at 4 °C.

oligonucleotide	sequence
120-mer WT fwd (-260 to -141)	*5'-[BIO] ACC CAG GAG CGA GGC CCC TGC CCC GCC CGA GGC GAG GTC CCG CCC AGT CGG CGC CGC GTG AAG AGT GGG AGA GAA GTA CTG CGG GGG CGG GGG CGG GGG CGG GGG CGG GGG CAG*
120-mer WT rev	*CTG CCC CCG CCC CCG CCC CCG CCC CCG CCC CCG CCC CCG CAG TAC TTC TCT CCC ACT CTT CAG GCG GCG CCG ACT GGG CGG GAC CTC GCC TCG GGC GGG GCA GGG GCC TCG CTC CTG GGT*
120-mer SCR fwd	*5'-[BIO]ACG CGT CCG AGG CCT GCG AGC CGG CCG TGC AGC GCT GCC GAG GGG TGC GAG CCC ACG CGG ACG CGC GAG GCC CGC ACC CTG CGA GGG TCG GAC GGA GCG ACG GTC CCC CAG CCT CGC GAG*
120-mer SCR rev	*CTC GCG AGG CTG GGG GAC CGT CGC TCC GTC CGA CCC TCG CAG GGT GCG GGC CTC GCG CGT CCG CGT GGG CTC GCA CCC CTC GGC AGC GCT GCA CGG CCG GCT CGC AGG CCT CGG ACG CGT*
80-mer WT-SMAD fwd (-334 to -255)	*5'[BIO]CGA ATG GAT GAG GGG TGG CAG CCG AGG TTG CCC CAG TCC CCT GGC TGC AGG AAC AGA CAC CTC GCT GAG GAG AGA CCC AG*
80-mer WT-SMAD rev	*CTG GGT CTC TCC TCA GCG AGG TGT CTG TTC CTG CAG CCA GGG GAC TGG GGC AAC CTC GGC TGC CAC CCC TCA TCC ATT CG*
80-mer SCR-SMAD fwd	*5'[BIO]ACG CCT GCG ATG CGA GGG TCA TCC CAC ATA ACG GCG CTA GAG GCC CTG CGA ACG CAC GTA GGG TCG AAG ATG GCC GAG GG*
80-mer SCR-SMAD rev	*CCC TCG GCC ATC TTC GAC CCT ACG TGC GTT CGC AGG GCC TCT AGC GCC GTT ATG TGG GAT GAC CCT CGC ATC GCA GGC GT*
80-mer WT-VDR fwd (-408 to -329)	*5'[BIO]AGC CTG GCC TTG GGC GAA GCC GAG GCA GGC AGG CAG GGC AAA GGG TGG AAG CAA TTC AGG AGA GAA CGA GTG AAC GAA TG*
80-mer WT-VDR rev	*CAT TCG TTC ACT CGT TCT CTC CTG AAT TGC TTC CAC CCT TTG CCC TGC CTG CCT GCC TCG GCT TCG CCC AAG GCC AGG CT*
80-mer SCR-VDR fwd	*5'[BIO]GCG GAC CGA TGG AAG CTA GGG ATC GTA AGC AGT CAG AAG GCA GGA TGA ACG ATA AGG CTG GAC GGA GCC ACG GAG GCA GG*
80-mer SCR-VDR rev	*CCT GCC TCC GTG GCT CCG TCC AGC CTT AGC GTT CAT CCT GCC TTC TGA CTG CTT ACG ATC CCT AGC TTC CAT CGG TCC GC*

Table 4: Oligonucleotide sequences used for mass spectrometry based DNA pulldown experiments with their respective length, 5'-modification and reverse strand. Genomic locations are given in relation to ATG.

For subsequent immunoblotting experiments, beads were resuspended in 20 µl Laemmli buffer and directly subjected to SDS-PAGE. For MS-based identification of protein complexes, beads were resuspended in 100 µl 25 mM ambic and 9 mM DTE was added to reduce disulfide bridges at 57 °C for 45 minutes, following an alkylation step by adding 17 mM IAA for 45 minutes at room temperature in the dark. Reaction was quenched by another addition of DTE for 15 minutes. Protein complexes were then directly digested on the beads overnight, by adding 0.1 µg trypsin per sample at 37 °C. The next day, digestion was completed by further adding 0.1 µg of trypsin per sample for 1 hour. Reaction was stopped by adding 1 µl of TFA to each sample.

For titration experiments various amounts of both nuclear extract (50 µg, 100 µg, 250 µg, 500 µg) and oligonucleotide (10 pmol, 50 pmol, 100 pmol, 250 pmol) were used in combination with one another. Titration experiments were performed with the monocytic cell line MM6 and subjected to label-free MS-quantification. Experiments with excess competitor were performed with the myeloid cell line HL-60 as described above, except for the use of DNA-competitor, which was changed to a 40 fold excess of used oligonucleotide of herring sperm DNA. These samples were likewise quantified by label-free MS-quantification. In general, all label-free MS-quantification experiments were performed in triplicate, whereas dimethyl labeling experiments were performed in duplicate.

3.2.4.4 Dimethyl labeling

Dimethyl labeling of peptides was carried out as described in Boersema *et al.* [238]. After washing the beads containing DNA-protein complexes, they were resuspended in 100 µl 50 mM TEAB buffer for ensuring compatibility for the following labeling reaction conditions. Samples were likewise reduced, alkylated and digested. After completion of digestion, samples were either labeled with light or heavy label for WT and CTR sample, respectively, with swatch of labels to avoid bias in labeling reaction. 4 µl of 4% v/v CH_2O or CD_2O was added to the according samples, mixed and further incubated with 4 µl of 0.6 M $NaBH_3CN$ at room temperature for 1 hour with slight shaking. Reactions were quenched by adding 16 µl of 1% v/v ammonia solution. To even further quench the reaction and provide suitable conditions for subsequent LC-MS/MS measurement, 8 µl of formic acid was added to each sample. The correlating light and heavy labeled samples were mixed, purified and dried by vacuum centrifugation for proceeding MS analysis.

3.2.4.5 Purification of peptides

After completion of digestion for both dimethyl labeled and label-free samples, peptides were purified with C18 spin columns according to the manufacturer's protocol. Each spin column was conditioned by centrifuging twice with 50% ACN and 5% ACN + 0.5% TFA. Samples were then bound to the C18 resin three times and washed afterwards again with 5% ACN + 0.5% TFA two times. Peptides were eluted by adding 20 µl of 70% ACN twice. Purified samples were dried by vacuum centrifugation for proceeding MS analysis.

3.2.4.6 LC-MS/MS

LC-MS/MS measurements were carried out with an EASY nLC 1000 coupled to a Q Exactive Plus mass spectrometer (both Thermo Fisher Scientific). Dried samples were reconstituted in buffer A (0.1% formic acid in water) and then loaded onto a 50 cm Poroshell EC120 C18 column. The separation gradient was set from 3 to 95% of buffer B (0.1% formic acid in ACN) in buffer A over 150 minutes at a flow rate of 250 nl/min. MS1 scans were recorded in positive ion mode at a resolution of 70000, the following highest 10 MS/MS spectra at a respective resolution of 17500 (AGC 5×10^5, max. injection time 55 ms, isolation window 1.8 Da).

3.2.4.7 Data analysis

All RAW data files were analyzed by open source software MaxQuant, version 1.6.0.1 and in-built search engine Andromeda [260]. All further DNA-pulldown data analysis sets were carried out using Perseus software, version 1.6.0.0 according to Hubner *et al.* [183] with minor modifications, as described in the following.

RAW data were grouped according to both cell conditions to be compared with (WT and CTR) of the respective cell lines. MaxQuant was operated with default settings. For label-based pull downs the following labels were enabled: DimethLys0, DimethLys4, DimethNter0, DimethNter4. For label-free pulldowns, match between runs was enabled, to align retention times. Oxidation (M), acetylation (protein N-term) and deamidation (NQ) were set as variable, carbamidomethyl as fixed modification. Tryptic digestion was set, with the option of 2 missed cleavages at maximum. For label-free quantification the iBAQ and LFQ option was enabled with a minimum ratio count of 1. The used FASTA-file contained all human UniProt entries and was downloaded on 19.06.2017. Search of peptides was performed with a tolerance of 4.5 ppm in main search and 20 ppm for fragment ions. PSM and protein FDR were both set to 0.01.

Protein tables achieved from the database search were afterwards filtered for contaminants, reverse hits and proteins only identified by site. Proteins were accepted as identified with at least 1 unique peptide. LFQ intensities and normalized H/L ratios were $\log_2$ transformed. Label-free quantification samples were grouped according to their WT or CTR affiliation and filtered for at least three valid values. Missing data imputation was done with default settings (width 0.3, downshift 1.8). Suitable distribution of imputed values was checked with histograms. Afterwards, a two sample t-test was applied and significant outliers were identified using a volcano plot, where threshold values of significance were adjusted to each cell line and cell state separately. Label-based quantification data imputation was done with an adjusted downshift of standard deviation to 0.5. Outliers were determined based on protein intensity (significance B), which was required to be 0.05 for forward and reverse experiments and visualized via scatter plot.

3.2.5 UV-VIS-spectroscopy

UV-VIS-spectra were recorded to give evidence for the existence of G4-DNA in oligonucleotides used for pulldowns. In order to do so, TMPyP4 was reconstituted to a stock solution of 50 µM prior to performing the experiments and kept in the dark. A serial dilution (0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.6, 0.8, 1, 2, 3, 4, 6, 8, 10 µM) of each already annealed oligonucleotide to be tested was prepared and incubated with a final concentration of TMPyP4 of 5 µM in a final volume of 100 µl for 5 minutes. Concentrations were successively increased until three more additions of oligonucleotide did not alter the shift of the Soret band any further. All UV-VIS spectra were recorded on the Tecan Spark in a range of 350-500 nm wavelength. Spectra were averaged out of three measurements after being corrected for their solvent effects. Oligonucleotide sequences used for the experiments were as follows:

WT sample: [BIO]-GATCCTGCGG(GGGCGG)$_5$CAG, positive control: c-myc: [BIO]-TGAGGGTGGGTAGGGTGGGTAA, negative control: GCTCGCCCCGCCCCGATCGAAT.

The percentage of hypochromicity of each oligonucleotide was calculated according to Wei *et al.* [144]. Percentage of hypochromicity there is set as % hypochromicity = $((\varepsilon_{free} - \varepsilon_{bound})/\varepsilon_{free}) \times 100$. ε_{free} was calculated based on Lambert-Beer's law with $\varepsilon_{free} = A_{free}/C$, C thereby corresponding to the free concentration of TMPyP4 being 5 µM. Absorbance was recorded at the Soret maximum of TMPyP4 (422 nm). ε_{bound} then was determined by using the equation $\varepsilon_{bound} = A_{bound}/C_{bound}$, with A_{bound} being the absorbance of fully bound TMPyP4 at the Soret maximum and C_{bound} being the concentration of bound TMPyP4, respectively, and thereby underlying equations of $C_{bound} = C - C_{free}$ and $C_{free} = C(1-\alpha)$. Here, α stands for the fraction of

bound TMPyP4 and was calculated with $\alpha = (A_{free} - A)/(A_{free} - A_{bound})$, A_{free} and A_{bound} representing the absorbance of free or fully bound TMPyP4 at the Soret maximum and A being the absorbance at any given time point.

3.2.6 CD spectroscopy

CD spectra were recorded at room temperature (25 °C) on a Jasco-810 spectrophotometer with the following settings: 1 mm optical path length quartz cell, instrument scanning speed 50 nm/min, response time 1s, range 200-350 nm, 200µl sample volume. Spectra shown are average spectra of five measurements after correction for solvent effects.

3.2.7 ELISA

For the determination of the *in vitro* presence of G-quadruplex structures indirect ELISA was carried out, according to Lam *et al.* [308] with minor modifications. Oligonucleotide sequences used were as follows: *WT sample: [BIO]-GATCCTGCGG(GGGCGG)$_5$CAG, positive control: c-myc: [BIO]-TGAGGGTGGGTAGGGTGGGTAA, negative control: [BIO]-GATCCTG(CGAATT)$_5$CAATTCAG.* After annealing of 5'-biotinylated single-stranded oligonucleotides to form possible secondary DNA structures, the DNA-sequences were bound to streptavidin coated 96-well plates. Prior to the incubation, the latter were conditioned three times with 200 µl of a washing buffer (25 mM Tris-HCl, pH 7.2, 150 mM NaCl, 0.1% BSA, 0.1% Tween 20) for 5 minutes each turn. 50 nM oligonucleotides were then bound to the conditioned plates for one hour at room temperature and washed again three times afterwards with ELISA buffer (50 mM K2HPO4, pH 7.4, 100 mM KCl). Primary anti-DNA G-quadruplex structures antibody was added in altering concentrations (0, 0.005, 0.01, 0.05, 0.1, 0.25 µg/ml) and left to bind for another hour at room temperature. Supernatant was removed and discarded and plates were washed three times with ELISA buffer, supplemented with 0.1% Tween 20, following incubation with secondary antibody (HRP-anti-DDDDK-tag) at a dilution of 1:100000 for another hour. After three additional washes with ELISA buffer containing 0.1% Tween 20, 100 µl TMB substrate solution was added to each well. Reaction was stopped by adding 100 µl 0.18 M H_2SO_4. Absorbance was read at 450 nm with the Tecan infinite m200®.

3.2.8 RT-qPCR

3.2.8.1 RNA extraction

RNA-extraction of HL-60 cells was performed according to an optimized RNA-extraction protocol for HL-60, kindly provided by members of the working group of Prof. Dr. Steinhilber as described in the following. All steps after harvest were carried out on ice at 4 °C. 10×10^6 cells were harvested by centrifugation at 1200 rpm for 5 minutes. Supernatant was removed and cell pellet was resuspended in 1 ml Trizol without further washing. After a short incubation of about 2 minutes, suspension was transferred into an Eppendorf tube and incubated on ice for 10 minutes. 200 µl of chloroform were added and mixture was vortexed until homogeneity was achieved. After another 10 minutes of incubation on ice, samples were centrifuged at 12000 × g for 30 minutes. Three different phases were to be seen after removing the samples from the centrifuge, the lower red one and the white middle one thereby containing DNA and proteins, respectively, the upper clear phase containing RNA. The latter was transferred to a new tube. To increase RNA yield and integrity, the addition of Trizol (300 µl) was repeated with direct incubation with 200 µl chloroform for 2 minutes at room temperature. This was followed by centrifugation at 12000 × g for 30 minutes. The upper phase again was transferred to a new tube and RNA was precipitated with 500 µl ice cold isopropanol, supplemented with 5 µl 3 M sodium acetate, pH 6.5. Samples were incubated on ice for 10 minutes and subsequently frozen at -20 °C overnight. The following day, RNA samples were centrifuged at 12000 × g at 4 °C. Supernatant was removed and discarded to leave behind a white RNA pellet, formed at the bottom of the tube. Pellet was washed with 1 ml 75% ethanol, vortexed and again centrifuged at 12000 × g. Supernatant was removed and pellet was left to air-dry completely. Afterwards it was resuspended in 32 µl RNAse-free DEPC-treated water and heated at 55 °C for 10 minutes to be brought into solution entirely.

3.2.8.2 DNAse digestion

RNA samples were incubated with both DNAse-I (2 µl per 32 µl RNA) and DNAse buffer (4 µl) with slight shaking at 37 °C. After 30 minutes, 2 µl of EDTA were added and incubated another 10 minutes at 65 °C. Precipitation of RNA was achieved by adding 4 µl sodium acetate and 200 µl ethanol (100%) and incubation at -20 °C overnight. Samples were centrifuged as before and supernatant was removed. Precipitate was resolved in 25 µl DEPC-treated water and concentration of RNA samples was measured on the Nanodrop 2000. RNA

integrity was monitored by ethidium bromide agarose gel electrophoresis (see Materials and Methods).

3.2.8.3 cDNA-synthesis

Kit (High-Capacity RNA-to-cDNA™ Kit) reagents were slowly thawed on ice and cDNA synthesis was performed according to the manufacturer's protocol. 2 µg of RNA was used per 20 µl reaction. 10 µl (2×) RT-buffer was mixed with 1 µl (20×) enzyme mix, the corresponding volume containing 2 µg RNA and the remaining volume of water to generate 20 µl maximum reaction volume. The reaction mixture was incubated at 37 °C for 60 minutes and stopped by heating up to 95 °C for 5 minutes. cDNA was stored at -20 °C until qPCR was performed.

3.2.8.4 qPCR

For qPCR experiments, quantification of the respective gene of interest was normalized to GAPDH as control gene. All experiments were carried out on StepOne PCR Systems in technical duplicates and biological triplicates. Gene expression was ascertained by calculating the $2^{-\Delta\Delta Ct}$ values afterwards. The following primer sequences were used: 5-LO primer *fwd – GAA TTA CTC CAA AGC GAT GG* and *rev – ATG ACC CGC TCA GAA ATA GTG* [309], GAPDH primer *fwd – TGC ACC ACC AAC TGC TTA GC* and *rev – GGC ATG GAC TGT GGT CAT GAG* [309]. A no template control was included for testing of contamination. Determination of mRNA level was performed by detection of SYBR Green fluorophore. For each approach 2 µl cDNA were mixed with 0.06 µl forward primer, 0.06 µl reverse primer 7.88 µl H_2O and 10 µl SYBR Green PCR Mastermix. The PCR program was set as follows: UDG activation 2 minutes, 50 °C, Dual-LOCK DNA polymerase 2 minutes, 95 °C, denature, 3 seconds, 95 °C, anneal/extend 30 seconds, 60 °C (40 cycles).

4. Results

4.1 Transcription factors in ALOX5 regulation

4.1.1 Choice of myeloid and B-lymphocytic cell lines

Since the 5-lipoxygenase expression is cell type specific, suitable cell lines for all of the performed promoter screening experiments were selected for their respective level of enzyme expression. As already described in section 1.1.3, 5-LO is mainly expressed in cells of myeloid origin, as well as B-lymphocytes to a lower extent. Due to the inducible protein expression after cell differentiation of the myeloid cell lineage, both undifferentiated and differentiated cell states had to be considered. Therefore, 5-LO expression in different cell lines of both lineages was verified by Western blot analysis in order to confirm the validity of the cell system. The cell lines that exhibited detectable levels of protein were used for further experiments, as were the 5-LO-negative control cell lines U937 and Hek293T.

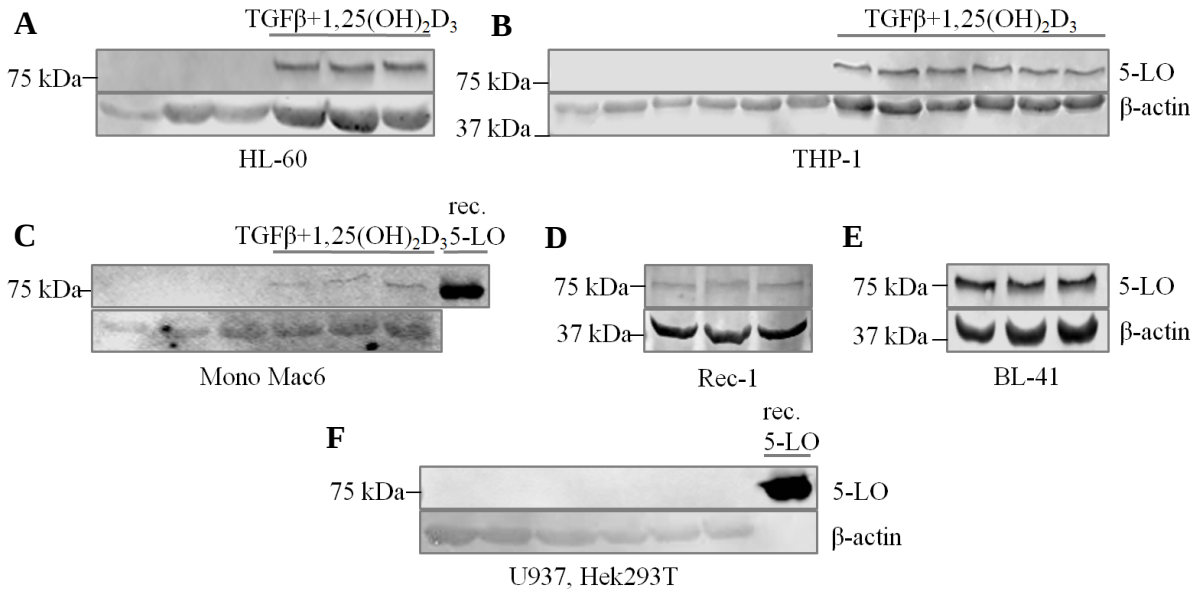


Figure 14: Immunoblot-based verification of 5-LO expression pattern. 10-20µl whole cell lysate, each containing 30µg total protein were tested for the myeloid cell lines (A) HL-60 in undifferentiated and differentiated, (B) THP-1 in undifferentiated and differentiated, (C) MM6 in undifferentiated and differentiated state. Differentiation was induced by adding 1 ng/ml TGFβ and 50 nM 1,25(OH)$_2$D$_3$ for 48h at 37 °C, 6% CO$_2$. Further cell lines comprised B-lymphocytes (D) Rec-1 and (E) BL-41, as well as 5-LO negative (F) U937 and Hek293T. Depicted are immunoblots of three subsequent cell passages, except for THP-1 cells, which comprise six replicates, since cells became slightly adherent after passage 2. Use of anti-5-LO antibody (6A12) 1:200, human recombinant 5-LO serving as positive, β-actin as loading control.

As expected, the myeloid cells HL-60, THP-1 and MM6 expressed 5-LO in detectable levels after differentiation with TGFβ/1,25$(OH)_2D_3$, the latter displayed the least prominent induction of enzyme expression level. Monocytes THP-1 exhibited comparable amounts of 5-lipoxygenase expression as promyelocytic HL-60 cells. Monocytic U937 did not provide a detectable level of 5-LO, probably due to their promoter methylation status [49]. The immunoblots further confirmed the constitutive enzyme expression of both B-lymphocytes Rec-1 and BL-41, the former showing less protein expression than the BL-41. Hek2937 as 5-LO negative cell line served as negative control. According to these results, HL-60, THP-1 and MM6 in both undifferentiated and differentiated state, as well as Rec-1 and BL-41 were chosen as representative cell lines for DNA-pulldowns.

4.1.2 Expression patterns of transcription factors

In order to confirm the presence of transcription factors for which evident or putative consensus binding sites exist in the proximal promoter sequence of the ALOX5 gene, the different cell lines were tested for a selection of TFs by immunoblotting. Furthermore, transcription factor WT1 was tested, since it is suspected to affect promoter activity according to reporter gene assays on the promoter sequence [310]. Table 5 gives an overview of the current knowledge on the TF-promoter interaction.

TF	DNA-binding domain	ALOX5-promoter binding	method
Sp1*	zinc finger, C_2H_2-type	direct [34]	EMSA
Sp3*	zinc finger, C_2H_2-type	direct [53]	ChIP
Egr-1*	zinc finger, C_2H_2-type	direct [46]	EMSA
VDR*	nuclear receptor, zinc finger, C4-type	direct [56]	EMSA/ChIP
RXR	nuclear receptor, zinc finger, C4-type	direct [56]	EMSA/ChIP
SMAD3/4*		direct [61]	EMSA/ChIP
AP-2α*		indirect [310]	reporter gene assay
WT1	zinc finger, C_2H_2-type	indirect [310]	reporter gene assay
GATA-1	zinc finger, GATA type	indirect [310]	reporter gene assay
Ets-1		indirect [310]	reporter gene assay

Table 5: Known or putative ALOX5-promoter binding transcription factors: classification and evidence of binding. TFs marked with an asterisk (*) possess consensus binding sites in the sequences used for the following DNA-pulldown. Source for localization of TFs: NCBI Gene Database

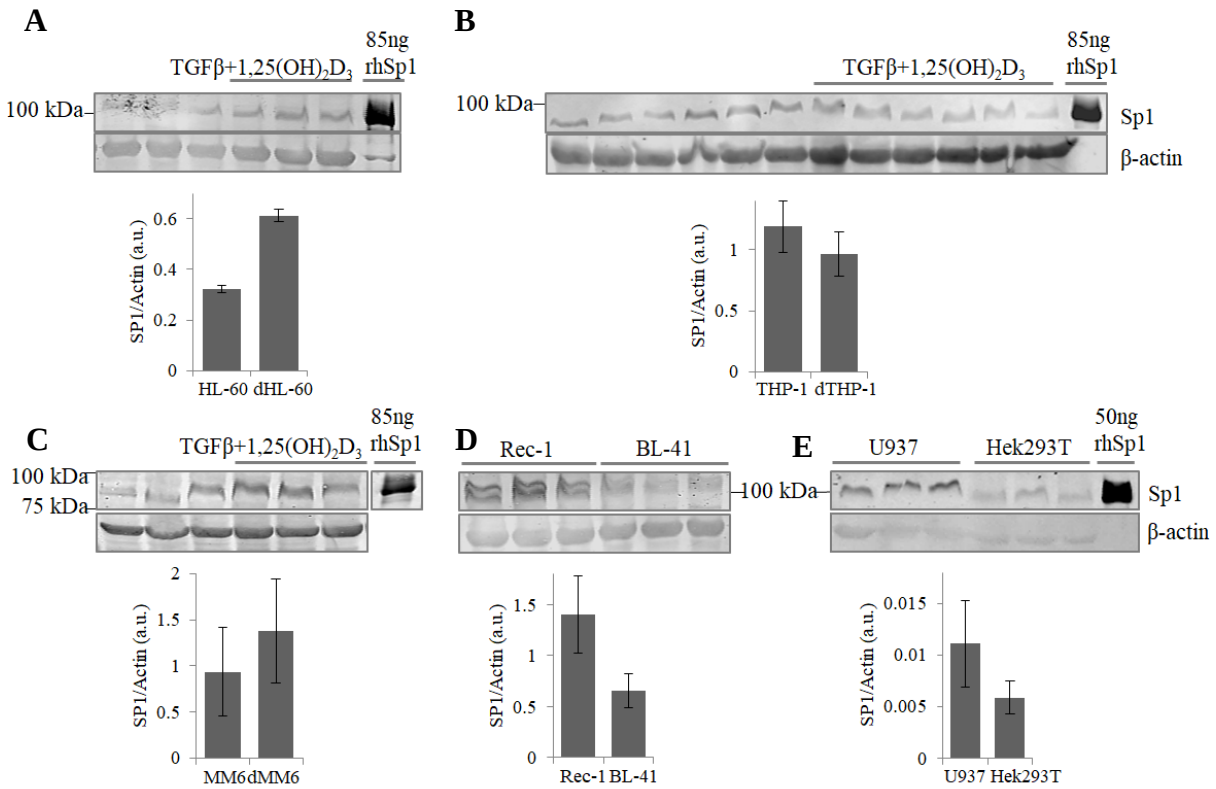


Figure 15: Immunoblot analysis of transcription factor Sp1 expression in different cell lines. 10-20µl whole cell lysate, each containing 30µg total protein were blotted against β-actin as loading control (anti-Sp1-ab 1:1000). All blots except the one including B-lymphocytes also comprised a positive control of the depicted amount of recombinant human Sp1 (rhSp1). Bar graphs (showing mean ± s. d.) represent a relative quantification based on densitometry between the depicted cell lines. Blots are representative of three subsequent cell passages of the used cell lines, except for THP-1, which include six biological replicates for undifferentiated (first six samples) and differentiated (last six samples) cells. Shown are (A) HL-60 in undifferentiated and differentiated (dHL-60), (B) THP-1 in undifferentiated and differentiated, (C) MM6 in undifferentiated and differentiated state, (D) Rec-1 and BL-41 and (E) 5-LO-negative controls U937 and Hek293T.

As depicted in the immunoblots, transcription factor Sp1 was consistently present in all cell lines and their respective differentiation state. The biological triplicates of myeloid cell lines HL-60 and MM6 displayed higher levels of Sp1 in their differentiated, than in their undifferentiated state, whereas protein expression level was reverse in THP-1 with a slight shift to the undifferentiated cells. B-lymphocytes Rec-1 exhibited a higher level of Sp1 than

BL-41, for the latter protein expression was hardly detectable. Monocytes U937 showed a comparable expression level to the myeloid cell lines. Sp1 was detectable in 5-LO negative cell line Hek293T, however protein levels were lower than those of the myeloid and B-lymphocytic cell lineage. Judging from the amount of recombinant Sp1 serving as positive control, the overall expression of the transcription factor lies within the lower nanogram range, corresponding to roughly estimated 0.1-0.01% of the amount of total cell lysate used.

Expression of nuclear receptor VDR was induced in the myeloid cell lines after treatment with TGFβ/1,25$(OH)_2D_3$, resulting in prominently detectable amounts in the respective differentiated states. While VDR protein level is barely to not present in the undifferentiated monocytes THP-1, MM6 and the promyelocytes HL-60, it is induced after differentiation up to 8.5-, 8- and 12-fold, respectively, according to densitometric detection. The B-lymphocytic cell lines displayed a different expression pattern of VDR, with BL-41 showing constant levels, whereas the protein level of Rec-1 deceeded the limit of detection. Faint but constant VDR levels were obtained for 5-LO-negative cell line U937; for Hek293T cells, detected bands did repeatedly alter slightly in their height in comparison to the control, so that no reliable identification can be concluded concerning VDR expression.

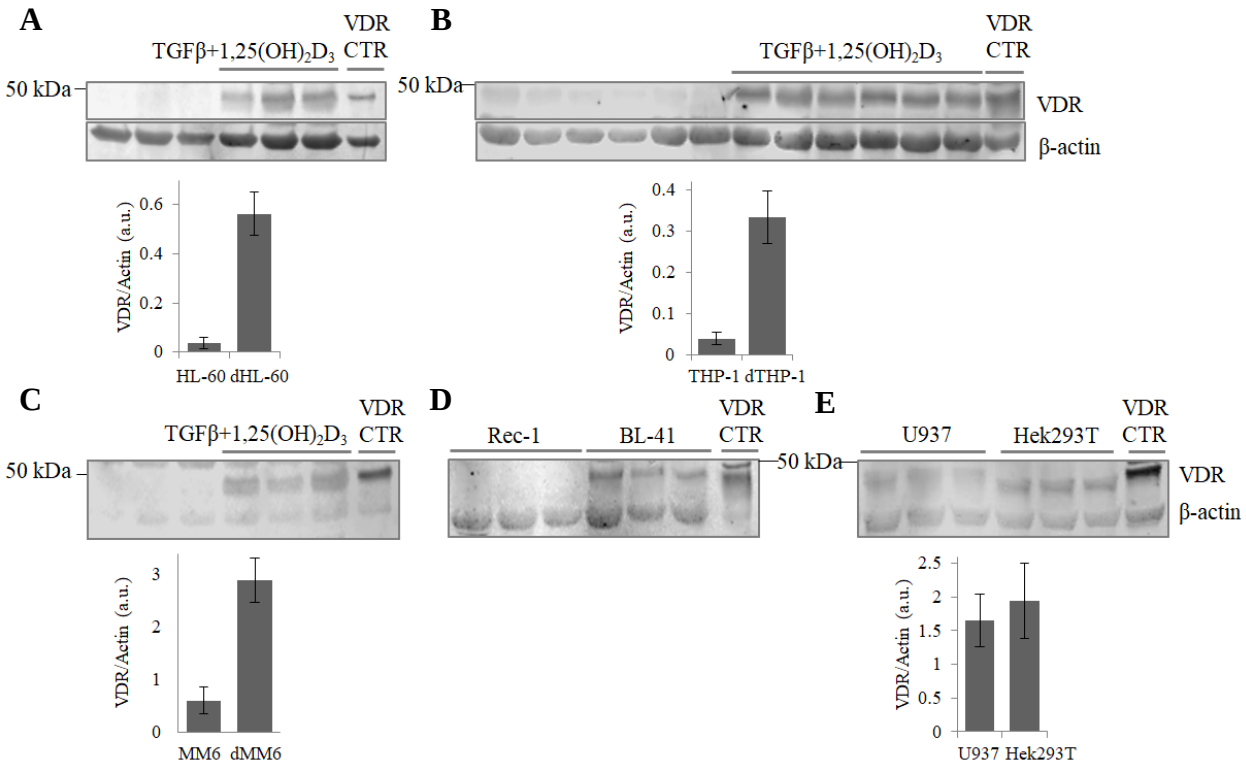


Figure 16: Immunoblot analysis of nuclear transcription factor VDR expression. 10-20µl whole cell lysate, each containing 30µg total protein were blotted against β-actin as loading control (anti-VDR-ab 1:200). Immunoblots comprised VDR-plasmid transfected cells, with overexpression of VDR, serving as control. Blots and bar graphs (mean ± s. d.), indicating a relative quantification based on densitometry between the depicted cell lines, are representative of three (six in case of THP-1) subsequent cell passages of (A) HL-60 in undifferentiated and differentiated (dHL-60), (B) THP-1 in undifferentiated and differentiated, (C) MM6 in undifferentiated and differentiated state, (D) Rec-1 and BL-41 and (E) 5-LO-negative controls U937 and Hek293T.

Mixed results were achieved for the nuclear receptor RXR, which was constantly expressed in HL-60 in both states to a comparable level and inducibly expressed in THP-1 after differentiation. Mature monocytes MM6 repeatedly did not exhibit any detectable protein amount, as did B-lymphocytes BL-41. In contrast, Rec-1, U937 and Hek293T displayed steady protein levels.

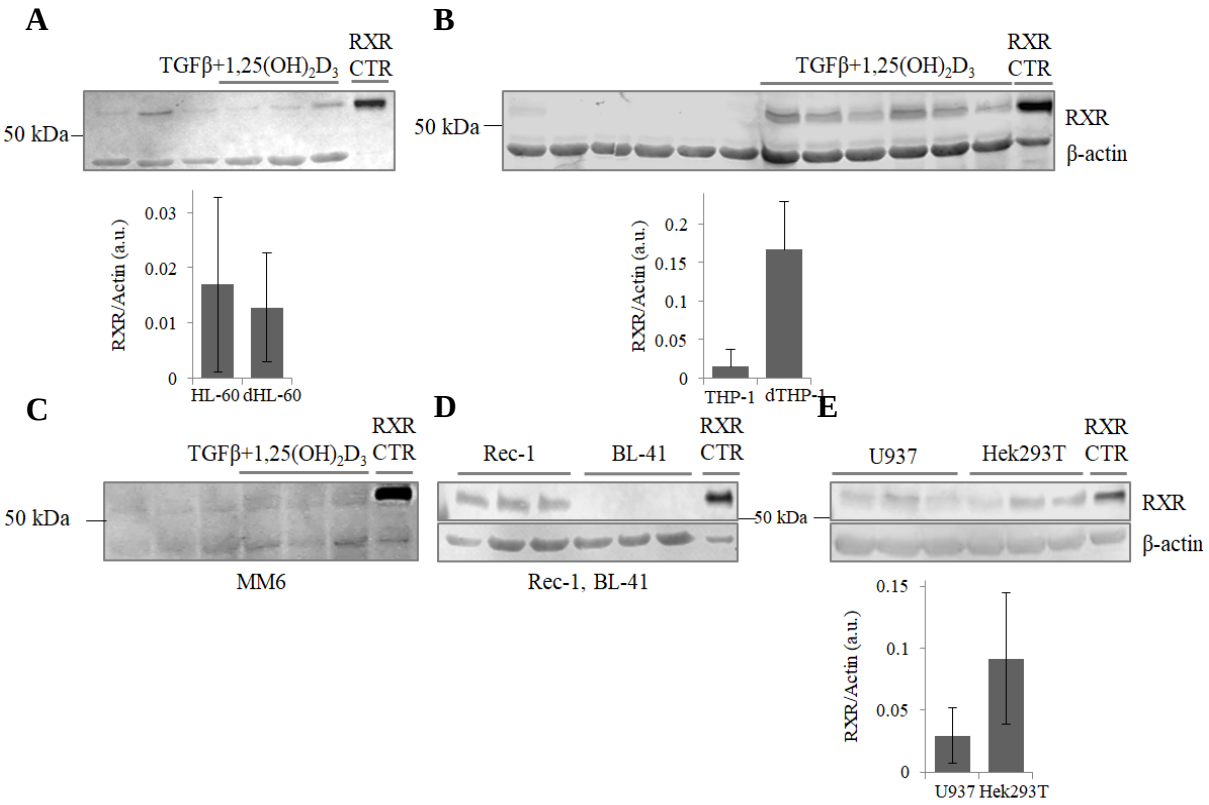


Figure 17: Immunoblot analysis of nuclear transcription factor RXR expression. 10-20µl whole cell lysate, each containing 30µg total protein were blotted against β-actin as loading control (anti-RXRα-ab 1:200). Immunoblots likewise comprised RXR-plasmid transfected cells, with overexpression of RXR, serving as control. Blots and bar graphs, indicating a relative quantification based on densitometry ± s. d. are representative of three (six for THP-1) subsequent cell passages of (A) HL-60 in undifferentiated and differentiated, (B) THP-1 in undifferentiated and differentiated, (C) MM6 in undifferentiated and differentiated state, (D) Rec-1 and BL-41 and (E) 5-LO-negative controls U937 and Hek293T.

Protein expression levels for transcription factors Egr-1 and WT1 were below the limit of detection under the used conditions for all cell lines tested. Similar results were obtained for transcription factor AP-2α, which was neither detected in the myeloid nor the B-lymphocytic cell lines, but solely in 5-LO negative Hek293T (supplementary figure 1).

4.2 Validation of the experimental setup for the identification of ALOX5 promoter interacting proteins

4.2.1 Workflow of DNA-pulldown

The major workflow used for the pulldown experiments was adapted from Hubner *et al.* [183] with minor modifications for identifying DNA-interacting proteins and is depicted in figure 18.

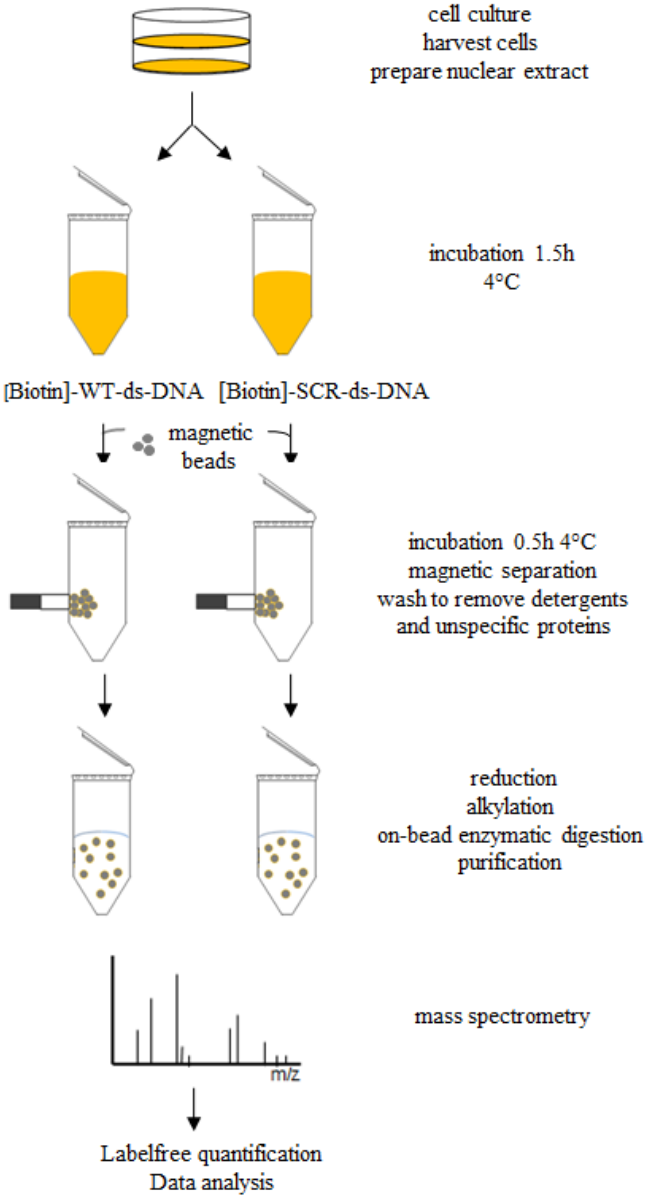


Figure 18: Workflow of the DNA pulldown used for the identification of DNA-interacting proteins by quantitative mass spectrometry.

For the identification of new regulatory proteins on genetic level, nuclear lysates of the aforementioned cell lines were incubated with synthetic double-stranded WT-DNA or CTR-

DNA and subjected to either mass spectrometric or immunoblotting-based identification of proteins. The incubation additionally contained nonspecific competitor DNA to reduce the amount of unspecific binders. Afterwards, formed DNA-protein complexes were coupled to magnetic beads as solid support, coated with streptavidin. The enriched protein complexes were washed after incubation to remove nonspecific binders and transferred to a new Eppendorf tube after the first wash. This ensures reduction of proteins that might have adsorbed to the tube wall. DNA-protein complexes were further washed with buffers, which did not contain detergents. This step is crucial to deplete detergents interfering with mass spectrometry, since their presence leads to saturation of C18-material and can suppress ion formation. For the identification of proteins via immunoblotting, beads were directly boiled in Lämmli-buffer, for MS-based identification, samples were reduced, alkylated and directly digested on the beads to minimize protein loss. Alternatively, after elution from the beads, proteins were subjected to SDS-PAGE and subsequent tryptic in-gel digestion for MALDI-MS/MS measurement.

4.2.2 Annealing of oligonucleotide sequences

Successful annealing of single-stranded forward and respective reverse strand of oligonucleotide is an essential step for providing functional ds-DNA, depleted of secondary structures that might form in solution (see section 1.2.1) and aggravate further experimental use. Common annealing protocols suggest buffer conditions containing low molar salt concentrations and a gradual thermal cooling of preheated oligonucleotides, which ensures correct formation of double-stranded DNA sequences. Single-strands were therefore mixed in equal amounts in a suitable buffer and heated up to 95 °C for a few minutes to disrupt hydrogen bonds present beforehand. Gradual cooling of the oligonucleotides then allowed for proper hybridization and formation of new hydrogen bonds of the respective complementary strands. Since the annealing process exacerbates with increasing numbers of base pairs, monitoring of the 80- and 120-mer oligonucleotides by agarose gel electrophoresis was inevitable. WT and corresponding SCR sequences in all cases contained the same amount of base pairs in a scrambled order to dispose of consensus binding sites present in the WT-sequence, but to maintain the same overall GC-/AT-content. Obviously, the used DNA stretches were able to form either secondary structures or dimers, as judged by multiple bands that were visible for the single-strands. However, they vanished after annealing, while generating ds-DNA, which clearly exhibited a sole band running at the anticipated height for the respective double-stranded sequences.

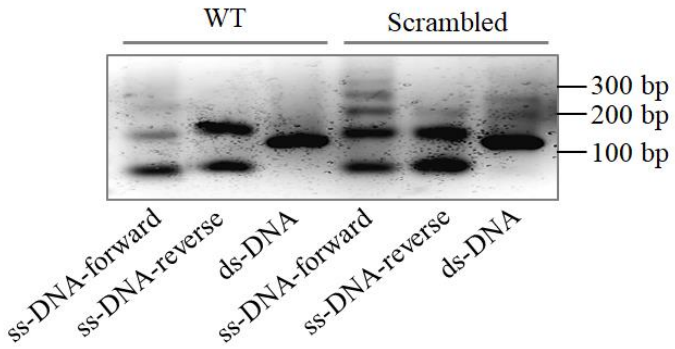


Figure 19: Validation of proper annealing of DNA oligonucleotides used in DNA pulldown. 3 µl of either single-stranded, and 5 µl of double-stranded DNA were loaded onto the 3% ethidium-bromide agarose gel, which was then run at a constant voltage of 75 V for 1 hour. The depicted results are representatives of three independent experiments.

4.2.3 Subcellular fractionation

In order to reduce sample complexity for subsequent DNA-pulldowns and focus on proteins deriving from the nucleus, whole cell lysates were fractionated to isolate nuclear lysates. Cell pellets were resuspended in hypotonic buffer to disrupt outer membranes and separate cytosol from crude nuclei. For extracting nuclear proteins, nuclei were in turn incubated with hypertonic buffer containing salt concentrations that maintain protein functionality and activity at maximum [311]. Cytosolic and nuclear fractions were subjected to immunoblotting to verify successful fractionation, histone H3 and β-actin served as control proteins for nucleic or cytoplasmic fraction, respectively. All nuclear extracts, except for the ones of B-lymphocytes BL-41, displayed clear separation from the cytosolic compartment, providing adequate removal of interfering cytosolic proteins. A light β-actin band could be found in the nuclear fraction of BL-41, which however was sufficiently weak to be negligible for subsequent DNA pulldowns, indicating that cytosolic proteins cannot be removed entirely from nuclear extracts. Cytosolic compartments of all the fractionated cell lines exhibited parts of the nuclear compartment, as judged by the detected signals for histone H3, which were due to the number of pestle strokes leading to the disruption of the nuclear membrane. A constant number of strokes indeed was used for all the cell lines to ensure processing homogeneity for all of the lysates and rather dispose of a small share of nuclear extract than intermingling the latter with cytosolic compartment. The overall subcellular fractionation provided decent separation of cytosol and nucleus to further use nuclear lysates for succeeding experiments.

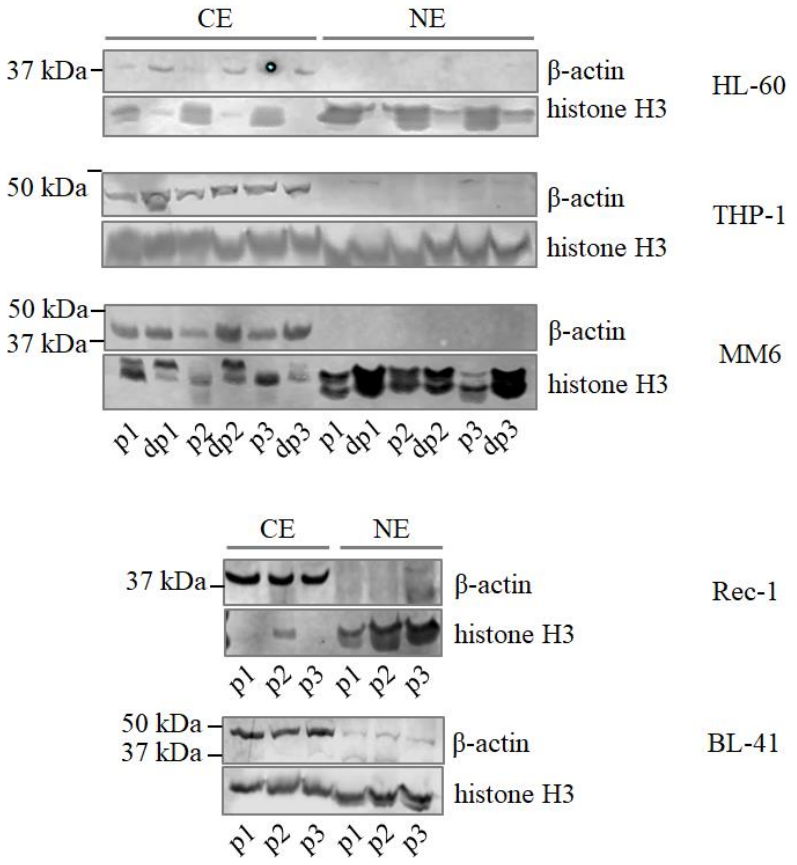


Figure 20: Validation of proper subcellular fractionation. Immunoblotted marker proteins for the respective fractions were β-actin (cytosolic fraction) and histone H3 (nuclear fraction). 3µl and 6µl of cytosolic and nuclear extracts were used for all blots, respectively. Depicted blots are representatives of three independent experiments (consecutive cell passages). CE = cytosolic extract, NE = nuclear extract, p = passage, dp = differentiated passage.

4.2.4 DNA pulldown using model proteins

4.2.4.1 Limit of detection

To assess the functionality of the experimental setup, transcription factor Sp1 was used as model protein for evaluation, since it is known to possess multiple consensus binding sites in, and binds to the tandem GC-box as already described in section 1.1.5.1. Therefore, knowing its detection limit in anticipated consecutive immunoblotting and SDS-gel electrophoresis experiments is essential for determining the amount of starting material necessary to perform consecutive DNA pulldowns. Consequently, different concentrations of human recombinant full size Sp1 were subjected to SDS-PAGE and western blotting, as depicted in figure 21. Protein amounts in gel electrophoresis ranged from 480 ng to 95 ng at minimum, corresponding to amounts of 6-1.18 pmol. 1.18 pmol Sp1 was the lowest amount detectable

after staining the gel with colloidal Coomassie stain, which is a prerequisite for a possible subsequent in-gel digestion and mass spectrometric detection.

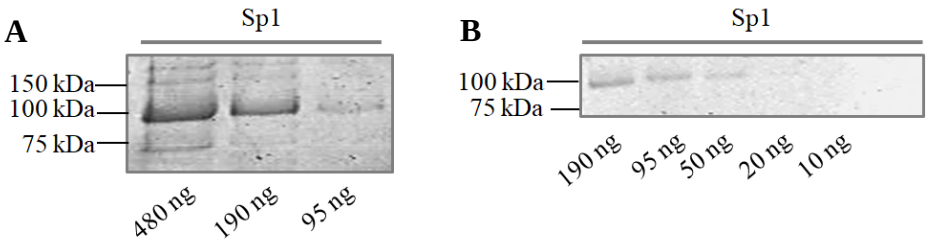


Figure 21: Limit of detection of human recombinant transcription factor Sp1 in both (A) SDS-PAGE gel electrophoresis stained with colloidal Commassie solution and (B) Western blot.

Immunoblots were able to detect lower levels of Sp1 with the used antibody, overall ranging from 190 ng to 10 ng total protein amount, corresponding to 2.35-0.125 pmol. The middle concentrations used could still be decently detected in western blots, resulting in a practical limit of detection at around 50-20 ng of total protein.

Assuming that the percentage share of transcription factors in total cell lysate ranges from 0.1-0.01% (see section 4.1.2), isolated nuclear extracts hence will contain an enriched share thereof. Thus, an accurate amount of necessary input material of nuclear extract cannot be determined this way, but a rough estimate can be deduced; based on a minimum of 0.01% proteins in whole cell lysate and a detection limit of 95 ng and 50-20 ng for SDS-PAGE and immunoblotting, respectively, at least 950 µg and 500-200 µg total cell lysate would be necessary for functional experimental setup for each approach. Since nuclear extracts contain enriched quantities of the proteins of interest, these amounts of NE should be sufficient for a successful DNA pulldown for both methods.

For further elucidation of the required amount of starting material in ESI-MS/MS based approaches, titration experiments with differing quantities of both oligonucleotide and NE were carried out via label-free quantification. Aiming at an adequate amount for significant identification of model proteins, 10, 50, 100 and 250 pmol of WT-oligonucleotides were incubated with 50, 100, 250 and 500 µg NE of monocytic MM6. Again, Sp1 was chosen as one of the model proteins indicating a functional DNA pulldown. Additionally, three more proteins were chosen for their known properties of binding to GC-boxes (Sp3 and KLF16) or showing high affinity for Sp1-consensus sequences and GC-rich DNA strands (MAZ). All of the 16 DNA pulldowns performed were visualized in heat maps, according to their input of

starting materials and number of identified peptides for the respective proteins; these are shown in figure 22.

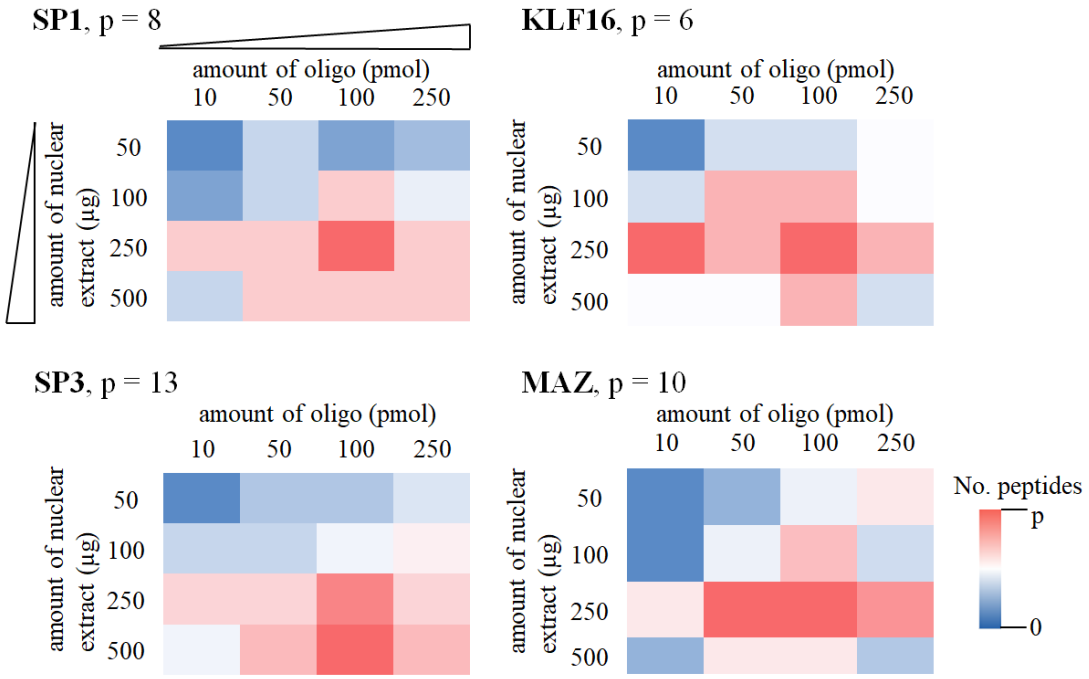


Figure 22: Optimization of the DNA pulldown protocol: Titration experiments with different amounts of oligonucleotide and nuclear extract. Color gradient ranges from blue for low numbers of identified peptides per protein, over white (intermediate) to red, depicting high numbers of identified peptides per protein. p = maximum number of peptides detected for the respective protein in the respective pulldown experiments.

For all proteins investigated, a trend could be ascertained showing a correlation of higher numbers of identified peptides to higher amounts of nuclear extract used. 50 and 100 µg of nuclear extract therefore were considered to be insufficient for providing the possible maximum of identified peptides per protein, probably due to protein loss during sample preparation. Interestingly, a higher amount of oligonucleotides did not necessarily lead to a higher number of peptides; the use of 250 pmol in combination with 500 µg NE even reduced the number of identified peptides for both KLF16 and MAZ. High concentrations of oligonucleotides can provide more putative binding targets for proteins exhibiting affinity for nucleic acids. This can possibly lead to inadvertent enrichment of background binders, thereby impairing selectivity of the experimental setup, and thus resulting in a lower number of identified peptides of the proteins of interest. In contrast, the use of 10 pmol of oligonucleotides only provided a low to intermediate identification rate with all quantities of

NE tested. Consequently, 50 and 100 pmol were suitable for capturing the maximum of peptides per protein possible. Since three of the four proteins exhibited highest numbers of identified peptides by using 100 pmol oligonucleotide, these were set as standard amount for the established DNA-pulldown protocol. In terms of nuclear extract, a minimum of 250 µg was considered to be essential for detecting proteins adequately sensitive, which could be increased up to the desired amount. However, since more than 400-500 µg of nuclear extract were hardly contrivable with cell cultural conditions, the standard amount of input NE was set to 400 µg. This also was in line with the former achieved estimated results for determining required amounts of input material by using SDS-PAGE and immunoblotting methods.

4.2.4.2 DNA pulldown and Western blot detection

In order to validate the estimated protein and oligonucleotide amounts determined as starting material in their respective suitability for a successful pulldown assay, the specific enrichment of transcription factor Sp1 from Rec-1 nuclear lysates was tested. The wildtype ALOX5 promoter sequence served as bait for capturing proteins, which consecutively were analyzed by immunoblotting and were compared to the ones isolated from the scrambled oligonucleotides. As shown in figure 23, Sp1 could easily be detected in the nuclear lysates, confirming that the setup would provide sufficient starting material for the pulldown assay.

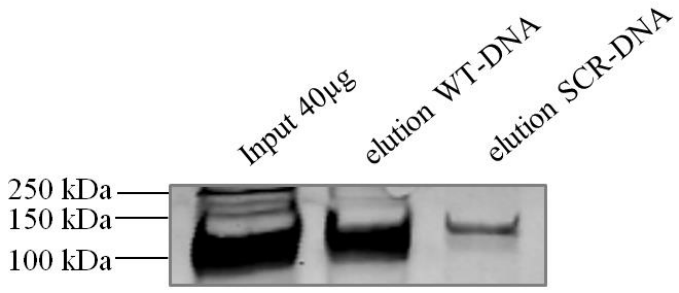


Figure 23: Immunoblot-based DNA pulldown. 400 µg nuclear extract were incubated with either WT or CTR ds-DNA and subjected to immunoblot (anti-Sp1-ab 1:1000). Input material corresponds to 1/10 of NE amount per approach. The shown immunoblot is one representative of three independent experiments.

The transcriptions factors' expected binding to the proximal ALOX5 promoter could furthermore be verified. The wildtype elution obviously exhibited a higher Sp1 content than the control elution, with 4-(±1.5)-fold enrichment over the scrambled sequence. Thus, a functional experimental setup for the DNA pulldown could be deduced from these results.

4.2.4.3 DNA pulldown with transcription factor WT1

After establishing a functional pulldown assay, the protocol was applied to transcription factor WT1 for elucidating its binding capability to the proximal ALOX5 promoter. As already mentioned in section 4.1.2, there is existing evidence of the transcription factor's effects on 5-LO expression, as judged by reporter gene assays [310]. WT1 represents a protein with multiple roles acting as transcription factor on transcription or post-transcriptional level, as well as contributing to the early development of specific organs, most prominent of the kidney. Its potential involvement in the generation of different diseases is dependent on its specific isoform, arising from alternative splicing [312, 313]. One specific isoform of WT1 containing the amino acid sequence lys-thr-ser (+KTS) was shown to activate the promoter comprising the tandem GC-boxes, whereas its absence results in sole activation of promoter constructs lacking the GC-boxes.

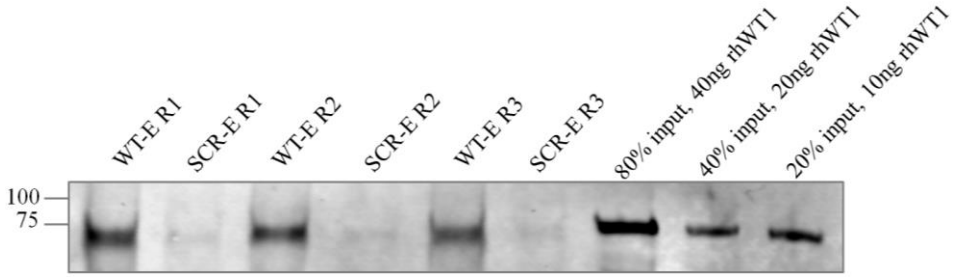


Figure 24: Immunoblot-based DNA pulldown with transcription factor WT1. 50 ng of recombinant human WT1 (+KTS) were used per approach and after binding to either WT or SCR ds-DNA subjected to western blotting. anti-WT1-ab 1:1000, R = replicate number, WT = wildtype, SCR = scrambled.

In order to provide evidence for its binding to the ALOX5 promoter sequence of 120 bp, a western blot based pulldown was performed, which also contained mass standards of human recombinant WT1 (rhWT1) to estimate its total amount. 50 ng of rhWT1 therefore were used for each approach in WT and SCR samples. As shown by the immunoblots (Figure 24), wildtype DNA resulted in a 9-(±0.5)-fold enrichment of rhWT1 over control DNA, thereby proving its specific binding to the promoter sequence. Based on densitometry, the total amount of human recombinant WT1 was estimated to be 23.5 ng (±6.5 ng).

4.3 DNA pulldown and MS analysis

4.3.1 In-gel digestion and MALDI-MS/MS

Since the classical approach for proteomic research included gel electrophoresis coupled to mass spectrometry in the first place, the anticipated DNA pulldown was initially performed using denaturing SDS-PAGE followed by in-gel tryptic digestion of proteins and identification by MALDI-MS/MS. Reproducible staining therefore is a prerequisite for comparable results obtained by in-gel digested proteins; the used colloidal Coomassie staining solution is reported to be able to visualize amounts as low as 1ng, which corresponds to the 10-fold quantity of potentially detected silver-stained proteins [314, 315]. However, as silver staining is only partly compatible with MS and lacks satisfying quantification linearity, the Coomassie stain provides identification rates, which are more reproducible [315].

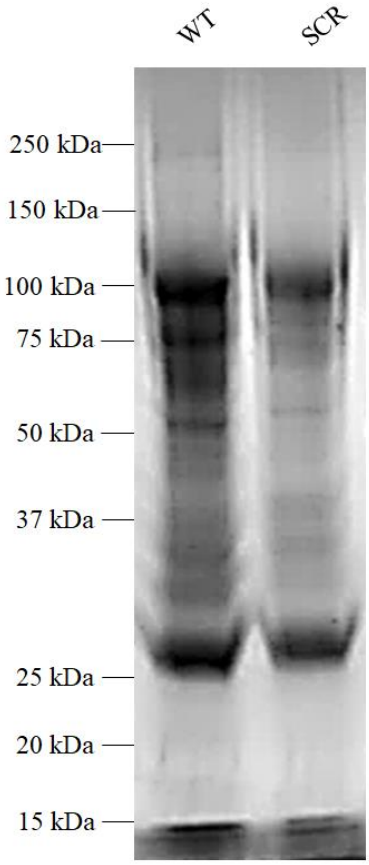


Figure 25: SDS-PAGE-based DNA pulldown of HL-60 nuclear extract with either WT or SCR ds-DNA. 400 µg of NE in a total volume of 20-30µl were used for each approach and subjected to gel electrophoresis. Gel was stained with colloidal Coomassie blue, extracted proteins were further subjected to in-gel tryptic digestion followed by MALDI-TOF-MS/MS.

DNA pulldowns (WT and SCR) were carried out with 400 µg of HL-60 nuclear extracts and stained with colloidal Coomassie blue solution overnight in three independent experiments. Protein elution fractions of WT and SCR of one representative gel are depicted in figure 25. Obviously, there exist only minor differences between both samples, whose lanes resemble one another, which cumber the successful discrimination of proteins found separately in one of the pulldowns. This is partly due to the low electrophoretic resolution achieved. Especially proteins in the lower mass range lack sufficient resolution, thereby impeding decent distinction of both lanes. Protein running behavior additionally seems to be influenced by the used bead matrix, as judged by the existent blur throughout the lanes, which generally resulted in mass spectrometric identification of streptavidin peptides. Identification of affinity-enriched proteins additionally provided only a low number of detectable protein bands altogether, possibly arising from too little starting material used. However, after increasing the amount of nuclear lysate for each sample to 800 µg, neither resolution nor overall protein identification could be improved (supplementary figure 4). Thus, there is only a low representation of protein complexes that are enriched via DNA pulldown in the performed one-dimensional gel electrophoresis, obviously due to a combination of insufficient amounts of used input material and lacking sensitivity, poor resolution and interfering streptavidin peptides that overlay target proteins during mass spectrometric measurement. A cumulative overview of the identified proteins is shown in table 6. Out of these 21 proteins, none is directly associated with transcriptional or gene regulatory effects. Almost all contribute to biogenesis of rRNA, ribosomal processing or are representatives of ribosomal proteins themselves, hence are part of the ribosomal machinery but basically masking DNA-binding proteins of interest. Further highly abundant proteins like histones, nucleolin and poly-(ADP ribose)-polymerase could be identified, which might overcast those of lower abundance additionally, which exacerbates identification of direct DNA-interacting regulators. Altogether, separation efficiency and sensitivity of the applied SDS-PAGE followed by MALDI-MS/MS does not suffice to provide a suitable result for the significant identification of DNA-interacting proteins.

kDa	protein	gene
208.7	protein RRP5 homolog	PDCD11
113	Poly [ADP-ribose] polymerase 1	PARP1
87.3	nucleolar RNA helicase 2	DDX21
76.6	Nucleolin	NUCL
73.6	nucleolar and coiled-body phosphoprotein 1	NOLC1
69.5	DNA repair protein XRCC1	XRCC1
66	nucleolar protein 56	NOP56
59.5	nucleolar protein 58	NOP58
54.1	Gamma-aminobutyric acid receptor subunit rho-2	GABRR2
34.8	rRNA/tRNA 2'-O-methyltransferase fibrillarin-like protein 1	FBLL1
33.8	rRNA 2'-O-methyltransferase fibrillarin	FBL
20.2	60S ribosomal protein L11	RPL11
18.6	60S ribosomal protein L21	RPL21
17.8	60S ribosomal protein L12	RPL12
17.2	H/ACA ribonucleoprotein complex subunit 2	NHP2
16.5	60S ribosomal protein L27a	RPL27A
15.5	histone H3.1t	HIST3H3
14	histone H2A type 1-B/E	HIST1H2AB
14	NHP2-like protein 1	SNU13
13.9	histone H2B type 1-B	HIST1H2BB
11.3	histone H4	HIST1H4A

Table 6: Proteins identified by SDS-PAGE-based DNA pulldown coupled to in-gel digestion and MALDI-MS/MS

4.3.2 LC-ESI-MS/MS

4.3.2.1 Choice of quantification method

In order to determine the quantification method of choice, both label-based and label-free DNA pulldowns were carried out and subsequently compared to each other in terms of identified proteins, their respective exclusivity in one of the strategies and their physico-chemical properties. Model cell line HL-60 was chosen in undifferentiated (HL-60) and differentiated (dHL-60) state. Dimethyl labeling was carried out after digestion of the enriched proteins by adding formaldehyde (light label) or deuterated formaldehyde (heavy

label) for the respective WT or SCR version of the ALOX5 promoter sequence. Labels were swapped once to avoid any bias of the actual labeling reaction. Log2-ratios of all quantified proteins of the duplicates were visualized in scatter plots and significant outliers were identified based on protein intensity (significance B: p≤0.05). Proteins gathering around the zero value represent nonspecific interactors, since both light and heavy labeled peptides are present in equal amounts after combination of WT and SCR sample. The outliers in the lower left and upper right quadrant are significantly interacting proteins of the SCR and WT pulldowns, respectively.

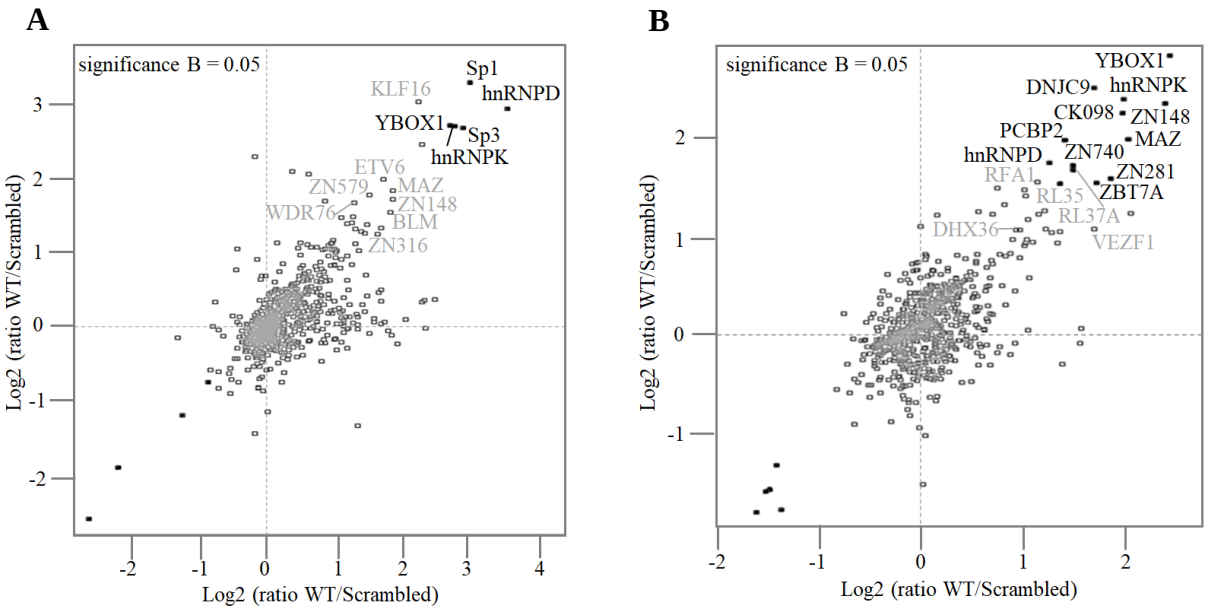


Figure 26: Dimethyl labeling-based quantification of enriched proteins and respective significant interacting proteins of the proximal ALOX5 promoter. 400 µg of nuclear extract of (A) HL-60 and (B) dHL-60 were incubated with WT or SCR DNA sequence and after pulldown and digestion labeled with light or heavy label, combined and subjected to LC-ESI-MS/MS. Significant outliers (black squares) were identified based on protein intensity (significance B), requiring a threshold of p = 0.05. Proteins that also occur in the label-free pulldown or represent artefacts are marked in grey. Scatter plots represent the respective log_2-ratios of LFQ intensities of each protein in both forward and reverse experiment.

In total, 1244 proteins were identified for the label-based quantification strategy for the undifferentiated and 1034 for the differentiated status, of which 735 (HL-60) and 673 (dHL-60) were kept for statistical analysis after filtering for contaminants, reverse hits and valid values. Proteins that were classified as significant outliers comprised 5 interactors for the undifferentiated and 13 for the differentiated state, as depicted in table 7. Outliers that were

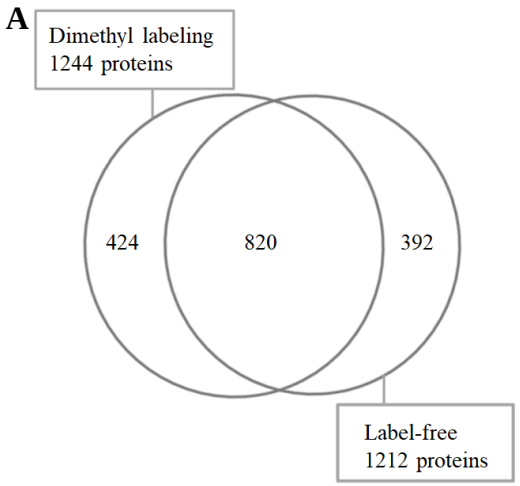
determined for the control DNA sequence are not given any further focus, since these do not contribute to the understanding of interacting proteins of the ALOX5 promoter sequence, but also cannot be avoided entirely. Among the identified interactors Sp1 and Sp3 were found for HL-60, thereby proving the functionality of the assay, as these transcription factors are known to interact with the GC-boxes present in the proximal promoter. However, these were solely found in the case of undifferentiated cells and were lacking for the differentiated ones. Nonetheless, the latter provided further transcription factors with C_2H_2-type zinc fingers as their DNA binding domain (MAZ, ZN281, ZN148, ZN740, ZBT7A). Overlap was found for hnRNP D, hnRNP K and YBOX1, all of them involved in transcriptional regulation and recognition of G4-DNA [115, 316]. Proteins that were additionally found in the comparative label-free DNA pulldowns or form clear artefacts (for instance ribosomal proteins RL35, RL37A) are marked in grey.

protein	gene name	function
HL-60		
Sp1	Sp1	transcription factor
Sp3	Sp3	transcription factor
HL-60 and dHL-60		
hnRPD	hnRNP D	transcription factor, G-quadruplex interacting protein
hnRPK	hnRNP K	transcription factor, G-quadruplex interacting protein
YBOX1	YBX1	transcriptional regulator, G-quadruplex interacting protein
dHL-60		
DNJC9	DNAJC9	may be co-chaperone
CK098	C11orf98	uncharacterized
PCBP2	PCBP2	oligo-dC binding protein
ZN148	ZNF148	transcription factor
ZN740	ZNF740	transcriptional regulator
ZN281	ZNF281	transcription repressor
MAZ	MAZ	transcription factor, G-quadruplex interacting protein
ZBT7A	ZBTB7A	transcription factor

Table 7: Overview of proteins identified by label-based quantification

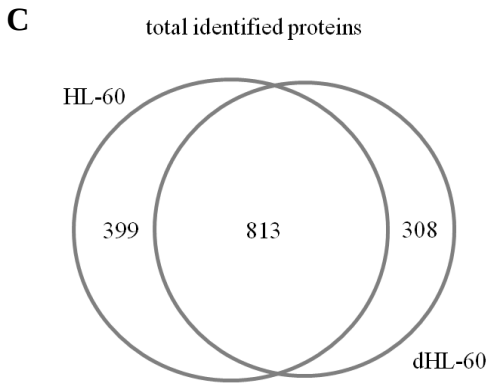
For choice of quantification methods the dimethyl labeling approach of HL-60 and dHL-60 was compared to label-free quantification. LFQ was likewise carried out with the wildtype DNA sequence of the ALOX5 promoter and the scrambled version thereof, serving as control. In contrast to duplicates for label-based quantification, triplicates were needed for subsequent t-test-based statistics. Results are visualized in volcano plots, according to their log2 intensity difference and -log10 p-value of permutational FDR-based two-tailed t-test. Thus, significant interactors are identified by determining a cutoff line, which depends on the FDR and basic minimal fold change (s0 value).

The DNA pulldowns of both cell states for label-free quantification resulted in a total amount of identified proteins of 1212 for HL-60 and 1121 for dHL-60, 744 (HL-60) and 613 (dHL-60) selected for statistical analysis after filtering the protein lists. Thus, similar amounts of proteins were enriched in both quantification strategies. These showed good overlap of 820 proteins (HL-60) and 695 (dHL-60) between both methods. Altogether 38 and 20 significant interactors were determined for HL-60 and dHL-60, respectively, applying a similar statistical evaluation ($p \leq 0.05$) as in the label-based strategy (for explicit identification see suppl. table 1). When comparing identified protein patterns between the cell states, no major difference could be distinguished, judged from an overlap of 813 proteins. Thus, the enriched protein entity resembles one another and indicates the absence of cell differentiation distinctions. Out of the enriched significant interacting proteins eleven (BLM, hnRNP K, KLF16, MSH3, RFA3, RFC1, Sp1, VEZF1, ZBT7A, ZN281, ZN579) are shared between the cell states, which supports the assumption that gene regulation is independent of differentiation state.



B

quantification method	cell state	identified proteins	significant interactors
Dimethyl labeling	HL60	1244	5
	dHL60	1034	13
Label-free	HL60	1212	38
	dHL60	1121	20



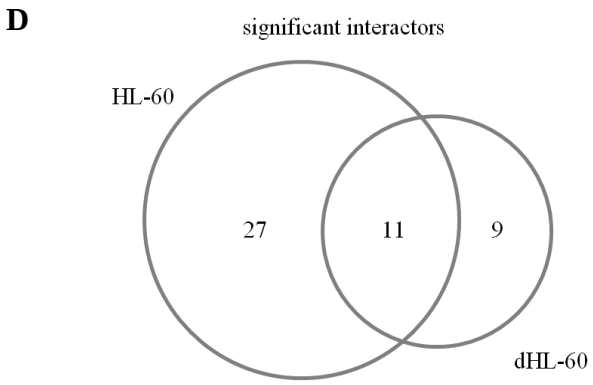


Figure 27: Label-free DNA pulldown of HL-60 and dHL-60. (A) Comparison of the total amount of identified proteins in label-based and label-free quantification strategy and overlap between the identified proteins. (B) Explicit representation of the overall identified proteins according to the respective quantification method and used cell state and respective number of significant interactors. (C) Comparison of label-free DNA pulldowns in terms of total number of identified proteins of HL-60 and dHL-60 and their corresponding overlap, as well as the (D) overlap of its respective significant interactors in both differentiation states.

Considering the functions of the identified proteins for HL-60 in both states, there is a clear trend towards DNA-interacting proteins, including transcription factors and transcriptional regulators. As in the case for the dimethyl labeling method, Sp1 and Sp3 could be found in undifferentiated HL-60, confirming a functional experimental setup. Furthermore, pulldowns revealed novel TFs and genetic regulators, which in part did overlap with the ones identified in the label-based approach (e.g. MAZ, ZN281, ZN148, YBOX1, hnRNP D/K, BLM, ZBT7A), but also provided additional interactors, whose function will be discussed cumulatively in sections 4.3.2.4 and 5.3.1 for all myeloid cell lines.

To exclude any bias present in the enriched protein patterns in both label-based and label-free method, physico-chemical properties of the total identified proteins were compared to each other. Therefore, isoelectric point, molecular weight and GRAVY-score were inspected. As shown in figure 28, a shift towards lower molecular weights was observed for the label-free approach as compared to the label-based. Further additional small distributions are found around 200-250 kDa and 270-320 kDa, however these findings do not alter the overall distribution and thereby equivalence of identified protein patterns is maintained. Finally, higher molecular weights of more than 400 kDa were more efficiently detected in label-free quantification as well. For neither of the parameters a strong difference was observed, indicating an unbiased approach for either of the used methods.

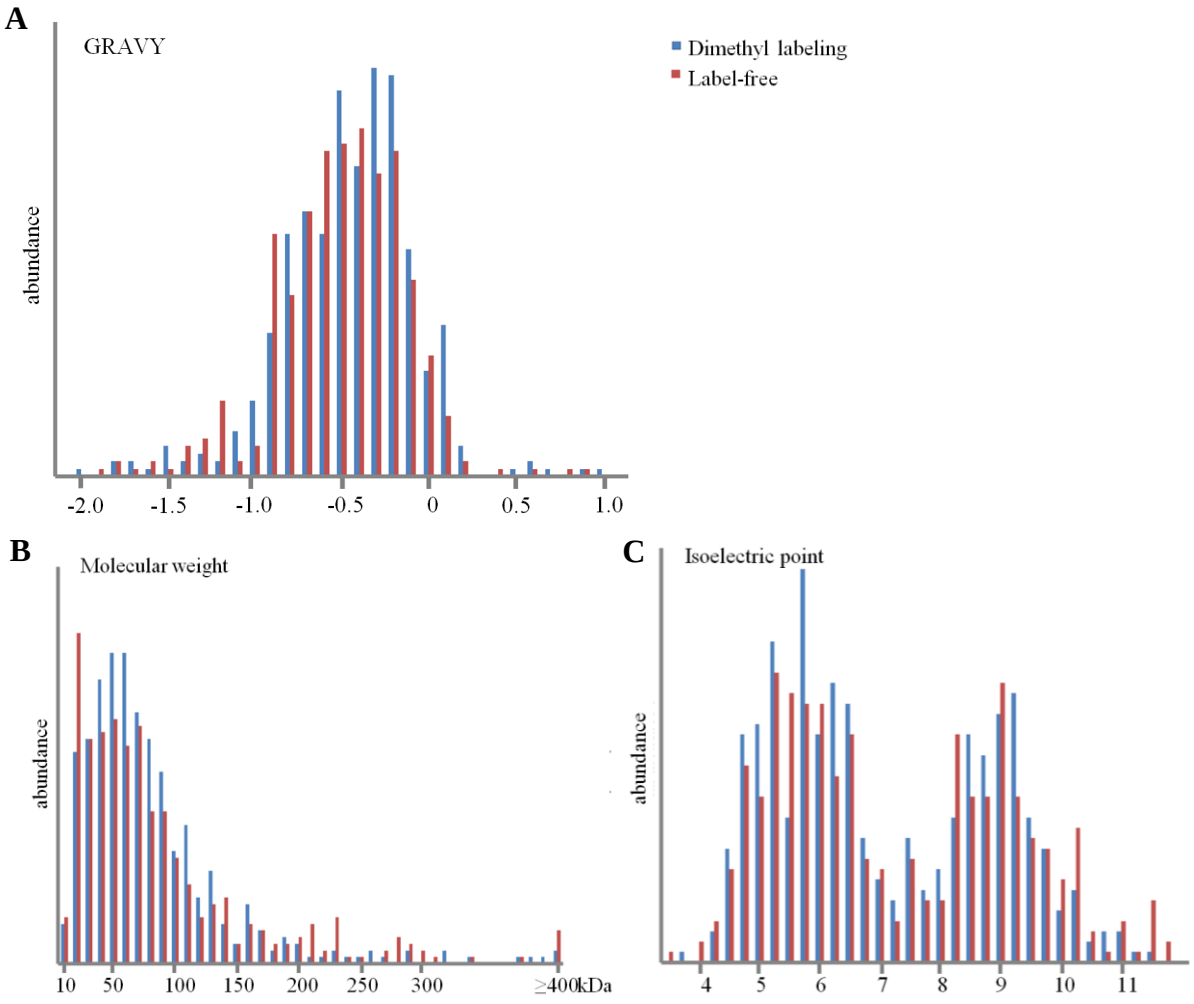


Figure 28: Comparison of physico-chemical properties of the identified proteins in label-based and label-free approach, displaying (A) GRAVY score, (B) molecular weight and (C) isoelectric point. All data is provided by https://web.expasy.org/protparam/.

The results obtained provided two equally suited methods for quantifying enriched proteins, with no alterations in identified protein patterns or significant interactors. Out of the significantly interacting proteins in the label-based pulldowns, 4 (HL-60) and 6 (dHL-60) were also found in the label-free approach. Dimethyl labeling provided a higher number of total identified proteins for undifferentiated HL-60 cells, which however did not contribute to a higher number of significant interactors. In contrast, label-free quantification clearly identified more interacting proteins in both undifferentiated and differentiated myelocytes. Thus, due to the higher number of significant interactors label-free quantification was chosen for all subsequent DNA pulldowns.

4.3.2.2 Evaluation of nonspecific competitor

The early approaches for the investigation of differentially expressed protein patterns by proteomics mainly focused on qualitative analysis, unspecific binding partners represented the major problem in distinguishing specifically enriched proteins (see section 1.3.2). This was mostly, but not always successfully addressed by adding high amounts of DNA competitors to prevent or minimize nonspecific binding. In order to evaluate whether such high amounts of competitor are still needed in quantitative approaches, or whether they would alter the observed protein composition, DNA pulldowns with excess competitor were carried out, again for the model cell line HL-60 in the undifferentiated state. The experimental setup was kept as usual ($p \leq 0.05$), with the exception of adding 40-fold sheared herring sperm DNA to the incubation. Results are given in figure 29. 1134 proteins were identified in total in good correlation to the pulldown performed with lower amounts of competitor, while the identification of significant interactors generated only about half of the number of proteins identified in the initial pulldown. Only 16 proteins were classified as significantly interacting with the ALOX5 promoter sequence, which can be affiliated to more stringent incubation conditions. However, functional investigation of the respective interactors revealed a similar pattern, including shared transcription factors of the Sp-family (Sp1, Sp3, Sp4), as well as transcriptional regulators (ZBT7A, ZN148, PRD10) and proteins involved in DNA repair (RFA1, DDB2). Proteins associated with transcriptional regulation that were exclusively found in the pulldown with 40-fold competitor only comprised TF SP110. Other identified proteins did not exhibit further function in gene regulation. Altogether, the 40-fold use of nonspecific competitor did not result in an altered protein composition of interactors or other significantly interacting proteins. As mentioned in section 1.3.2 quantitative analysis superseded the need for excess competitor, and are presumably sufficient to maintain transient

complexes. For these reasons, subsequent pulldowns were performed with the standard setup as evaluated in section 4.2 (1.35-fold poly-dIdC as competitor), without adding any further excess of competitor.

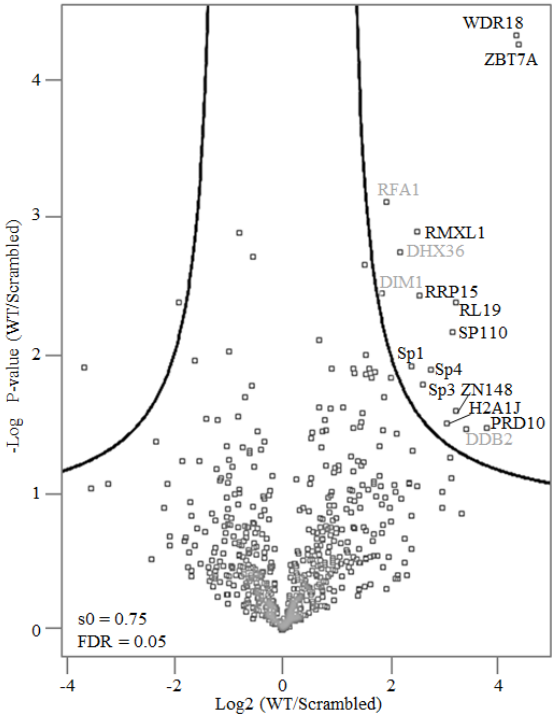


Figure 29: Label-free DNA pulldown of ALOX5 core-promoter sequence and 40-fold excess competitor. Significant interactors were identified based on two-tailed t-test and are presented in the upper right corner. Proteins that were identified below a threshold of at least 3 peptides (one of them being unique) are marked in light gray. The ratio of LFQ intensities of WT and SCR (x-axis, log_2-transformed) is plotted against the respective p-values (y-axis, log_{10}-transformed).

4.3.2.3 Evaluation of bioinformatic processing parameters

Additional means to substantiate a functional DNA pulldown or to ascertain the feasibility of statistical data processing are provided by different steps in data analysis, which is executed after LC-MS/MS measurement. These processing criteria do not have to be fulfilled necessarily in their entirety, but they contribute to a rough and quick primary estimation of data suitability. Representative examples will be given for the LFQ pulldown of differentiated HL-60.

Distribution of protein intensities

Since the achieved raw data generally do not follow a certain distribution on the first glance, it is necessary to apply a statistical test that is suitable for their respective analysis. In this context, parametric tests, like the Student's t-test, are often preferred over non-parametric

tests due to their higher power and therefore are commonly used for proteomics data [317]. In order to make unprocessed protein intensities accessible for Student's t-test, they have to be transformed, generally by logarithmic transformation. This usually provides normal distributions, which are accessible for parametric testing [318]. To test for normal distribution, data can be visualized for example in histograms, which should show a Gaussian distribution for each of the individual samples. As depicted in figure 30(A), each of the dHL-60 pulldowns shows overall normal distribution, with slight left-handed bias in four out of six samples.

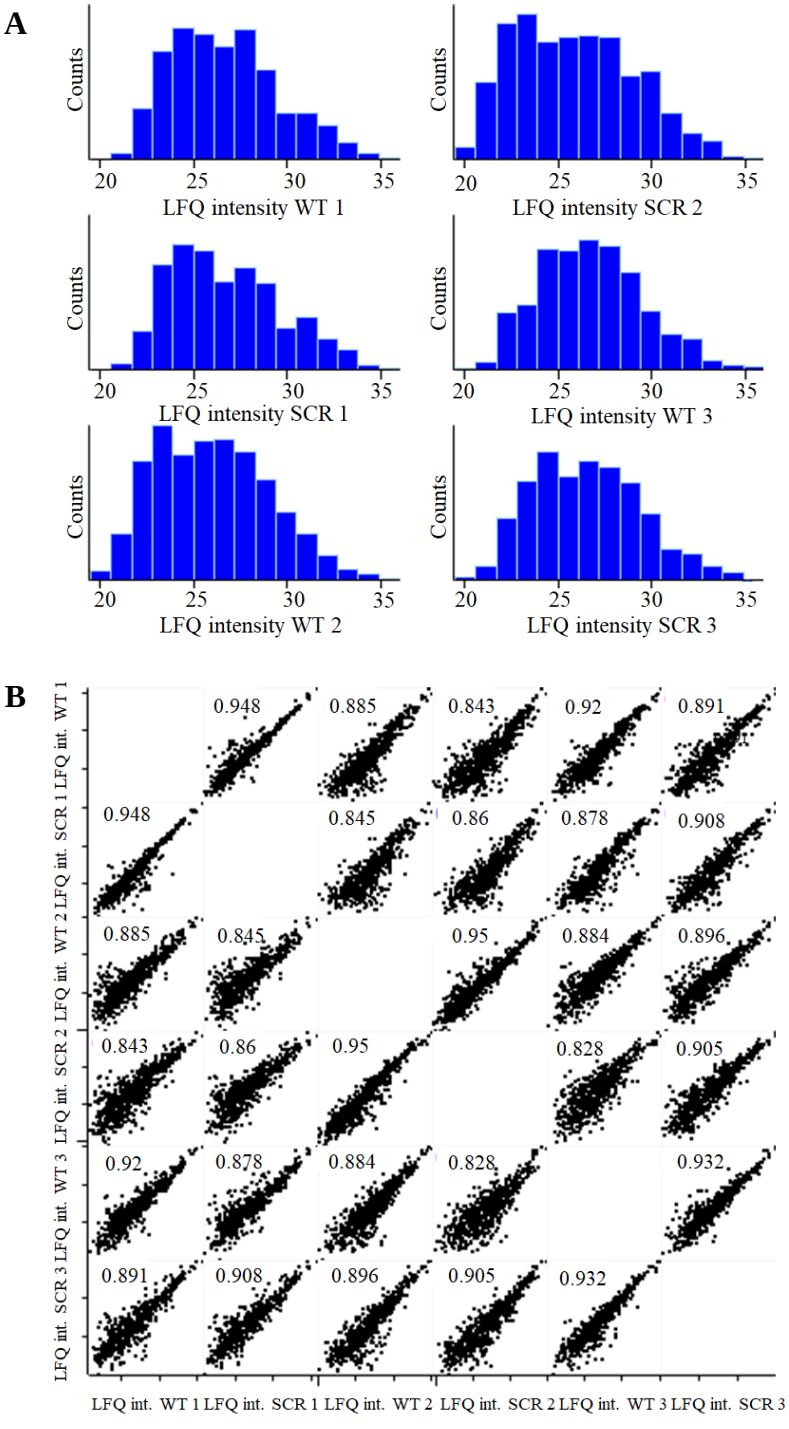


Figure 30: LFQ intensities of the identified proteins in the label-free DNA pulldowns. (A) Distributions of proteins after log_2-transformation. Depicted are the separate distributions of the wildtype replicates (WT 1, 2, 3) and their corresponding control (SCR 1, 2, 3) experiments. (B) Correlation of the protein intensities of the same WT and SCR pulldown replicates, indicated with the respective Pearson correlation coefficients.

Correlation of protein intensities in different samples

Another aspect for testing reproducibility of experiments or identifying potential outliers is provided by the correlation between samples of biological triplicates and samples of wildtype correlating to control. Therefore, protein LFQ intensities from the different pulldowns are plotted against each other and visualized in a multi scatter plot. Pearson coefficient can be determined as a means of correlation. Since the pulldowns are expected to exhibit similar protein patterns identified, deriving from background binders with only marginal numbers of interacting proteins being enriched, correlation should basically be high. dHL-60 affirmatively displayed substantial correlation, with Pearson coefficients ranging from 0.82-0.95, respective WT to SCR sample thereby obviously gave higher coefficients (0.828-0.95) than WT to WT (0.86-0.92) sample of different biological replicates. Inferior correlation between samples on the other hand would point at alterations in sample processing steps, being indicative for incongruent experimental performance.

Missing value imputation

As already mentioned in section 1.3.3.2, proteins of low abundance or those falling below the detection limit can benefit from missing value imputation, as this process includes those into the statistical data analysis, rather than rejecting them in the first place. Imputation of randomized values should simulate LFQ values of low abundance within the respective intensity distribution of each sample and therefore should not form a separate distribution. Imputed values are calculated from the normal distribution achieved by narrowing its width and height in dependence of its standard deviation and again can be visualized in a histogram. Figure 31 depicts imputational distribution for the dHL-60 pulldown for two differently set parameters of width and downshift, presenting both functional and improper imputation, the latter leading to the formation of a separate distribution of imputed values (for SCR 1, WT 3 and SCR 3), and thereby biased outcome of experimental data and contorted identification of significant interactors.

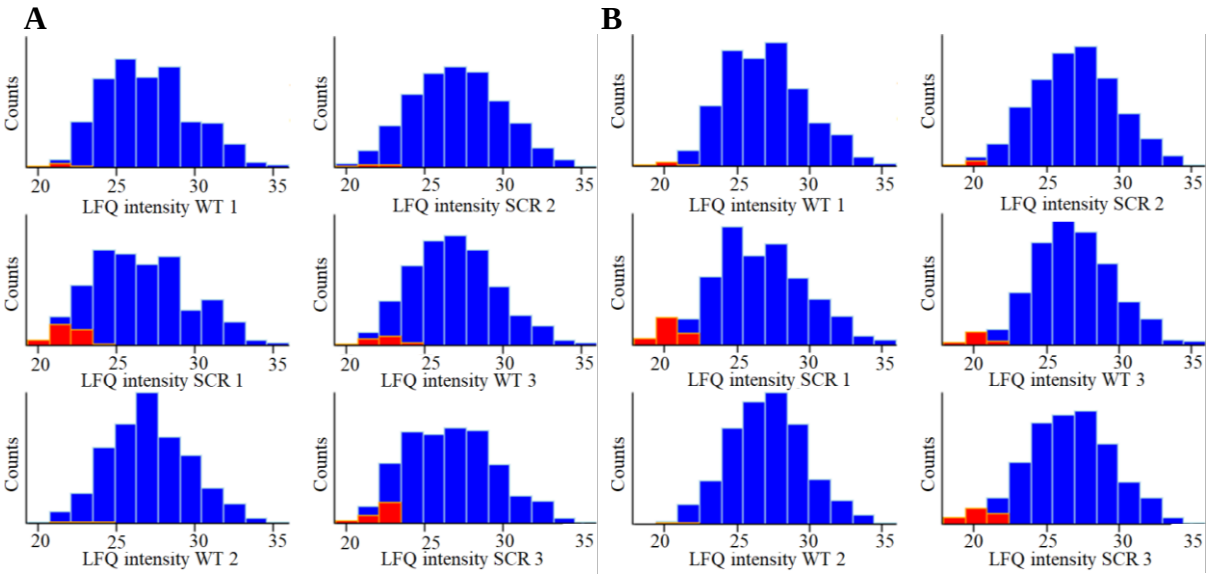


Figure 31: Statistical data analysis: Missing value imputation. Separate LFQ intensity distributions of the identified proteins in the label-free DNA pulldowns after log_2-transformation. Depicted are the wildtype replicates (WT 1, 2, 3) and their corresponding control (SCR 1, 2, 3) experiments. (A) Proper imputation (width 0.3, down shift 1.8) and (B) improper imputation (width 0.3, down shift 2.5). Imputed value distribution is given in red.

Visualization of proteins in volcano plots

After performing t-test based statistical analysis, proteins are generally visualized in volcano plots for identifying significant interactors. The overall representation of data in these plots contributes additionally to evaluating a functional pulldown assay. First, theoretically no significantly enriched proteins should be found in the upper left corner, as those basically interact with the SCR-version of the ALOX5 promoter. Thus, absence of false positives is indicative for choosing a suitable control sequence. However, due to the unintentional generation of new binding sites and the basic affinity of numerous proteins for nucleic acids, this feature cannot be guaranteed in all cases. Furthermore, these proteins do not affect or hinder the actual identification of truly interacting proteins of the wildtype sample and are therefore negligible. Second, background binders optimally form a narrow distribution around the zero value, thereby verifying reproducible experimental performances. Third, significance for the identification of interacting proteins depends on a cutoff curve, which can be adjusted to individual needs of the pulldowns carried out. This threshold is based on the minimum fold change obtained (s0 value) and a given FDR. Both of the parameters contribute to setting the threshold by altering the curvature of the line when being changed. In case of the dHL-60 pulldown, s0 and FDR were set to 0.75 and 0.05, respectively. While setting these values, one

has to ponder the stringency and plausibility with which significant interactors are thereby identified. Changing the values to higher ones for example (FDR = 0.25, s0 = 1) results in additionally identified possible interactors, mainly due to less stringent conditions determined by the FDR settings. These additionally found proteins enlarge the ones found with former settings to 32 proteins in total (in contrast to 20 proteins in the former pulldown). However, most of their respective protein classes do not exhibit a potential role in transcriptional regulation. These include particularly DNA repair and replication proteins, along with unspecifically interacting single-stranded nucleic acid-binding proteins, which altogether diminish the share of plausible interactors involved in transcriptional regulation. When changing the settings to lower values (FDR = 0.025, s0 = 0.4) most of the proteins are lost, with only 5 remaining, which however solely represent transcriptional regulators. Thus, altering these parameters can contribute to different outcomes of identification of significant interactors, which mostly has to be evaluated based on already known interactors or plausibility of the identified proteins. Referring to the published literature, s0 and FDR values generally range from 0.5-3 and 0.001-0.1, respectively. Since this method provides putative interactors altogether, the liable characterization and functionality of each individual protein has to be tested and confirmed in another way.

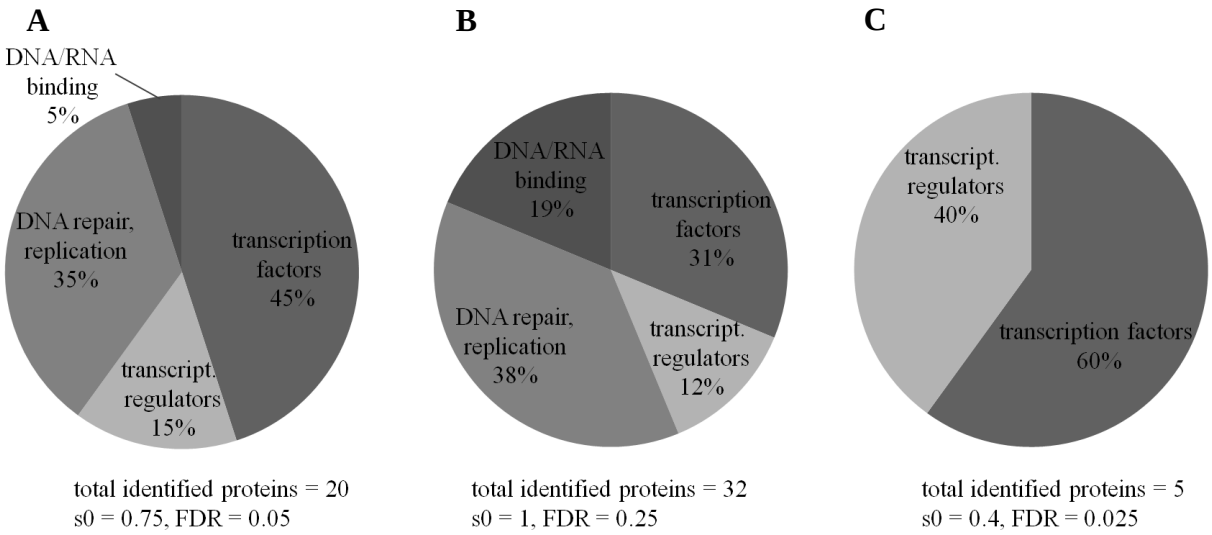


Figure 32: Classification of proteins depending on variably chosen threshold values for statistical analysis and their respective share of total identified proteins for (A) standard, (B) higher and (C) lower values. Classification is based on entries of the Uniprot Database.

LFQ intensities of individual proteins

Taking a closer look at the identified interactors and their respective LFQ intensity values for the individual WT and SCR pulldowns can give suggestions as to whether these proteins might be significantly interacting or not. This is partly based on whether reproducibility of intensities achieved in their respective 6 pulldowns. As shown in figure 33, some of the proteins found in the dHL-60 pulldown do not occur consistently throughout all of the performed experiments, which alters their significance in being identified as interactor.

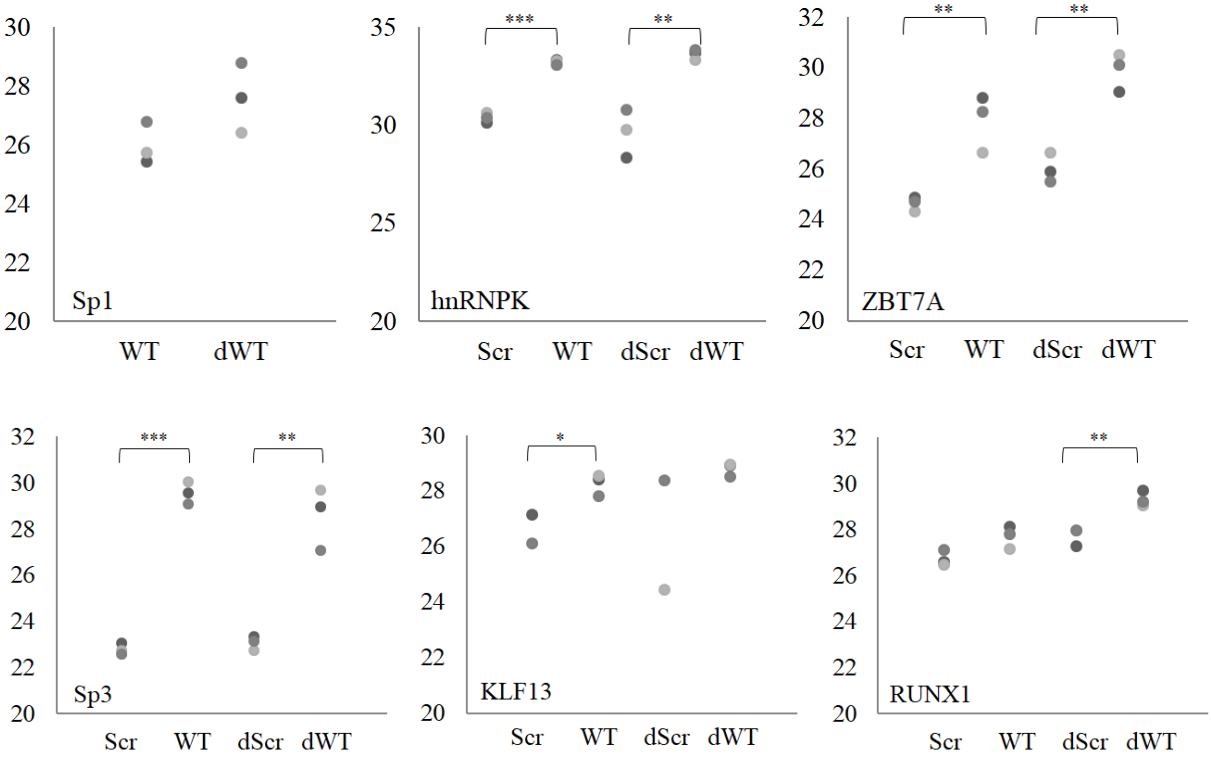


*Figure 33: LFQ intensities of individual pulldowns of selected proteins for wildtype (WT) or differentiated wildtype (dWT) and corresponding control (SCR, dSCR) experiments. $p \leq 0,05$ *, $p < 0,01$ **, $p < 0,001$ *** significance of performed t-test.*

When comparing HL-60 and dHL-60 pulldowns for known interacting transcription factor Sp1, hnRNP K or putative TF ZBT7A for example, which all were significantly identified in both experiments, the values of dHL-60 display similar to even higher variance in their log2-transformed intensities than the ones of the HL-60 cells, however still higher overall difference in intensities, suggesting that both variance and absolute value contribute to exceeding significance thresholds. In contrast, the narrow distribution and difference in intensity of Sp3 or KLF13 seems clear for HL-60 for being significantly interacting, whereas variance of wildtype and control pulldowns of dHL-60 is quite high, although the difference

of intensities for WT and SCR is similar to the HL-60 case, resulting in the TFs not being significant in the dHL-60 pulldown. On the other hand, a narrow distribution of single intensities will not necessarily lead to significance, if wildtype and control intensities do not differ sufficiently from one another, as seen by RUNX1, which is only interacting in dHL-60. Thus, obtaining significance seems to depend, among other features, on intensity values of single pulldown replicates and their respective variance, as well as their total difference in WT and CTR sample.

As already mentioned, these parameters do not provide exact proof for a functional experimental setup and data analysis, but optimally they allow to accept or reject prior assumptions made for the following data analysis. All of these processing steps were carried out for each of the used cell lines after pulldowns were performed, thereby testing their plausibility for analysis and interpretation of visualized data.

4.3.2.4 DNA pulldowns with core-promoter sequences

After establishing the general conditions of the pulldowns to be performed, additional to HL-60 and dHL-60, experiments were carried out with THP-1 and dTHP-1, MM6 and dMM6, as well as Rec-1 and BL-41. All of the pulldowns were performed with three subsequent cell passages, serving as biological triplicates. The overall identification of significant interactors generated a list of 66 proteins (suppl. table 1), deriving from all the different cell lines and cell states cumulatively. Search parameters were set to accept proteins identified with at least one unique peptide. To increase stringency of identification of interactors, parameters were afterwards set to identify at least one unique peptide in a minimum of three peptides identified per protein. Significant interactors falling below this threshold are shown in light grey in the volcano plots of the respective pulldowns.

The myeloid cell lineage generated a higher amount of enriched proteins than the B-lymphocytes, still they also differ in their number of identified interactors. Since cell maturation of HL-60, THP-1 and MM6 increases in this order [319, 320], a trend can be ascertained of less proteins being detected with higher maturation state achieved. Promyelocytes HL-60, which can either differentiate into granulocytes or monocytes/macrophage-like cells depending on the differentiation agent used [319], therefore display the highest number of interactors identified. Thus, cell maturation may influence the number of proteins being involved in transcriptional regulation, whereas cell differentiation status does not seem to alter the amount of enriched proteins, as judged by the comparison of

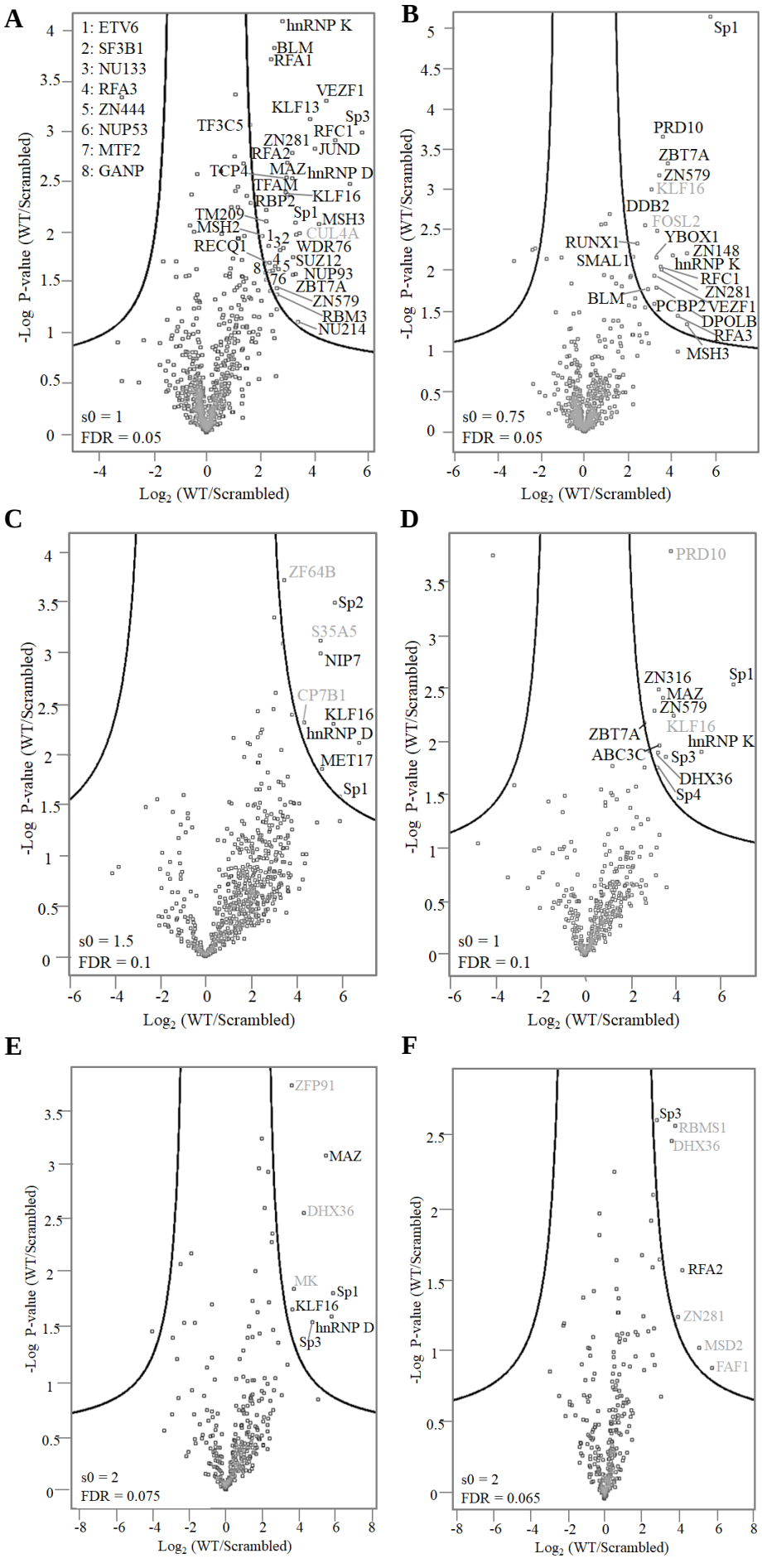

A
1: ETV6
2: SF3B1
3: NU133
4: RFA3
5: ZN444
6: NUP53
7: MTF2
8: GANP
hnRNP K
BLM
RFA1
VEZF1
KLF13
Sp3
TF3C5
RFC1
ZN281
JUND
RFA2
TCP4
MAZ
hnRNP D
TFAM
KLF16
RBP2
Sp1
TM209
MSH3
MSH2
CUL4A
RECQ1
WDR76
SUZ12
NUP93
ZBT7A
ZN579
RBM3
NU214
s0 = 1
FDR = 0.05
-Log P-value (WT/Scrambled)
Log2 (WT/Scrambled)
B
Sp1
PRD10
ZBT7A
ZN579
KLF16
DDB2
FOSL2
YBOX1
RUNX1
ZN148
SMAL1
hnRNP K
RFC1
ZN281
BLM
PCBP2
VEZF1
DPOLB
RFA3
MSH3
s0 = 0.75
FDR = 0.05
C
ZF64B
Sp2
S35A5
NIP7
CP7B1
KLF16
hnRNP D
MET17
Sp1
s0 = 1.5
FDR = 0.1
D
PRD10
ZN316
Sp1
MAZ
ZN579
KLF16
ZBT7A
hnRNP K
ABC3C
Sp3
DHX36
Sp4
s0 = 1
FDR = 0.1
E
ZFP91
MAZ
DHX36
MK
Sp1
KLF16
hnRNP D
Sp3
s0 = 2
FDR = 0.075
F
Sp3
RBMS1
DHX36
RFA2
ZN281
MSD2
FAF1
s0 = 2
FDR = 0.065

Figure 34: Label-free DNA pulldowns with ALOX5 core-promoter containing sequence in myeloid cell lines. Volcano plots represent specific interactors enriched in their upper right corner; the respective threshold values of each individual pulldown are given in the plots. LFQ intensities of WT and SCR samples (x-axis, log_2-transformed) are plotted against the respective p-values (y-axis, log_{10}-transformed) of the identified proteins. Volcano plots are given for (A) HL-60, (B) dHL-60, (C) THP-1, (D) dTHP-1, (E) MM6 and (F) dMM6.

the myeloid cell lines in both of their respective differentiation states. However, by comparing the protein patterns, no major difference can be deduced from the performed DNA pulldowns. The myeloid cell lines all share similar protein compositions, which repeatedly contain transcription factors of the Sp/KLF-family, as well as hnRNPs. Further proteins recurring in more than one cell line include TFs MAZ, ZBT7A and ZN281, as well as transcriptional regulators PRD10, ZN579 and the helicase DHX36. The results obtained for both B-cell lines show good correlation to those achieved for the myeloid cell lineage and contribute to the overall protein pattern consisting of Sp/KLF-family transcription factors. Ribonucleoproteins could likewise be detected, suggesting similar regulatory machinery involved in all of the cell lines used. Only a single protein (F120A) could exclusively be identified in the B-lymphocytic cell lines, which however does not exhibit gene regulatory functions.

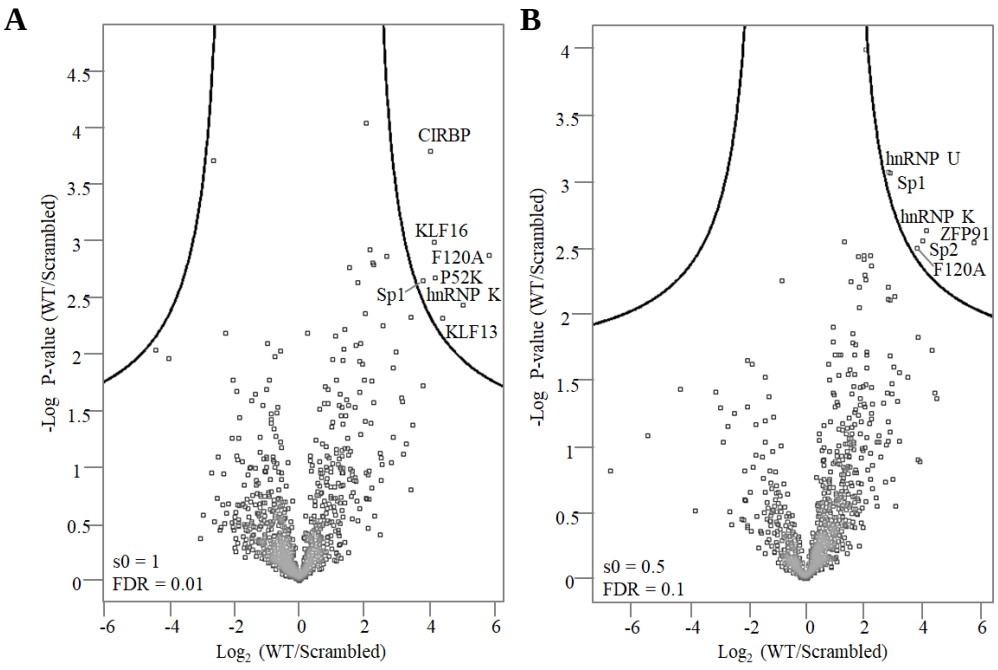


Figure 35: Label-free DNA pulldowns with ALOX5 core-promoter containing sequence in B-cell lines (A) Rec-1 and (B) BL-41. Analysis was done according to the residual DNA pulldowns.

The overall identification revealed several protein classes, as depicted in figure 36A. More than one third exclusively contributes to transcriptional regulation in the first place (transcription factors and transcriptional regulators, which are partly not characterized any further). 5% belong to hnRNPs that only comprised three different proteins, hence a marginal percentage share, however they were detected in all cell lines used. A rather large part of 11% is provided by proteins involved in replication or DNA repair, which, considering the actual experimental setup appears to be logical and inevitable. Another protein class of interest, as these are involved in recognition of secondary DNA structures is represented by helicases enriched in the myeloid cell lines. Accordingly, more than one third of the identified interactors (mainly TFs) display zinc fingers as their active site or DNA binding domain, respectively (figure 36B). In this respect, four different zinc finger types could be detected, however the major part is provided by the C_2H_2-type. Only a small number of transcription factors contained other DNA binding domains, such as the RUNT or ETS domain. Various proteins, including the helicases or DNA repair proteins, additionally are ATP-dependent or require metal ions for functionality.

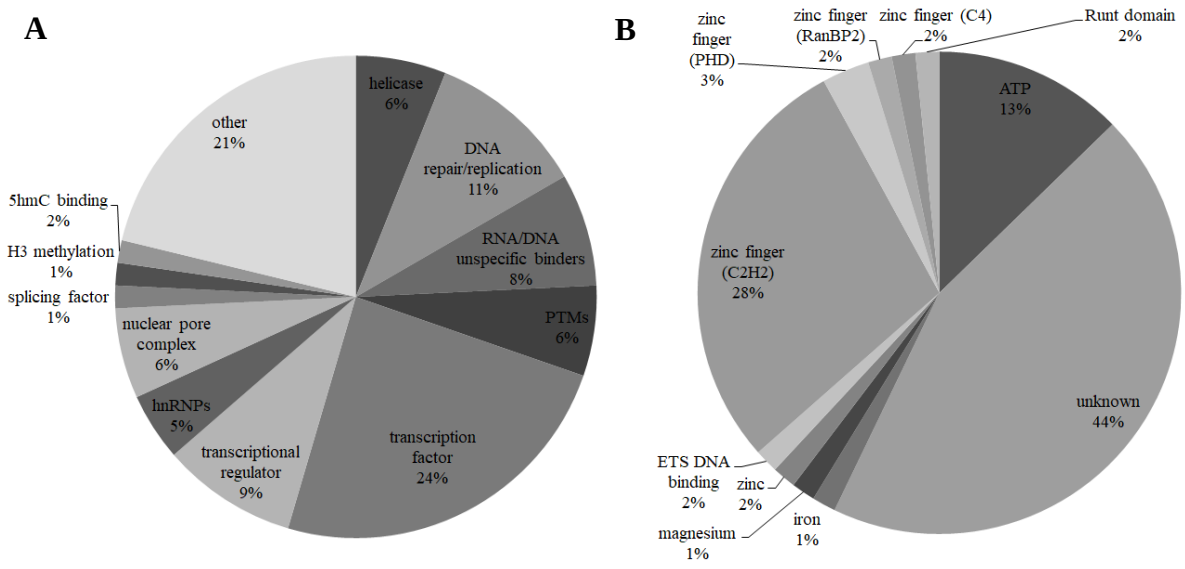


Figure 36: Classification of identified proteins in all pulldown experiments, according to their (A) protein classes and (B) active or functional binding site. Classification is based on data from Uniprot database.

protein	HL-60	dHL-60	THP-1	dTHP-1	MM6	dMM6	Rec-1	BL-41	unique peptides	function
BLM	x	x							26+17	DNA helicase, G4-DNA preferred substrate for unwinding [131]
DHX36				x	x	x			3+2+1	ATP-dependent helicase, unwinding of G4-DNA/RNA [136]
hnRNP D	x	x	x						9+24+6	transcription factor with DNA/RNA-binding domains, binds G4-DNA in human telomeres [115]
hnRNP K	x	x		x			x	x	46+45+16+65+68	transcription factor, both activator and repressor, interacts with G4-DNA [122]
KLF13	x						x		4+6	transcription factor, activates expression of RANTES in T-cells [321]
KLF16	x	x	x	x	x		x		2+1+8+1+5+6	transcription factor, competes with Sp1 for binding sites
MAZ	x			x	x				9+2+3	transcription factor with several functions in G4-DNA recognition
PRDM10		x		x					4+2	transcriptional regulator
Sp1	x	x	x	x	x		x	x	7+5+9+10+5+17+18	activating transcription factor
Sp2			x					x	7+8	transcription factor
Sp3	x			x	x	x			8+11+7+1	transcription factor
VEZF1	x	x							9+10	transcriptional regulator
ZBTB7A	x	x		x					6+22+5	transcription repressor
ZNF281	x	x				x			14+29+2	transcription repressor
ZNF579	x	x		x					3+5+8	transcriptional regulator

Table 8: Summary of recurrently enriched proteins (given by their gene name) in DNA pulldowns of all cell lines after filtering for selected criteria.

Altogether, a large share of almost 50% of the classified proteins was enriched repeatedly in the performed DNA pulldowns. To decipher the interactors most prominently found, protein identification criteria were set to accomplish at least two independent quantification events in the different experiments, as well as exhibiting a clear functional contribution in gene regulation. This resulted in a list of 15 recurring proteins depicted in table 8. In summary, the obtained results suggest similar protein patterns involved in the regulation of the proximal ALOX5 promoter, regardless of cell line or differentiation state.

4.3.2.5 DNA pulldowns with core-promoter proximal sequences

In order to enlarge the proximal promoter span covered and identify further interacting proteins, additional DNA pulldowns were carried out using oligonucleotide sequences encompassing 80 base pairs, further upstream of the sequences already used. These DNA stretches either comprised a vitamin D_3 response element (-408 to -329 in relation to ATG) or a SMAD binding element (-408 to -329 and -334 to -255 in relation to ATG, respectively), thus providing binding sites for vitamin D_3 itself and TGFβ effector proteins, respectively. Sequences were designed to contain an overlap of 6 bps to the preceding sequence, in order to not separate consensus sites, which might otherwise be present at the interface of two DNA stretches (figure 37). Promyelocytes HL-60 and dHL-60 were used as model cell line, as these probably contribute to identifying a higher number of proteins than the more mature myeloid cell lines (see section 4.3.2.4).

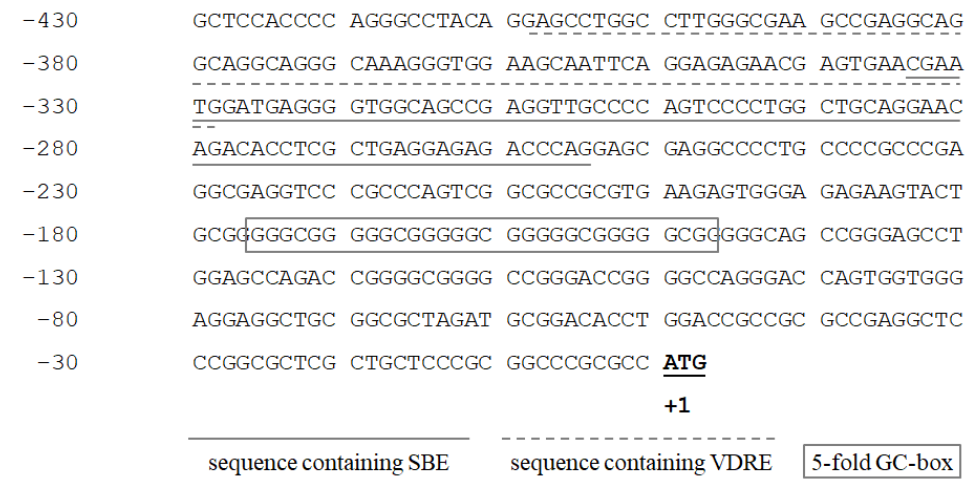


Figure 37: DNA sequences of core-promoter proximal sequences. The 5-fold GC-box is given for reference.

Both of the performed HL-60 pulldowns generated more significantly enriched proteins than the respective dHL-60 pulldowns, the overall distributions were wider for both

pulldowns carried out with the SBE containing sequences. Furthermore, a higher number of proteins interacting with the control oligonucleotides could be observed, compared to the 120-mer wildtype and control pulldowns, suggesting possible generation of binding sites while scrambling the respective sequences. Control oligonucleotides of both 80-mer pulldowns were accordingly repeated using different scrambled sequences with similar outcome. As already stated in section 4.3.2.3, these proteins do not cumber identification of wildtype interacting proteins and represent a share ranging from 1.1-2.9% of all identified proteins in the respective pulldowns. Judging from available literature for label-free quantification, this percentage share is estimated to be higher in some of the investigations, if presented after all.

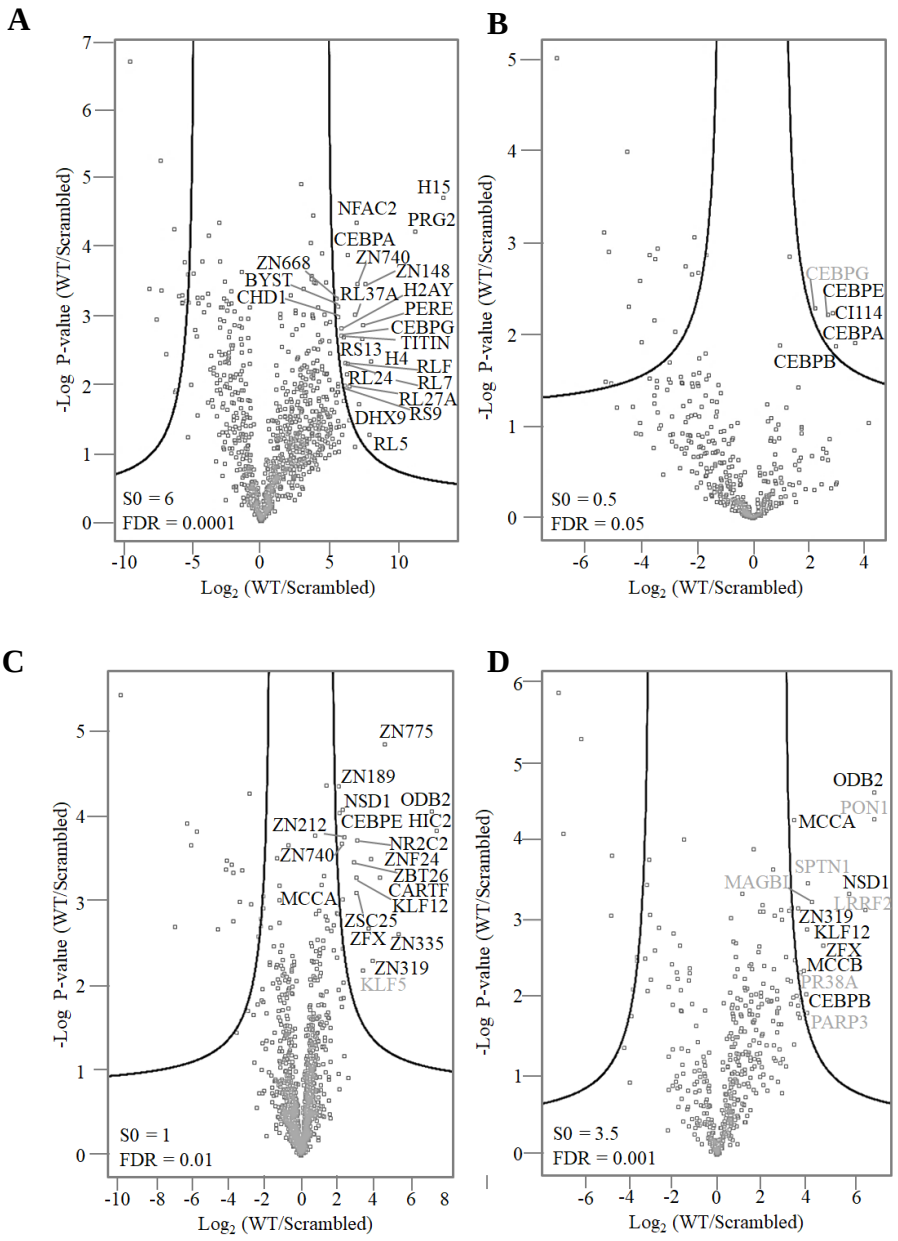


Figure 38: DNA pulldowns of the ALOX5 promoter proximal sequences. Label-free quantification and experimental setup was performed analogously to the former pulldowns. Log_2-transformed LFQ intensities of the individual proteins are plotted against their log-transformed p-value of both WT and SCR sample. Oligonucleotide sequences used comprised either the Smad binding element in (A) HL-60 and (B) dHL-60 or the VDRE in (C) HL-60 and (D) dHL-60.

Remarkably, oligonucleotide sequences comprising the SMAD binding elements enriched different transcription factors of the CCAAT/enhancer binding proteins (C/EBPs) in both differentiation states. This TF family comprises six proteins (C/EBPα-ζ), which are generally involved in cell proliferation and differentiation, as well as inflammatory responses [322]. Most of the identified interactors in dHL-60 cells comprise these proteins, suggesting their strong participation in transcriptional regulation of the ALOX5 gene. Pulldowns performed with undifferentiated HL-60 likewise identified two of the C/EBPs, additional to certain zinc finger proteins, as well as nuclear factor of activated T-cells NFAC2.

DNA sequences containing the VDRE also enriched members of the C/EBP family of transcription factors, additional to further TFs of the KLF-family and transcriptional zinc finger proteins. Interestingly, three proteins coexisted in both undifferentiated and differentiated cells, including TF KLF12, probable transcriptional activator ZFX and histone methyltransferase NSD1.

Some of these pulldowns furthermore comprised several ribosomal or mitochondrial proteins identified as interactors, which clearly are not involved in transcriptional regulation, hence their marking is in light gray. An overview of the identified interactors with functionality in gene regulation and identification parameters set as mentioned above (at least three peptides of which one has to be unique) is given in table 9. For the representation of all other identified proteins see suppl. table 2.

protein	function	gene
80-mer containing SMAD-binding element		
HL-60		
NFAC2	nuclear factor of activated T-cells, involved in inducible expression of cytokine genes in T-cells	NFATC2
ZN148	zinc finger protein, transcriptional regulator	ZNF148
HL-60 and dHL-60		
C/EBP α	TF, coordinating proliferation arrest and differentiation of myeloid progenitors among others	CEBPA
dHL-60		
C/EBP β	TF regulating expression of genes involved in immune and inflammatory responses	CEBPB
C/EBP ε	transcriptional activator, required for promyelocyte-myelocyte transition in myeloid differentiation	CEBPE
80-mer containing VD-response element		
HL-60		
nucl. receptor subfam. 2 group C member 2	orphan nuclear receptor, transcriptional repressor or activator; repressor of nuclear receptor signaling pathways such as RXR or VDR	NR2C2
Zinc finger protein 24	TF, involved in maintenance in progenitor stage by promoting cell cycle	ZNF24
Ca-responsive TF	transcriptional activator	CARF
Krüppel-like TF 5	transcriptional activator	KLF5
ZN335	zinc finger protein, may regulate transcription by recruiting histone methyltransferase complexes	ZNF335
HL-60 and dHL-60		
histone methyltransferase	H3 lysine-36 and H4 lysine-20 specific, can both negatively or positively influence transcription	NSD1
Krüppel-like TF 12	transcriptional repressor	KLF12
zinc finger X-chromosomal protein	probable transcriptional activator	ZFX

Table 9: Selection of proteins identified as proximal ALOX5 promoter interactors in HL-60 and dHL-60 cells. Classification is based on UniProt database entries

For the most relevant transcription factors identified in all DNA pulldowns, confirmatory evidence is provided by Chip-Seq datasets, as shown in figure 39. These published datasets corroborate binding of the depicted proteins to the ALOX5 promoter sequence in different cell lines and enhance the proteins' potential role in ALOX5 gene expression.

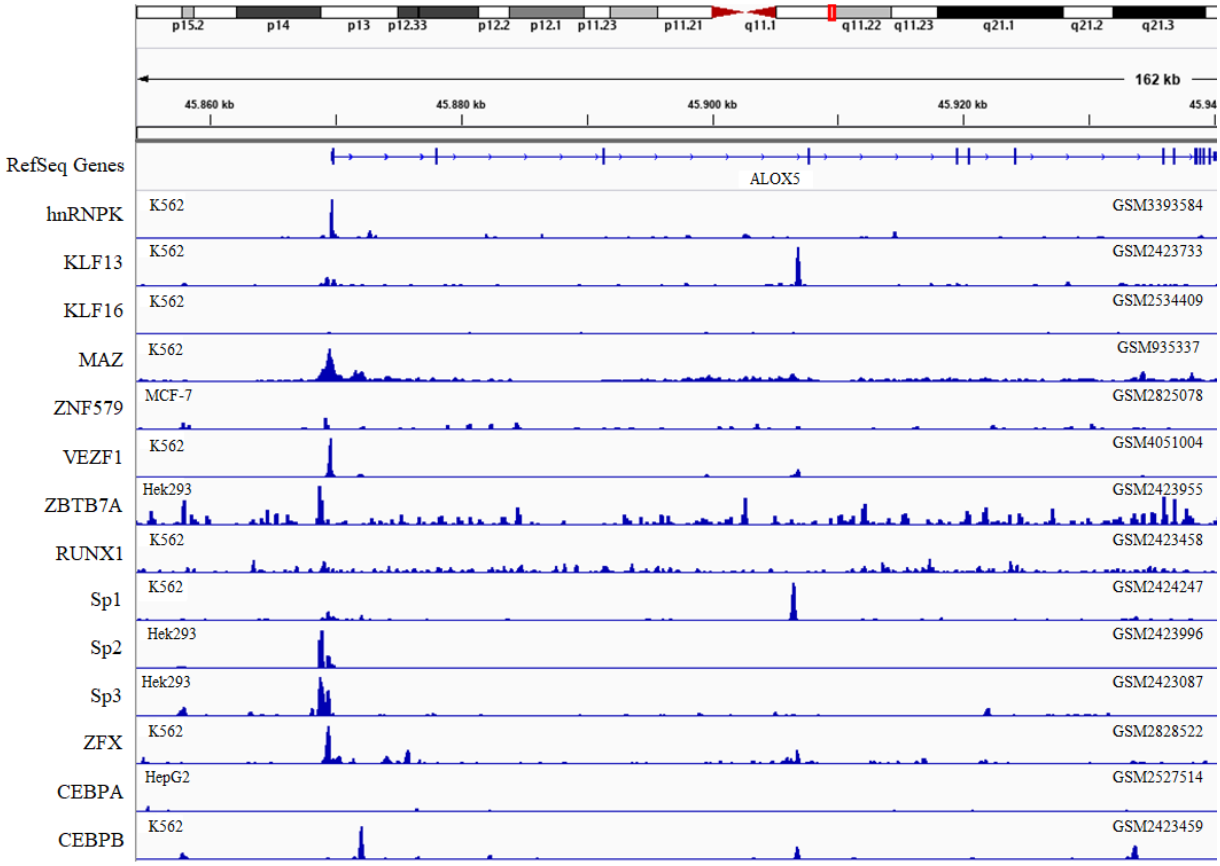


Figure 39: GEO ChIP-Seq data, visualized by IGV viewer. The depicted transcription factors are confirmed to bind to the ALOX5 gene promoter and/or downstream binding sites within gene introns.

4.4 Secondary DNA structures

4.4.1 G-quadruplex-associated proteins enriched in DNA pulldowns

Since several of the identified proteins found in the performed pulldowns encompassing the tandem GC-box (120-mer wildtype of core promoter and control) execute functions associated with G4-DNA recognition or their unwinding/stabilization, taking a closer look at the respective interactors was started in order to elucidate their potential role in ALOX5 gene regulation in correlation with secondary DNA structures.

As already stated in section 1.2.2.3, there exists a correlation between Sp1 binding sites and an elevated feasibility of forming quadruplex structures. Consequentially, the basal regulation by Sp1 additional to the present GC-boxes in the proximal ALOX5 promoter sequence supposedly promote the possibility of developing secondary structures of that kind. Especially the tandem GC-box represents a good candidate for the potential formation of G4-DNA. Furthermore, proof is provided for the involvement of certain enriched interactors with quadruplex unwinding capability (hnRNP D, MAZ and the specific resolving helicases BLM and DHX36), as well as stabilizing properties, which include hnRNP K for stabilization of the quadruplex's C-rich complementary strand and MAZ. A previous study additionally revealed a shared amino acid composition similar and exclusively found between known quadruplex interacting proteins, which includes the aforementioned factors, as well as YBOX1 that also could be associated with the presence of G4-DNA before [316, 323]. In this context, it seemed obvious to pursue the possible formation of G-quadruplex structures within the proximal ALOX5 promoter sequence.

protein	HL-60	dHL-60	THP-1	dTHP-1	MM6	dMM6	Rec-1	BL-41
BLM	x	x						
DHX36				x	x	x		
MAZ	x			x	x			
hnRNP D	x		x		x			
hnRNP K	x	x		x			x	x
YBX1		x						

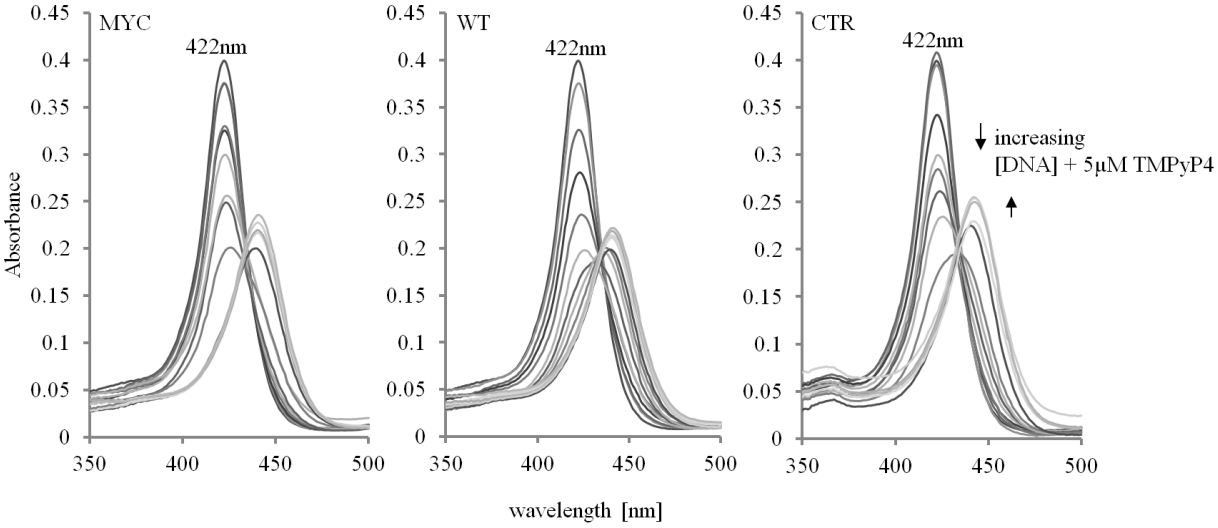
Table 10: Expression pattern of G-quadruplex interacting proteins in all cell lines used. Proteins are given by their gene names.

4.4.2 *In vitro* analysis for G-quadruplex structures in the ALOX5 promoter

4.4.2.1 Steady-state UV-VIS-spectroscopy

In order to investigate the putative formation of G4-DNA in the proximal ALOX5 promoter, interactional behavior of a combination of G-quadruplex stabilizing agent TMPyP4 and a 46-mer containing the tandem GC-box was monitored by UV-VIS-spectroscopy. TMPyP4 is a common agent for either treatment of cells for stabilizing G4-DNA or performing spectroscopic studies *in vitro* (see section 1.2.2.5). It is capable of intercalating with or stacking upon the formed G-tetrads, thereby hampering their unwinding. The extent of interaction of both stabilizing agent and quadruplex-containing oligonucleotide contributes to the generation of a more or less stabile construct, which in case of strong G4-TMPyP4 interaction generates a remarkable and sharp isosbestic point in the respective UV-VIS-spectra. Furthermore, the Soret band absorption spectra tend to shift in hypochromicity depending on the interaction of both partners to higher (for strong interaction) or lower (for light interaction) extent.

The used wildtype 46-mer, comprising the tandem GC-boxes of the ALOX5 gene was compared to a sequence deriving from the c-myc promoter, known to form quadruplex structures, serving as positive control, as well as a random 22-mer of oligonucleotides, unable to develop G4-DNA, which served as negative control. All oligonucleotides were annealed in annealing buffer containing 100mM of potassium ions to ensure proper folding of the quadruplex structures. TMPyP4 was diluted with differing concentrations of the three single-stranded oligonucleotides until no further change of absorption spectra could be observed upon three successive additions of DNA sequences.



oligonucleotide	sequence	Shift Soret band [nm]	Hypochromicity %
WT	[BIO]-GATCCTGCGG(GGGCGG)$_5$CAG	21	54
MYC	[BIO]-TGAGGGTGGGTAGGGTGGGTAA	18	49.1
CTR	GCTCGCCCCGCCCCGATCGAAT	20	39.4

Figure 40: In vitro analysis of G-quadruplex formation: Steady-state UV-VIS-spectroscopy. The quadruplex stabilizer TMPyP4 (5 µM) was incubated with increasing concentrations of the chosen oligonucleotides WT, MYC and CTR and absorbance (arbitrary units) was recorded in a range of 350-500 nm. Absorption spectra shown are average spectra of three experiments after correction for solvent effects. Used oligonucleotides and their respective sequences, induced Soret band shifts and hypochromicities are given in the corresponding table.

As depicted in figure 40, all of the used oligonucleotide sequences displayed interaction with the quadruplex stabilizing agents, although to different extent. The free TMPyP4 exhibited its absorbance maximum at 422 nm. The amount of absorbance was downshifted with raised amounts of oligonucleotide added, until an isosbestic point was reached for all sequences. Afterwards, absorbance values increased again, until no further change in absorbance was detectable. Isosbestic points achieved ranged in all cases from 431-435 nm, thereby giving evidence for an adjusted equilibrium between TMPyP4 and oligonucleotide. The Soret band of the porphyrin analogon experienced a bathochromic shift towards higher wavelengths, which ranged from 18 nm to 20 nm to 21 nm for the positive control, negative control and wildtype sequence, respectively. Since all of these parameters exhibited similar characteristics, the major criterion for distinguishing the extent of interaction was provided by the downshift of absorbance of the Soret band, which is given by the percentage of

hypochromicity. Positive control c-MYC resulted in a downshift of absorbance of 49.1%, indicating a strong interaction of quadruplex DNA and TMPyP4. The negative control sequence displayed 39.4% of hypochromicity, which is in good agreement to published data investigating the effects of different secondary DNA structures on absorption spectra of TMPyP4 [145]. Intriguingly, the wildtype sample exhibited a decrease in absorbance of 54%, thereby even exceeding hypochromicity of the positive control. Evidently, the WT-DNA sequence is capable of forming a strong complex with the porphyrin, which therefore is more effectively interacting with it than with any of the other samples. Judging from these results obtained, there is high evidence of G-quadruplex structures formed in the wildtype sequence comprising the tandem GC-box of the ALOX5 proximal promoter.

4.4.2.2 CD spectroscopy

For providing further evidence for G-quadruplex formation by a direct spectroscopic method, CD spectra were recorded (figure 41). Spectra were recorded for all of the oligonucleotides used in both UV-VIS spectroscopy and ELISA. The control and mutant sequence both displayed spectra with a maximum positive peak at about 280 nm, whereas the positive control c-MYC and the ALOX5 promoter sequence exhibited a positive peak at 262 nm and 266 nm and a negative at 241 nm and 242 nm, respectively, which are characteristic for G-quadruplex formation [70]. Additionally, the latter two sequences displayed another G4-DNA characteristic positive maximum at a wavelength of about 210 nm. The ALOX5 promoter fragment furthermore showed a positive peak at 293 nm, pointing to a G-quadruplex topology, which does not derive from sole parallel or antiparallel G4-DNA formation (see also section 1.2.3). Altogether, the CD spectra revealed G-quadruplex formation in the ALOX5 core promoter sequence and corroborated the data obtained from UV-VIS spectroscopy.

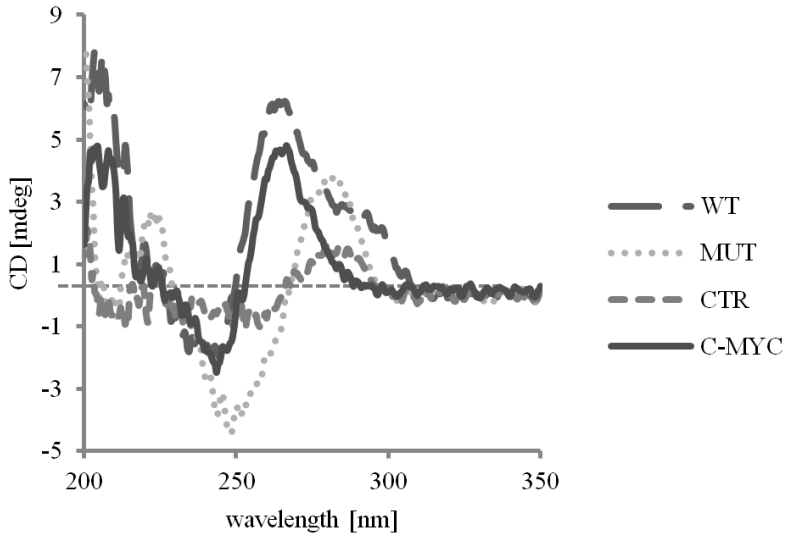
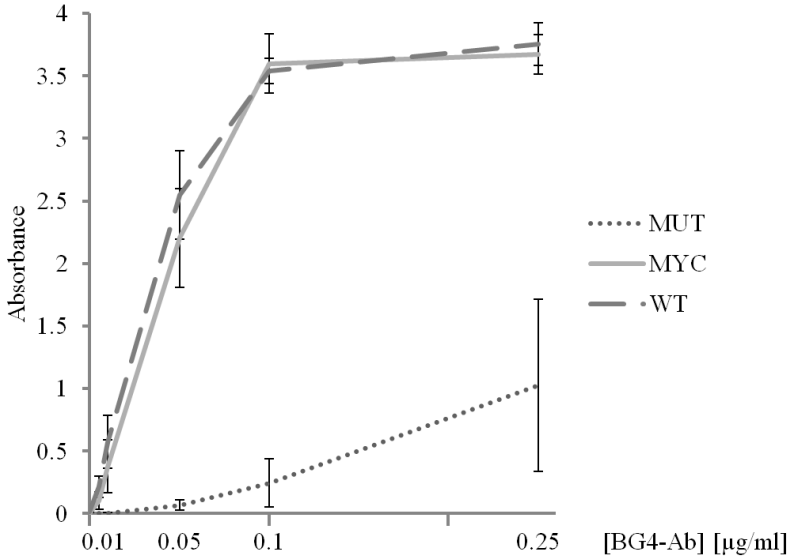


Figure 41: In vitro analysis of G-quadruplex formation: CD spectroscopy

4.4.2.3 Enzyme linked immunosorbent assay

In order to confirm the existence of *in vitro* formed G4-DNA in the ALOX5 gene, ELISAs were performed as an independent immune-based method. As already stated in section 1.2.3, there exist commercially available antibodies, which are able to detect G4-DNA quite specifically [152]. ELISAs were carried out, again utilizing the ALOX5 promoter wildtype 46-mer, containing the tandem GC-box, as well as the positive control used in the UV-VIS-spectroscopic experiments. The negative control consisted of a scrambled version of the WT sequence, likewise improbable of forming quadruplex structures. 50 nM of each of the oligonucleotides were immobilized in streptavidin-coated 96-well plates and subsequently washed and probed with varying concentrations of primary antibody. Secondary antibody used encompassed HRP-conjugate for consecutive spectroscopic detection.



oligonucleotide	sequence
WT	[BIO]-GATCCTGCGG(GGGCGG)$_5$CAG
MYC	[BIO]-TGAGGGTGGGTAGGGTGGGTAA
MUT	[BIO]-GATCCTG(CGAATT)5CAATTCAG

Figure 42: In vitro analysis of G-quadruplex formation: ELISA. Immobilized WT, c-MYC and CTR oligonucleotides were incubated with differing concentrations of anti-G4-DNA-ab, BG4. The secondary HRP-antibody was then added, followed by TMB substrate. Absorbance (arbitrary units) was measured at 650 nm. n = 3, mean ± s.d.

As shown in figure 42, absorbance values of the WT and c-MYC DNA sequences exhibited good correlation to one another, whereas the negative control did not display a high affinity for the G4-DNA antibody. The latter obviously did show some affinity to the negative control in higher concentrations, however to a much lower extent than when being used for the WT and positive control sample. Thus, a small amount of unspecific binding could be detected. In contrast to that, WT and c-MYC displayed high affinity for the antibody, with spectra similarly congruent to one another, giving evidence for the formation and subsequent detection of G-quadruplexes present in both of the oligonucleotide sequences. Consequently, the ELISA did provide an alternative and additional method of giving proof for the *in vitro* existence of G4-DNA in the ALOX5 proximal promoter sequence.

4.4.3 *In cellulo* analysis for G-quadruplex structures in the ALOX5 promoter

4.4.3.1 Cell viability and cytotoxicity of G-quadruplex stabilizing agents

Since the *in vitro* data obtained pointed at the existence of G-quadruplex structures formed by the tandem GC-box, efforts were started to evaluate the importance of these secondary structures in cellular experiments. Therefore, quadruplex stabilizing agents TMPyP4 and pyridostatin were used to monitor their effects on 5-LO mRNA and protein expression in the model cell line HL-60. As already described in section 1.2.2.5, treatment with these agents in the micromolar range is commonly used to examine regulatory effects of G4-DNA on transcriptional level.

Both TMPyP4 and PDS (and respective analoga) are shown to exhibit cytotoxic effects on various different cancer cell lines, including for example leukemic K562, colon carcinoma cell line HCT8, fibrosarcoma cell line HT1080 or breast cancer cell line MCF7 [113, 142, 149, 324]. In order to examine cytotoxicity and impacts on cell proliferation of both agents on promyelocytes HL-60 in their differentiated state, cells were incubated with either 50 µM TMPyP4 or 10 µM PDS after differentiation and monitored for the respective parameters in periodical intervals of 24 h over three days. An untreated version of dHL-60 that was likewise cultured served as negative control. All cultures were seeded out at 0.2×10^6/ml for reference. Cells were both counted with an automatic cell counter and manually to determine the proliferation rate and calculate cell survival, respectively.

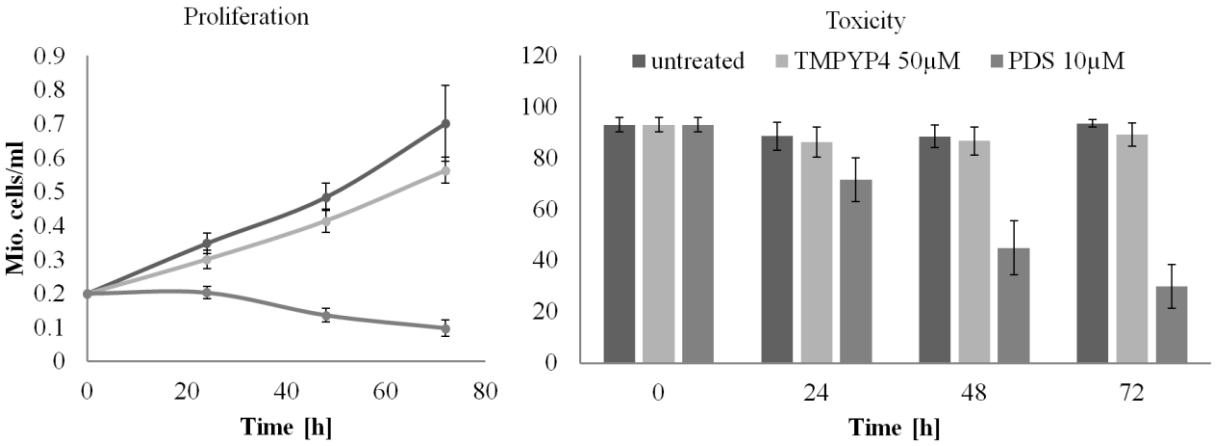


Figure 43: Influence of G-quadruplex stabilizing agents on dHL-60 cell viability. Proliferation rate (left) and the ratios of dead and living cells (right) were determined by manual cell counting after treatment with trypan blue staining solution. Depicted are three independent experiments with mean ± s.d.

Proliferation rate of untreated dHL-60 was higher than expected, considering a doubling time of around 40 hours for the undifferentiated state (according to Deutsche Sammlung von Mikroorganismen und Zellkulturen DSMZ). Differentiation technically can contribute to decelerated proliferation, however in this case it did not have a high impact on slowing down the proliferation rate. The TMPyP4 treated cells exhibited a related behavior, besides the fact that proliferation rate was slightly diminished, as compared to the untreated cells. Nonetheless, TMPyP4 treated dHL-60 displayed a sturdy proliferation, which was rather not expected after treatment with a cytotoxic agent. In contrast, proliferation of PDS treated cells did not even start after differentiation and culturing, as judged by the slanting curve obtained. In this case, dHL-60 retained a steady cell density for about 24 h before retardation of the overall proliferation could be observed, thereby confirming potent efficacy on cell viability of pyridostatin.

In terms of cell survival and cytotoxic effects on dHL-60, a comparable image was pictured for the untreated and treated cells. Untreated dHL-60 altogether maintained their share of living cells for over three days, fluctuating between 88-94% of all cells counted. Treatment with TMPypP4 did not have a high impact on cell survival and similarly ranged between 86-89%, with only marginal deterioration compared to the untreated cells. A different behavior was represented by PDS treated cells, whose share of living cells diminished quickly and significantly after three days of culturing. Cytotoxicity clearly could be verified, considering a starting cell survival share of 93%, which was reduced to only 45 and 30% after 48 and 72 hours, respectively.

4.4.3.2 5-LO mRNA expression after treatment with G4-DNA stabilizing agents

In order to monitor effects of G-quadruplex stabilizing agents on 5-LO mRNA expression levels *in cellulo*, RT-qPCR experiments were carried out after treating dHL-60 cells with TMPyP4 or pyridostatin. Untreated cells again served as control. Addition of differentiating agents TGFβ/1,25$(OH)_2D_3$ and G4-DNA stabilizing agents was simultaneously carried out and cells were subsequently incubated for two days, with samples being taken every 24 hours. Differentiation of HL-60 obviously was required to initiate 5-LO expression. After harvesting around 15 million cells per sample, RNA was extracted and transcribed into cDNA, which consecutively was used for qPCR, monitoring 5-LO and GAPDH expression levels, the latter serving for normalization. Respective gene expression levels were plotted in bar graphs according to their $2^{-\Delta\Delta Ct}$ values.

As depicted in figure 44, mRNA expression levels of the examined dHL-60 cells did not generate great distinctions in their different previous treatments. As expected, untreated cells experienced a time-dependent increase in 5-LO mRNA expression, with about twice as much being detected after 48 hours than 24 hours. Treatment with TMPyP4 provided miscellaneous results, given the fact that slightly less expression levels were achieved after 24 hours in comparison to the untreated cells, but mRNA levels after 48 hours even exceeded those of the control experiment. PDS treated cells on the other hand displayed similar expression rates as the TMPyP4 treated cells after 24 hours. Indeed, 5-LO mRNA expression slightly decreased after 48 hours of treatment, but still solely resulted in a non-significant decline of 13% compared to the control.

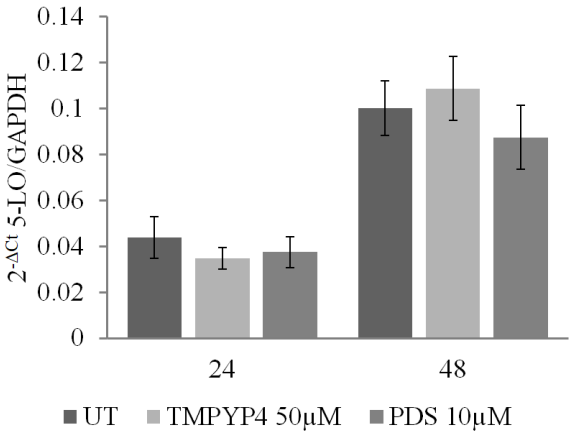


Figure 44: 5-LO mRNA expression in dHL-60 cells after treatment with G-quadruplex stabilizing agents TMPyP4 and pyridostatin (PDS). Cells were cultured for two days with their respective treatment and harvested every 24 h. Experiments n = 3 mean, ± s.d.

4.4.3.3 5-LO protein expression after treatment with G4-DNA stabilizing agents

Along with the 5-LO mRNA expression level, protein expression was examined. Therefore, samples were taken from the same cell cultures that were prepared for the RT-qPCR experiments, again after 24 and 48 hours of incubation. Cell pellets were lysed and protein amount was subsequently quantified. 20 µg of whole cell lysate was used for immunoblotting of 5-LO for each cell treatment. 5-lipoxygenase level was normalized to β-actin as loading control and quantification was carried out based on densitometry. Values determined for untreated dHL-60 cells were set to 100%, as reference for determining the reductive extent of the treated cell cultures.

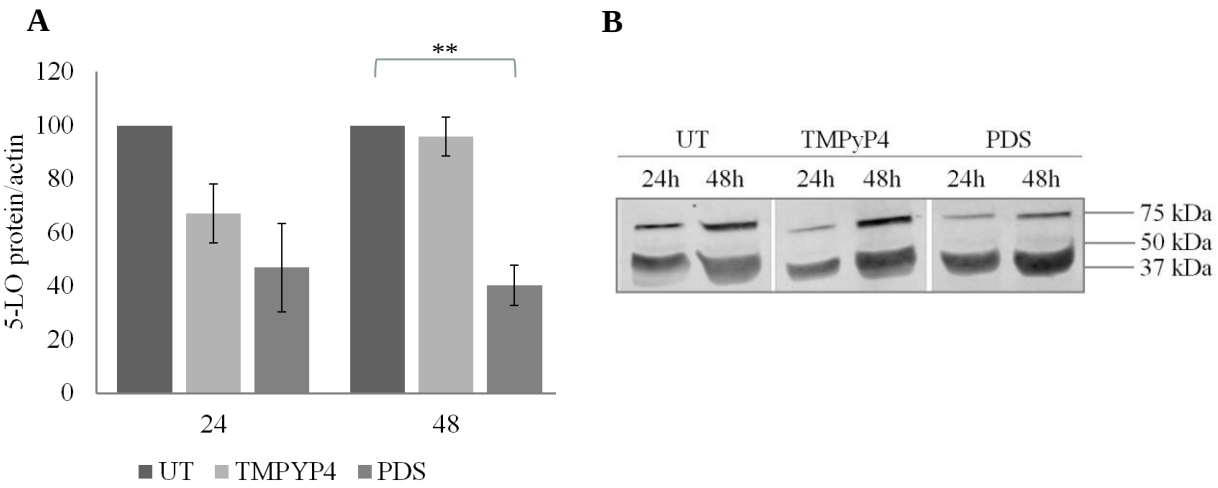


Figure 45: 5-LO protein expression in dHL-60 cells after treatment with G-quadruplex stabilizing agents TMPyP4 and pyridostatin (PDS). Cells were cultuered for two days with their respective treatment and harvested every 24 h to determine (A) protein expression based on densitometry and (B) corresponding immunoblot (anti-5-LO-ab, 1:1000, Cell Signaling Technologies). Experiments n = 3, mean ± S.E.M.

Unlike the results obtained for mRNA expression levels, broader effects on 5-LO protein expression level could be observed. TMPyP4 treatment again provided assorted results, as 5-LO levels decreased after 24 hours to about 67%, but crucially increased again after 48 hours to deliver almost 96% of the control expression. In contrast to that, PDS treatment steadily diminished protein expression levels time-dependently, resulting in a reduction to 47% and 40% in comparison with the untreated cells after 24 and 48 hours, respectively. Whereas the immunoblot itself only gave an estimate of decreased protein expression, densitometry clearly provided evident diminution of protein expression after treatment of PDS.

5. Discussion

5.1 Transcription factors in ALOX5 regulation: binding and expression patterns

5.1.1 Cell specific 5-LO expression and presence of TFs

The proximal 5-lipoxygenase gene promoter possesses specific characteristics, which resemble features of so called housekeeping genes or are attributed to gene promoters, whose transcription is regulated by CpG islands [35, 325], mainly featuring lack of TATA- or CAAT-boxes. CpG islands are conserved DNA stretches comprising high contents of CG-repeats that mostly contain binding sites for ubiquitously present transcription factors like Sp1, thereby generating transcription initiation sites [325]. Further transcriptional regulation is exerted by cytosine methylation within CpG islands, which in general contributes to gene silencing, as for the ALOX5 promoter in certain cell types [49]. DNA methylation thus provides an explanation for cell type specificity of 5-LO expression. In line with that, the uniquely occurring sequence, comprising five GC-boxes arranged in tandem, which are located in close proximity to the translation start site (ATG start codon) generate consensus sequences for transcription factors Sp1 and Egr-1 [34]. Both proteins were shown to be essential for basal promoter activity [45]. Although without transcriptionally regulating promoter activity, the heterodimer VDR/RXR was shown to bind to consensus sequences in the ALOX5 core promoter previously [56]. Finally, there exist binding sites for TF AP-2 along with evidence for the influence of TF WT1 on promoter activity [34, 310].

Along with these findings, it was inevitable to initially test on expression levels of some of the involved transcription factors, to provide preliminary data for further experiments considering their plain existence in all cell lines used. In analogy to cell type specific 5-LO expression, myeloid (both undifferentiated and differentiated) and lymphoid cell lines were used to evaluate the dependence of cell type and differentiation state on protein expression levels. The ubiquitous transcription factor Sp1 was steadily expressed in all cell lines, regardless of differentiation state of the myeloid lineage. In contrast to that, inducible factor Egr-1 could be detected in neither of the tested cell types by immunoblotting, although it is evidently playing a role in 5-LO basal transcription. The probable explanation for this is the time-dependent occurrence of this transcription factor after splitting of the respective cell cultures. In case of monocytic MM6 and dMM6, Egr-1 could be detected for about 6 hours after cell splitting before protein levels decreased to amounts, which were not detectable any further [45]. Considering the nuclear receptor VDR, its expression increased strongly after differentiation of myeloid cell lines. Indeed, VDR expression is induced above a basal level in promyelocytic/monocytic cell lines after treatment with $1,25(OH)_2D_3$ and goes along with

suppression of proliferation while initiating differentiation of promyelocytes into monocytes [326]. This behavior was observed for all myeloid cells tested to a comparable extent. Among the B-cells, only BL-41 displayed low detectable amounts of VDR. According to studies examining the expression pattern of the nuclear receptor, B-lymphocytes do not exhibit detectable levels of VDR, unless they experience certain activation signals, which in turn do not necessarily lead to a $1,25(OH)_2D_3$ response [327, 328], which could be an explanation for lacking protein levels in Rec-1 B-cells. Somewhat mixed results were obtained for RXRα. This nuclear receptor is widely expressed in different cell types and protein expression thereof can be induced during monocyte differentiation and induction of differentiation by $1,25(OH)_2D_3$ to different extents [329, 330]. In line with this, THP-1 exhibited induced expression of RXRα, whereas the protein level was more or less constant for HL-60. Surprisingly, MM6 did not show detectable protein levels at all, even after several repetitions of the respective immunoblotting procedure. This is clearly in contrast to their responsiveness to VDR/RXRα binding and may lead to the assumption that RXRα may be downregulated by differentiation. B-lymphocytes generally express RXRα to a rare extent [331], as could be confirmed for BL-41, which displayed no detectable protein amounts, whereas Rec-1 exhibited steady levels of the nuclear receptor. Interestingly, both B-cell lines either expressed the VDR or RXRα, while their opposite partner for heterodimerizing was absent, respectively. Since binding of the heterodimer VDR/RXRα was shown for certain functional VDREs throughout the ALOX5 gene in monocytes to promote $1,25(OH)_2D_3$ response in reporter gene assays [59], the presence of only one of the nuclear receptors in Rec-1 and BL-41 might contribute to a lacking $1,25(OH)_2D_3$-dependent induction of 5-LO expression in these B-lymphocytes.

Furthermore, none of the tested cell lines showed detectable protein expression levels of transcription factors WT1 and AP-2, suggesting that their respective concentrations are too low in whole cell lysates to be visible in immunoblots under the conditions used.

5.1.2 Affinity of transcription factor WT1 for the proximal ALOX5 promoter

Transcription factor WT1 supposedly occupies various roles in transcriptional regulation of several target genes and a dual role in cancerogenesis. It can either act as tumor suppressor or oncogene in leukemic tumorigenesis. The effects as TF are based on enhancing or repressing transcription of target genes [313]. Its respective functionality depends on its protein isoform, deriving from alternative splicing mechanisms (see section 4.2.4.3). These result in the insertion or deletion of exon 5 and a three amino acid sequence KTS, termed isoform (-/-),

(+/-), (-/+), (+/+), respectively. Isoforms +KTS and –KTS are generally expressed in a ratio of 2:1 [332] and isoforms containing +KTS are assumed to show lower DNA binding affinity, since the three amino acid insertion disrupts the spacing between zinc fingers three and four. Sequence similarities to the family of Krüppel-like TFs are nonetheless given due to C_2H_2-type zinc fingers for instance, which recognize GCG_5CG DNA sequences [312], thereby forming the possibility of interacting with the tandem GC-box of the ALOX5 promoter with overlapping binding sites. As neither the immunoblots of the whole cell lysates nor the MS-experiments provided detectable levels of WT1, a human recombinant protein isoform containing the KTS sequence was used for a DNA pulldown with isolated protein to increase the available protein concentration. The results clearly supported previous results obtained of WT1 inducing ALOX5 promoter constructs in its +KTS isoform, provided that the tandem GC-box is available. Apparently, under the tested conditions expression levels of WT1 are too low to prove DNA binding, which however becomes apparent in case of higher concentrations. Since there is evidence for an increase in promoter activity in reporter gene assays by WT1, the binding capacity of recombinant protein is not due to a mere *in vitro* affinity effect, but supports functional relevance. Additionally, activity was only induced in case of a fully present 5-fold GC-box (for +KTS). Considering that there exist naturally occurring mutations of this promoter sequence with *in vivo* relevance in certain pathogenic pathways (see section 1.1.5.1), further efforts in terms of functional studies are needed to elucidate the proteins' isoform-dependent regulatory role in 5-LO transcription and attributed diseases.

5.2 Evaluation and choice of experimental MS-based setup

5.2.1 Sample preparation and separation/ionization method for complex samples

The establishment of a convenient mass spectrometry setup for identifying DNA interacting proteins proved to be rather challenging in the first place, considering the multitude of options to choose from in terms of sample preparation, ionization technique and quantification method. A variety of already established methods provided a good starting basis, which mainly included in-gel digestion of SDS-PAGE separated proteins and liquid chromatography coupled to electrospray ionization and protein quantification. In order to decide for which one of the MS methods to go for, initial attempts for both techniques were carried out and compared to each other. Obviously, the gel-based approach coupled to MALDI-MS/MS resulted in an extremely poor yield with respect to protein identification, it was therefore excluded for the investigation of DNA-protein-complex identification. The underlying

reasons for this are manifold, the main drawback being a mere qualitative outcome of the experimental method (see also section 1.3.3). The major problem of gel-based approaches arises from its lack of sensitivity and lower identification rate compared to nLC-ESI systems [333]. 1D- or even 2D-gel electrophoresis of complex samples most prominently depicts proteins of high abundance, which interfere with the detection of low abundant ones [220, 334, 335], as was confirmed by the results obtained. These mainly included highly abundant nucleolin and PARP, ribosomal or histone-deriving proteins that were consistently found in multiple gel bands in the lower mass range of both WT and SCR sample and impeded the detection of other proteins. None of the identified proteins could be attributed to transcriptional regulation, or was exclusively found in the WT approach. Residual proteins could be attributed to RNA processing (DDX21, FBLL1/FBL), DNA repair (XRCC1) or ribosomal biogenesis (NHP2, SNU13). Another crucially disturbing factor was given by streptavidin, which contributed to gel smears obtained, thereby reducing gel resolution, additionally it generated permanent MS signals in almost every spectrum measured. These findings are consistent with the resolution capacity of gel electrophoresis, which is reported to represent 1000 proteins at utmost maximum, compared to 10000 protein in LC-MS/MS [336], indicating that transcription factors of low abundance are likely to be lost during protein separation. This could be counteracted by providing high concentrations of starting material, however higher amounts compared to the already used microgram amounts are unlikely to be successfully cultivated in cell culture conditions and would demand pooling of different cell passages.

Finally, a major prerequisite for quantifying proteins in a gel matrix requires steady and sensitive proteins dyes. For exact quantification, these must provide dying linearity, which is not always given by silver staining, although sensitivity is the highest possible. Silver staining additionally exhibits reduced sequence coverage compared to other dyes and is only semi compatible with further downstream MS [315, 337]. The used colloidal Coomassie stain enables detection of proteins as low as 1 ng [314], nonetheless faintly stained proteins could not always be identified reliably, probably due to ineffective peptide extraction efficiency. The latter indeed is quite low with a percentage share of 65-70% peptide recovery rate [338].

The use of 2D-gel electrophoresis might have provided an enlarged identification rate compared to the performed 1D-gel electrophoresis. However this was not taken into account, as especially membrane and nuclear proteins are often prone to be lost because of low solubility, hence their accessibility in the second gel dimension and overall identification is limited [339]. Along with these findings, the gel-based approach coupled to MALDI-MS/MS

was not pursued any further because of its inferiority to LC-ESI-MS/MS and subsequent quantitative analysis.

5.2.2 Quantification method, competitor and data analysis

After deciding to use a quantitative proteomics approach, the next step to evaluate the anticipated experimental setup was the choice of an adequate quantification method. Since five different cell lines were to be examined, metabolic labeling was excluded from the beginning, due to too high costs and time investment. The cost factor also applied to chemical labeling reagents iTRAQ and TMT; for practical reasons dimethyl labeling and label-free quantification were chosen as comparative strategies.

Generally, label-based techniques are still considered to be more accurate and precise in terms of protein quantification, however label-free methods now provide sufficient quantitation accuracy to match up to labeling strategies, provided a change in protein intensity abundance occurs at a factor of two at least. Lower fold changes typically are likely to be quantified erroneously [340]. Since label-free quantification is not limited to relative quantification experiments, its application in complex biological or clinical terms is more favorable than labeling techniques, thus it became the method currently used in most cases [341]. In line with this, both label-based and label-free strategies of the performed pulldowns of model cell line HL-60 resulted in similar findings, concerning identified interactors and distinguishability of unspecifically and specifically interacting proteins. All of the investigated parameters like total amount of identified proteins, overlap of identified proteins and their respective physico-chemical properties did not differ to a large extent, indicating a similar suitability of DNA-enriched protein quantification for both methods. A single difference that could be observed was a higher dynamic range of acquired molecular weights for the label-free approach, since it did capture more proteins in the lower (around 20 kDa) and upper (>400 kDa) mass range. Except for that, the sole crucial distinction between the two strategies arose from the significantly identified interactors, which were more prominently found in the label-free pulldowns, thereby enlarging the number of putatively interacting proteins with the chosen DNA sequence for further functional studies. This was basically the main reason for the label-free techniques to be pursued, although both quantification strategies performed equally well.

The successful discrimination between true and false interactors is a prerequisite for the identification of novel DNA-protein complexes. Identifying truly interacting proteins by quantitative abundance ratios provides sufficient confidence for distinguishing non-specific

from specific interactors without the need for stringent preparation conditions. Based on their equal affinity to the matrix in wildtype and control samples, contaminants are filtered out, although they will be widely detected in pulldown experiments (see also sections 1.3.2 and 1.3.3). This overcomes drawbacks of qualitative studies, which tried to dispose of nonspecific interactors by adding excess amounts of competitor oligonucleotides or harsh washing conditions, thus rendering these steps redundant, while applying mild sample purification [340]. In order to evaluate the influence or confirm these findings of excess competitor on protein identification and composition, a comparative pulldown with cell line HL-60 was performed with 40-fold sheared herring sperm DNA. The overall number of identified proteins was comparable in both states, but excess competitor resulted in a lower number of identified significant interactors, probably due to more stringent binding conditions. However, the obtained protein pattern identified did not alter to a high extent, leaving the question, whether high stringency conditions are necessary after all. Among the 16 detected interactors in the pulldowns containing 40-fold competitor, 9 were also found in other pulldowns performed without excess competitor, the remaining mostly representing RNA binding or processing proteins (RMXL1, DIM1), as well as ribosomal (RRP15, RL19) or histone proteins (H2A1J), which, considering their function, are negligible after all. This left two proteins of potential interest in the obtained results, being WDR18, which is involved in the regulation of ZN148 [342] and SP110, a transcription factor expressed in human leukocytes, which acts as activator of gene transcription. SP110 is furthermore assumed to be involved in nuclear hormone receptor coactivation [343]. However, the similar results obtained in pulldowns with and without excess competitor indicated a functional discrimination between non-specific and specific interactors, so that high stringency conditions are not necessarily needed and relinquishment thereof may even contribute to maintaining dynamic or low affinity binding partners that otherwise might be lost.

Given the complex nature of the data obtained from these experiments, bioinformatic analysis is essentially indispensable. Although one presumably partly forfeits autonomous analysis, bioinformatic software provides a reliable tool for dealing with these complex datasets of more than a thousand identified proteins in one pulldown. Nonetheless, a critical examination of the involved parameters (if possible) is essential for enhancing confidence of the results received. Since none of the features possess universal benchmarks (see section 4.3.2.3) and have to be adjusted for each individual pulldown approach, certain insecurity always remains when identifying significant interactors. Protein distribution and matching correlation support the assumption of a successfully performed experiment, however they do

not represent sole parameters for determining its success or failure. Missing value imputation on the other hand, is highly recommended for not losing interactors of low abundance, thereby introducing a bias, and capturing the entire range of a global proteome. According to studies, missing peptide values can range from 20% up to 50% for complex datasets [344], which points out the necessity of this particular bioinformatic step. In line with the former findings, the representation of enriched proteins in volcano plots and their respective identification as significant interactors depends on customized values of FDR and s0, which should be pondered before being set, as these obviously change the outcome of the experiment. Thus, especially the proteins detected close to the threshold curve are affected by changing these values and should be interpreted carefully and reasonably. Then again, statistical data analysis of pulldown experiments currently provides a confident groundwork for further functional studies, which are inevitable for specific protein characterization and elucidation of protein interaction networks.

5.3 Significant interactors of the proximal ALOX5 promoter

5.3.1 DNA pulldowns of core-promoter sequences

The overall identification of significant interactors of the 120-mer containing the tandem GC-boxes revealed a total of 66 proteins, including already known transcription factors, but also novel binding partners putatively involved in the regulation of 5-LO expression. The similarities in protein patterns enriched in the different pulldowns point to a similar gene regulation, regardless of cell lines or differentiation state. Over one third of the enriched proteins contain zinc fingers of different kinds as their DNA-binding domain (mainly C_2H_2-type), indicating a high affinity of these TFs to the GC-rich sequences of the proximal promoter. The main class thereof comprise transcription factors of the Sp/KLF-family, out of which Sp1 and Sp3 are proven regulators of 5-LO expression [34, 53]. Additionally, Sp2 and Sp4 were detected in two and one quantification event, respectively, which, considering their sequential relation to Sp1/Sp3, seem to be plausible in regulating ALOX5 promoter activity. There exist interactions of members of the Sp/KLF-family, since both exhibit similar to identical consensus binding sites (GT- or GC-box) [38]. Accordingly, the identification of KLF13 and KLF16 as interacting partners is absolutely logical, however their specific role in 5-LO regulation has yet to be determined. KLF13 is ubiquitously expressed and functions as transcriptional repressor, able to compete with and to antagonize Sp1-mediated transcriptional activation [321]. Its involvement in inflammatory processes is pictured by different studies and includes regulation of chemokines (activation of RANTES) and subsequent activation of

T-cells [321, 345] and presumably activation of macrophages [346]. In line with this, KLF13 was shown to be a target gene of microRNA-125a-5p (miRNA-125a), thereby possibly contributing to the regulation of macrophage responses and differentiation [345, 347]. In this context, the small non-coding miRNAs of about 20 nucleotides length are post-transcriptional regulators of mRNA expression, which generally bind to the 3'-UTR of target mRNAs [347]. 5-LO itself is shown to be target gene for miRNA-125b-5p, which possesses the same core sequence as miRNA-125a-5p and directly inhibits 5-LO protein expression [347, 348], indicating an involvement of miRNA-125 in the regulation of inflammatory processes by repression of both KLF13 and 5-LO. KLF16 on the other hand is fairly characterized, it is known to be able to compete with and replace Sp1 from binding sites, thereby displaying inhibitory effects [349]. It is furthermore shown to regulate metabolic-gene expression [350]. Both KLF13 and KLF16 couple to members of histone deacetylase complexes and thus contribute to chromatin remodeling and associated transcriptional repression [351].

Further identified proteins comprise VEZF1 and MAZ, zinc finger transcription factors with shared sequence homologies in their DNA-binding domain, which bind GC-repeats [352]. VEZF1 could be associated with the regulation of DNA methylation patterns in murine promoters containing CpG islands [353] and with slowing of RNA polymerase II dependent transcript elongation [354]. The latter similarly applies for MAZ, which was shown to pause Pol II elongation and promote transcript termination *in vitro* [355]. Interestingly, Chip-Seq data sets revealed enhancer sites in intron C, D and G of the ALOX5 gene [36], in which both Pol II and MAZ simultaneously bind, suggesting a potential regulatory function of MAZ on transcript elongation (suppl. figure 5). However, whether these effects are actually relevant in 5-LO transcriptional regulation needs to be determined in further studies. Besides its additional regulatory impacts on G-quadruplex dependent gene transcription (see section 1.2.2.3), transcription factor MAZ is also associated with the pathogenesis of inflammatory diseases including rheumatoid arthritis and atherosclerosis, it is involved in the regulation of serum amyloid A expression and potential upregulation of TGFβ [356, 357]. In this context, inflammatory regulation by the already mentioned miRNA-125b is not only mediated by inhibiting 5-LO protein expression, but equally MAZ expression [358]. Furthermore, several studies gave evidence for an interplay of MAZ and Sp1 on transcriptional regulation of target genes, both G4-dependent and -independent [82, 129, 130, 359, 360], which might be applicable to ALOX5 gene regulation as well. All of these findings point to their regulatory role in 5-LO expression, which makes it a target of high interest for determining its functional relevance in future studies.

Along with transcription factor MAZ, binding of RUNX1 and ZBT7A to the proximal ALOX5 promoter is proven by Chip-Seq datasets available, which further confirms the results obtained here by DNA pulldown assays (suppl. figure 6). Although RUNX1 could only be detected in one quantification event (dHL-60), binding seems plausible, since the induction of promoter activity of ALOX5 promoter constructs by RUNX1 was shown in reporter gene assays. Accordingly, fusion proteins of RUNX1 could be associated with elevated ALOX5 gene expression in acute myeloid leukemia [361]. As RUNX1 exhibits widespread functions in cell proliferation, differentiation and functional hematopoiesis, its involvement in 5-LO has to be determined in further experimental studies. Transcription factor ZBT7A likewise displays various functions in cell proliferation and differentiation, its gene was identified as proto-oncogenic previously. It is able to compete with and replace Sp1 from binding sites, thereby repressing transcription of target genes. Additionally, repression of transcription can be achieved by recruitment of corepressors leading to histone deacetylation [362]. ZBT7A is yet another direct target of miRNA-125a, which in this context is proposed to have anti tumorigenic activity by synergistically targeting both ZBT7A and WT1 [363]. Apart from possible direct effects on 5-LO expression by binding to its core promoter, ZBT7A additionally is capable of interacting with TGFβ mediator SMAD4, resulting in the inhibition of TGFβ-dependent transcriptional activation, and thereby might exhibit secondary regulatory impact on the ALOX5 expression, as TGFβ responsive gene [364].

A large part of the novelly identified significant interactors, including for instance ZN148, ZN316 and ZN579 account for poorly characterized zinc finger proteins that are supposed to be involved in transcriptional regulation. PRD10 might function as transcription factor with methyltransferase activity, given the fact that it contains C_2H_2 zinc fingers for direct DNA interaction as well as a conserved SET domain in its N-terminal region. It is presumably involved in the development of arthritis [365]. Zinc finger protein ZN148 can either activate or repress transcription of genes, the latter often by competing with Sp1 for binding sites. Its expression can be downregulated by certain inflammatory cytokines [366] and it directly interacts with tumor suppressor p53, thereby affecting cell growth and proliferation [367]. ZN281 shares sequence homologies with ZN148 and possesses a similar DNA binding domain. It is involved in the regulation of pluripotency and cell differentiation, by cooperating with other developmental transcription factors [368]. Furthermore, ZN281 is shown to be involved in the development of intestinal inflammation and to repress the transcription of a variety of inflammatory genes, which makes it a plausible regulator of ALOX5 expression [369, 370].

5.3.2 DNA pulldowns of core-promoter proximal sequences

Since the above results of the 120-mer DNA pulldowns did not point to a potential cell type specific regulation of the ALOX5 promoter, pulldowns with 80-mer oligonucleotides were performed using one model cell line, for which HL-60 cells were chosen, since these provided the highest amount of identified interactors of all cell lines tested. 80-mer oligonucleotides comprised sequences upstream of the 120-mer used, with overlap of 6 bp to each of the utilized promoter sequences, in order to not divide putatively existing protein binding sites.

The sequence containing the SMAD binding element did not reveal actual TGFβ mediators, however provided a large variety of CCAAT/enhancer binding proteins enriched in the pulldowns. C/EBPs are a family of transcription factors that widely contribute to functional cellular processes, such as cell growth, proliferation and differentiation, as well as inflammatory responses. Due to their widespread involvement therein, dysfunction results in numerous pathological pathways. There exist six members of the family, sharing their DNA binding domain, basic leucine zippers, with which they are able to homo- or heterodimerize, a prerequisite for DNA binding [371]. Their expression patterns vary depending on the respective protein, C/EBPα and γ are expressed in several tissues, C/EBPβ in myelomonocytic cells among others and C/EBPε exclusively in myeloid and lymphoid cells. C/EBPα, β and ε are pivotal for granulocyte/macrophage differentiation and activation, hence for a functional immune response. Furthermore, C/EBPα, β and δ have been shown to induce FLAP expression, thereby exhibiting indirect effect on 5-LO mediated leukotriene production [372]. In addition, expression of C/EBPβ and δ is induced by inflammatory signals like LPS, IL-1 and IL-6 and further cytokines, and subsequently contributes to transcriptional regulation of inflammatory target genes, including cyclo-oxygenase-2, TNFα or monocyte differentiation antigen CD14 [322]. In line with that, the heterodimer C/EBPα/β was proven to mediate monocyte maturation and differentiation after treatment with TGFβ/1,25$(OH)_2D_3$. Intriguingly, regulation of expression of CD14, which in analogy to the ALOX5 gene is a TGFβ/1,25$(OH)_2D_3$-responsive gene, could be attributed to this heterodimer, indicating that C/EBPα/β occupies a critical role in exerting synergistic impacts of TGFβ/1,25$(OH)_2D_3$ and thereby regulates TGFβ-mediated signaling in monocyte differentiation [373]. These findings might not be limited to the CD14 gene and further investigation of ALOX5 gene regulation in this behalf seems to be worthwhile. Supporting evidence for C/EBPβ as a regulator of TGFβ signaling was further given for the differentiation of myoblasts [374].

C/EBPγ does not contain an activation domain and requires heterodimerization with either C/EBPα or β, thereby modulating their respective activity. Accordingly, the C/EBPα-C/EBPγ

ratio is proposed to be responsible for proper lymphoid differentiation in murine cellular systems and a decrease of granulocytic differentiation in case of higher C/EBPγ levels [375]. High levels of C/EBPγ in heterodimeric C/EBPβ-C/EBPγ on the other hand promote cell proliferation and suppress transcription of several inflammatory genes including IL-1 or IL-6 [376]. Finally, C/EBPε is an essential transcription factor in normal granulopoiesis and furthermore determines differentiation into granulocytes or macrophages [377, 378]. All of these findings promote a plausible and pivotal role of this family of transcription factors in the regulation of ALOX5 transcription with further confirmatory evidence of C/EBPα and β binding to the proximal promoter sequence given by Chip-Seq data (see section 4.3.2.5). Further elucidation of their exact regulatory effects and protein characterization needs to be exerted in coming studies.

Significant interactors identified in the 80-mer oligonucleotide pulldowns containing the putative VDRE provided another protein pattern, which however also includes further transcription factors of the KLF-family, as well as zinc finger proteins and methyltransferase NSD1, some of which were detected in both cell states (HL-60 and dHL-60). Out of the zinc finger proteins enriched, only ZN335 is sparsely characterized. It is reported to enhance ligand-dependent transcriptional activation by nuclear hormone receptors, such as RXRα [379] and furthermore interact with members of a H3K4 methyltransferase complex, including MLL. In this way, it is proposed to recruit methyltransferase complexes to gene promoters, thereby contributing to transcriptional regulation [380]. Interestingly, another histone methyltransferase complex member interacting with MLL could be identified as significant interactor, namely NSD1. NSD1 possesses a SET domain, which exhibits catalytical histone methyltransferase activity and is responsible for methylation of lysine 36 on histone 3 and lysine 20 on histone 4 [381], monomethylation of the latter, as marker for transcript elongation, also being induced by TGFβ/1,25$(OH)_2D_3$ [58]. Additionally, NSD1 analogously contains two nuclear receptor binding domains, one of which interacts with the receptor ligand-independently, while the other does in the ligand-bound version of RXRα for instance, suggesting a transcriptional regulation of target genes by these properties [382]. In line with this, the used DNA sequence of the ALOX5 promoter comprises a putative VDRE, which is known to bind heterodimeric VDR/RXRα [56] and which possibly is responsible for the methyltransferase's interaction. NSD1 thereby might either modulate nuclear receptor signaling or contribute to recruitment of the MLL-associated protein complex, which indeed was shown to induce ALOX5 promoter activity in association with VDR/RXR [54].

Transcription factor ZFX represented another significantly interacting protein, whose binding to this region of the ALOX5 gene is confirmed by Chip-Seq data (see section 4.3.2.5). It acts as transcriptional activator, most prominently binding to CpG islands of promoter regions [383], which correlates to the CpG islands present in the 5-LO promoter. Protein expression of ZFX was furthermore shown to increase over time after cell treatment with TGFβ, suggesting an induced regulatory effect of the former on 5-LO expression after differentiation of myeloid cell lines [384]. Finally, two other members of the transcription factors of the Krüppel-like family were identified as significant interactors. KLF12 is a scarcely characterized transcription factor involved in cell proliferation, whereas KLF5 is a transcriptional activator or repressor, contributing to the regulation of inflammatory response, cell proliferation and apoptosis [385]. Its functionality is in part controlled by TGFβ and $1,25(OH)_2D_3$ signaling. Its expression levels and subsequent impacts depend on the respective cellular context and increase after stimulation with inflammatory stimuli, however decrease after cell treatment with $1,25(OH)_2D_3$ [386]. KLF5 is able to promote pro-inflammatory effects of IL-1β and nuclear factor kappa-B, which in turn can be downregulated by cellular treatment with TGFβ, which impairs the transcription factor's expression in certain cell types [387]. Further support, that KLF5 is an essential cofactor in TGFβ signaling but exhibits different effects in a different cellular context, is given by the fact, that upon TGFβ treatment KLF5 shifts its function from repressor to coactivator in the expression of CDKN2B, a cell cycle inhibitor and effector of TGFβ-dependent growth inhibition. Accordingly, without TGFβ activation, KLF5 is responsible for inhibition of CDKN2B expression, whereas with TGFβ activation, it mediates induction of CDKN2B expression. These opposing effects are shown to rely on TGFβ-mediated acetylation of KLF5 by recruiting acetyltransferase p300. KLF5 thereby is assumed to experience altered capability of binding to the CDKN2B gene promoter sequence and reverse its transcriptional effect [388, 389]. Similar results of opposite effects were obtained in studies examining the TGFβ-mediated inhibition of c-Myc expression, which relies on KLF5. In this study, KLF5 was shown to be essential for c-Myc expression without TGFβ activation, however, after TGFβ activation KLF5 turned out to be a crucial factor for mediating TGFβ-mediated suppression of c-Myc expression [390]. Thus, KLF5 is a crucial effector protein in TGFβ-induced gene expression, which could also play an essential role in 5-LO expression. When contemplating the regulatory activity of KLF5 in TGFβ-mediated upregulation of CDKN2B, analogous or similar effects might apply to ALOX5 gene control. Furthermore, KLF5 interacts with various other transcriptionally active proteins, including TGFβ effectors SMADs, as well as transcription factors C/EBPβ and

RXRα (suppl. figure 7), which seems plausible considering both the identified interacting proteins and its highlighted function in gene control.

5.4 G-quadruplex formation in the proximal ALOX5 promoter sequence

Based on the identified interacting proteins of the proximal ALOX5 promoter sequence containing the tandem GC-box, a formation of G-quadruplex structures of this DNA sequence seemed probable. Six proteins were recurrently enriched in the different pulldowns that are involved in recognizing or resolving G4-DNA, including YBOX1, hnRNP D and hnRNP K, BLM, DHX36 and MAZ (see sections 1.2.2.3 and 1.2.2.4). As described in section 1.2.2.3, putative G-quadruplex forming sequences exhibit higher probability to be located in close proximity to G-rich sequences and correlate to the existence of GC-boxes and its interacting transcription factor Sp1. These findings support the proximal ALOX5 promoter to be a potential candidate for these secondary DNA structures. Additionally, the interaction of TF Sp1 and MAZ, which was demonstrated in several studies, promotes a G4-DNA-dependent regulation of oncogene transcription and thereby occupies a central role in a possible G-quadruplex mediated transcription of 5-LO. In line with this, as shown for the h-RAS promoter (see figure 6), a synergistic interplay of both transcription factors is inevitable for G4-structure unwinding facilitated by MAZ and subsequent Sp1-induced transcriptional activation [82]. Intriguingly, the h-RAS promoter displays high similarity to the 5-LO promoter and also lacks TATA- or CCAAT-boxes, shares the same extremely high GC content of 80% and possesses several GC-boxes of the sequences GGGCGG and its reverse equivalent CCGCCC. Two of the latter are equally arranged in tandem with another following closely behind with a sole basepair missing for properly representing the third tandem box, but which are responsible for G-quadruplex formation. The altogether G4-DNA mediated transcriptional regulation may therefore not only be limited to the h-RAS gene, but might also equally apply to ALOX5 gene control.

Another regulatory effect deriving from transcription factor MAZ is given by the fact that it is responsible for G-quadruplex controlled transcription of c-myb [81]. Since the c-myb protein product itself is repressing 5-lipoxygenase gene expression [391] MAZ indirectly contributes to the transcriptional regulation thereof.

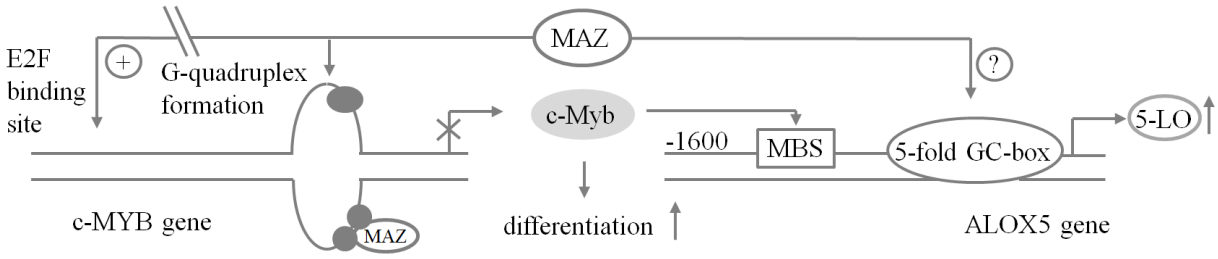


Figure 46: Proposed role of transcription factor MAZ in the regulation of 5-LO expression. MAZ regulates c-MYB transcription both G-quadruplex-dependent and -independent and activates [128] *or represses* [81] *c-Myb expression, respectively. Since transcription factor c-Myb itself binds to the ALOX5 gene (MBS = c-Myb binding site) and thereby suppresses 5-LO expression, MAZ exhibits secondary effects on the latter, in that G4-stabilization results in diminished c-MYB transcription and following upregulation of 5-LO expression. These effects are obviously reversed in case of induced c-MYB expression by binding of MAZ to the E2F site in the c-MYB gene. Proposed direct effects of MAZ are deriving from interaction with putative G-quadruplex structures forming in the 5-fold GC-box.*

The aforementioned assumptions could eventually be confirmed by means of *in vitro* experiments. According to the results obtained, the proximal ALOX5 promoter sequence containing the tandem GC-box is indeed able to form tetraplex structures, which could exert regulatory functions on gene transcription. The studies performed were the first to give evidence for G-quadruplex formation in this DNA region, however its actual topology and conformation has yet to be determined. The achieved results from *in cellulo* performed experiments displayed somewhat ambiguous activity of G4-DNA targeting small molecules TMPyP4 and PDS. Treatment of differentiated HL-60 cells with both agents provided different outcomes concerning cell viability and toxicity, although both agents are reported to exhibit cell toxic functions. Accordingly, PDS treated cells forfeited viability, whereas treatment with TMPyP4 did neither result in impaired viability nor cytotoxic effects, suggesting a potentially lacking or attenuated cellular uptake of the latter. Indeed, although cellular uptake of the porphyrin analoga was demonstrated in different cell lines by monitoring their accumulation by fluorescence emission [142, 392], the actual mechanism seems to be faintly characterized. In line with this, TMPyP4 stimulation did not exhibit any effect on 5-LO mRNA or protein production in the model cell line either. These findings could certainly indicate the absence of *in vivo* formation of G-quadruplex structures, however since the *in vitro* evidence is quite obvious, another approach for targeting potential secondary structures seems worthwhile. Cellular uptake was furthermore shown to be dependent on the specific porphyrin used [393], thus an exchange of G4-stabilizing porphyrin might result in

another outcome. Additionally, TMPyP4 binding affinity was found to depend on the respective quadruplex topology, which could be another explanation for diminished binding capacity (see section 1.2.2.5). PDS treated cells exhibited a different behavior of 5-LO expression, which was associated with slightly decreased mRNA-, as well as significantly decreased protein levels after 48 hours, indicating existing repressive secondary effects of probable G4-DNA stabilization, which affect protein expression to a higher degree than mRNA expression levels.

Altogether, the discrepant outcomes of the *in cellulo* experiments are mostly due to the inconsistent and unexpected results obtained from cellular treatment with porphyrin TMPyP4, which might be altered and even improved by choosing a different G-quadruplex stabilizing agent. Nonetheless, further efforts are inevitable for elucidating the actual transcriptional impact of potential *in vivo* G-quadruplex structures on 5-LO expression. However, the combined findings of identified proteins associated with G4-DNA structures, the *in vitro* evidence for their respective formation and repressive effects on 5-LO protein expression induced by PDS all are indicative for a highly probable regulatory function of G-quadruplexes in the proximal ALOX5 promoter sequence.

5.5 Conclusion and outlook

The overall identification of the proximal ALOX5 promoter sequence revealed known and novel potential interactors and regulators of 5-LO expression. Similar protein patterns of the oligonucleotides containing the tandem GC-box indicate a mutual transcriptional regulation in the myeloid lineage, independent of differentiation or maturation state. With regard to the number of significantly enriched proteins, a trend towards lower amounts with higher cell maturation could be determined, suggesting additional regulatory effects in promyeloid cell lines. The most consistently enriched interactors also apply for the B-lymphocytes tested, although less interacting proteins were found in total. However, overall ALOX5 gene regulation based on promoter binding proteins seems to be cell unspecific. Considering both 120-mer and 80-mer oligonucleotide pulldowns, most of the relevant transcription factors found are involved in various pivotal cellular functions like cell proliferation, growth and differentiation, and an interplay of some of these transcription factors seems likely, indicating a highly complex regulation of 5-LO expression. Although several newly found proteins could potentially be involved in ALOX5 gene control, there exist a few sovereign proteins, which, considering their functional involvement, represent crucial regulators thereof, including transcription factors of the KLF-family (KLF5 and KLF13), the CCAAT/enhancer

binding protein family (C/EBPα, β, ε) and further transcription factors ZBT7A and MAZ. Some of these (C/EBPα, and β, KLF5) additionally exhibit features that might help to promote continuative efforts for deciphering the TGFβ/1,25$(OH)_2D_3$-mediated regulation of 5-LO expression. Others (mostly MAZ and its respective interaction with Sp1) represent interesting candidates for elucidating the potential G-quadruplex formation in the proximal ALOX5 promoter and its regulatory impacts on gene transcription. Altogether, the listed proteins therefore display a starting point of utmost interest for subsequent functional studies, which will provide further insight into the regulation of 5-lipoxygenase expression.

6. Summary

The enzyme 5-lipoxygenase (5-LO) occupies a central role in the biosynthesis of inflammatory leukotrienes and thus takes part in the pathogenesis of related diseases. Its occurrence is mainly restricted to cells of the immune system including granulocytes, monocytes/macrophages or B-lymphocytes and can be induced by cell differentiation of myeloid cells after treatment with differentiating agents, such as DMSO, retinoic acid or the combination of TGFβ/1,25$(OH)_2D_3$. The latter contribute to the highest level of induction of mRNA and protein expression. Its cell specific occurrence is at least partly due to DNA methylation in cells that do not exhibit 5-LO activity and genetic regulation is further dependent on histone acetylation. 5-LO expression is controlled by transcription factors binding to the promoter sequence of the ALOX5 gene that induce basal promoter activity, as well as promoter independent effects including transcript initiation and elongation, which are mostly attributed to TGFβ/1,25$(OH)_2D_3$ signaling. The ALOX5 gene resembles a typical housekeeping gene, hence lacks TATA- or CAAT-boxes for transcriptional regulation, but displays a high GC-content with eight GC-boxes, five of which are arranged in tandem, that provide binding sites for transcription factors Sp1, Sp3 and Egr-1.

The proximal ALOX5 promoter is furthermore a target for additional factors, such as TGFβ effector proteins SMADs or the vitamin D receptor and possesses additional consensus sequences for transcriptional regulators, including NF-κB or PU.1. However, as yet no actual binding of these proteins to the promoter sequence was demonstrated and an unbiased screening for identifying further ALOX5 promoter interacting proteins, which might have impact on 5-LO expression, is still lacking. For this purpose, the present study focused on the identification of significantly interacting proteins, employing DNA-affinity enrichment coupled to label-free quantitative proteomics, spanning a sequence of about 270 base pairs of the proximal ALOX5 promoter. For the elucidation of potential cell specific differences in protein patterns and compositions, DNA pulldowns were performed by using oligonucleotide stretches comprising the core promoter sequence including the 5-fold GC-box, which were incubated with different cell lines and differentiation states of myeloid, as well as B-lymphocytic lineages. In order to compare different mass spectrometric quantification strategies that would allow for identification of interactors, dimethyl labeling and label-free techniques were used. Since the label-free approach outperformed the label-based one in initial experiments, it was established as standard quantification strategy in all DNA pulldowns performed. The pulldowns of myeloid cell lines in both undifferentiated and differentiated state and B-lymphocytes resulted in a cell-unspecific protein pattern whose

composition was similar, regardless of cell lineage. Additionally, further DNA sequences comprising either a vitamin D response element or a SMAD binding element were investigated in the promyelocytic model cell line HL-60 in both undifferentiated and differentiated state. The identified proteins confirmed known interaction partners and furthermore revealed novel potential regulators of the 5-LO promoter. Out of these, the most prominently identified and promising proteins included transcription factors of the KLF- and CCAAT/enhancer binding protein-family. In this context, KLF5 and KLF13 are both involved in the regulation of inflammatory processes, the former additionally being an effector protein of TGFβ-signaling, whose functional characterization is of utmost interest in terms of regulation of 5-LO expression. Further protein characterization will be inevitable for the CCAAT/enhancer binding proteins C/EBPα, C/EBPβ and C/EBPε. These transcription factors are involved in the regulation of inflammatory processes and heterodimers thereof (C/EBPα/β) are known to control TGFβ/1,25$(OH)_2D_3$-mediated effects of the CD14 gene.

Several of the identified proteins of the pulldowns containing the tandem GC-box represented interactors of G-quadruplex DNA, including the helicases BLM and DHX36, the ribonucleoproteins hnRNP D and hnRNP K and transcription factor MAZ. Since G-quadruplexes form in G-rich DNA sequences as secondary DNA structures and exhibit substantial regulatory effects on the transcription of their target genes, the potential formation thereof in the ALOX5 core promoter sequence was investigated in a second project. Out of the proteins mentioned above, MAZ is shown to exert resolving effects on G4-DNA and synergistically induce Sp1-dependent gene activation of oncogene h-RAS, which displays analogous promoter characteristics to the ALOX5 gene. A DNA stretch comprising the tandem GC-box was used for elucidating the potential of secondary DNA structure formation. Intriguingly, both immune-based and spectroscopic methods provided clear evidence for the *in vitro* G-quadruplex formation of the proximal promoter sequence for the first time. In order to provide additional information on a possible regulatory effect of existing G-quadruplex structures on 5-LO transcription, differentiated HL-60 cells were subsequently treated with two distinct G4-DNA stabilizing agents. A porphyrin analogon (TMPyP4) did not exhibit any effects on 5-LO mRNA and protein expression after cell treatment. A second G4-DNA stabilizing agent (pyridostatin) on the other hand revealed significant reduction on 5-LO protein expression after cellular treatment. These mixed results render further experiments inevitable, in order to provide a clear assertion as to whether 5-LO expression is regulated by G-quadruplex structures or not.

Altogether, this study enlarges the knowledge of ALOX5 proximal promoter interacting proteins by corroborating the binding of already known transcription factors and identifying novel interactors. It yields essential groundwork for subsequent functional studies of proteins involved in 5-LO transcription and introduces G-quadruplexes as a new potential mechanism in ALOX5 gene regulation.

7. Zusammenfassung

Das Enzym 5-Lipoxygenase (5-LO) stellt das Schlüsselenzym im Biosyntheseweg der Leukotriene dar, einer Gruppe pro-inflammatorischer Mediatoren, die an der Entstehung vielerlei entzündlicher Erkrankungen wie Asthma oder rheumatoider Arthritis beteiligt sind. Darüber hinaus findet sich eine erhöhte 5-LO Expression in einigen Tumorarten. Die 5-LO katalysiert hierbei über mehrere Schritte die Umsetzung frei vorliegender Arachidonsäure in entsprechende Leukotrien A_4 durch Oxidation. Das Enzymexpressionsmuster beschränkt sich hauptsächlich auf Zellen des Immunsystems myeloiden und lymphatischen Ursprungs, wobei ihr Vorkommen unter anderem auch in Langerhans-Zellen und Zelltypen des Gehirns nachgewiesen werden konnte. Die 5-LO-Expression in diesen Zelltypen kann durch die Präsenz einiger Stimuli maßgeblich gesteigert werden. Darunter fallen vor allem Agenzien, welche die Zellreifung induzieren wie beispielsweise DMSO, Retinsäure oder die kombinierte Gabe von TGFβ/1,25$(OH)_2D_3$. In diesem Zusammenhang finden sich erhöhte 5-LO-Proteinlevel in Granulozyten und Monozyten/Makrophagen, die aus der Differenzierung ihrer entsprechenden myeloischen Vorläufer entstanden sind.

Die Enzymregulation auf transkriptioneller Ebene wird über verschiedenste Mechanismen gesteuert, die bis heute nicht vollständig geklärt sind. Maßgeblich an regulatorischen Effekten beteiligt sind epigenetische Faktoren wie DNA-Methylierung und Histonacetylierung, sowie die Transkriptionskontrolle durch unterschiedliche Transkriptionsfaktoren. Der Methylierungsstatus ist einer der Hauptgründe für die zellspezifische Enzymexpression, wobei eine vollständige DNA-Methylierung im Promoterbereich für die fehlende Enzymaktivität in 5-LO-negativen Zelllinien verantwortlich gemacht wird. Die Regulation durch Transkriptionsfaktoren kann sowohl durch deren Bindung in der Promoterregion, als auch durch Bindung in mehreren Introns des ALOX5-Gens erfolgen. Die posttranskriptionelle Expressionsregulation wurde hierbei bislang auf verschiedene SMADs als TGFβ-Effektoren, sowie Vitamin-D-Rezeptor Response Elemente im distalen Part des 5-LO Gens zurückgeführt. Die Enzyminduktion nach Differenzierung myeloider Zelllinien nach Zugabe von TGFβ/1,25$(OH)_2D_3$ konnte in diesem Zusammenhang bereits durch Mechanismen der Tranksriptionsinitiation und -elongation als Promoter-unabhängige Effekte erklärt werden.

Der Kernpromoter und die Promoter-proximalen Bereiche des ALOX5-Gens weisen Bindungsstellen für verschiedene Transkriptionsfaktoren auf, die unter anderem die basale Promoteraktivität aufrechterhalten. Der 5-LO Promoter weist hierbei besondere strukturelle Charakteristiken auf. Er zeichnet sich durch einen erhöhten Anteil an GC-Basenpaarungen aus, die CpG Inseln bilden und den Großteil der Transkriptionsfaktor-Konsensussequenzen

stellen. Durch den hohen GC-Anteil von etwa 80% und das Fehlen von TATA- und CAAT-Boxen ähnelt die Promoterregion typischen Housekeeping-Genen. Vor allem im proximalen Bereich nahe der Transkriptionsinitiationsseite finden sich acht GC-Boxen, von denen fünf in Tandem arrangiert vorliegen. Diese bieten Konsensussequenzen für Transkriptionsfaktoren der Sp/KLF-Familie, sowie Egr-1, wobei die Bindung für Sp1, Sp3 und Egr-1 in Studien bereits gezeigt werden konnte. 5'-upstream der Transkriptionsinitiationsseite finden sich außerdem Bindungsstellen für die Transkriptionsfaktoren NF-κB oder PU.1. Für deren Bindung, oder der Bindung weiterer regulatorischer Faktoren fehlt bisher allerdings der Nachweis. Die vorliegende Datenlage erforderte in diesem Zusammenhang einen neutralen Screening-Ansatz der proximalen ALOX5 Promotersequenz, um DNA-interagierende Transkriptionsfaktoren zu identifizieren und die bisher Hypothesen-getriebene Charakterisierung der Genregion zu erweitern.

Zu diesem Zweck war es Zielsetzung dieser Arbeit, ALOX5-Promoter-interagierende Proteine mittels DNA-Affinitätschromatographie (DNA-Pulldown) und anschließender Massenspektrometrie zu identifizieren. Dazu wurden verschiedene MALDI-MS und ESI-MS Techniken zwecks ihrer Eignung zur Identifizierung neuer Proteinkomplexe durchgeführt und miteinander verglichen. Die Komplexität der durch DNA-Affinitätschromatographie angereicherten DNA-Interaktoren erforderte schlussendlich eine ESI-MS basierte Methode, die anschließend mit bioinformatischen Methoden gängiger quantitativer Proteomics-Ansätze ausgewertet wurde. Um eine geeignete MS-Quantifizierungsstrategie zur Isolierung DNA-interagierender Proteine aus komplexen Proben zu finden, wurden sowohl markierungsfreie als auch markierungsbasierte massenspektrometrische Methoden untersucht und miteinander verglichen. Die markierungsbasierte beruhte hierbei auf dem Prinzip des Dimethyllabelings der nach dem Proteinverdau erhaltenen Aminofunktionen der Peptide. Hinsichtlich ihrer Effizienz der Identifizierung signifikanter Interaktoren des 5-LO-Promoters war das Dimethyllabeling der markierungsfreien Quantifizierung deutlich unterlegen, weshalb der markierungsfreie quantitative Ansatz als Standardmethode für weiterführende DNA-Pulldowns gewählt wurde.

Die verwendeten Promotersequenzen des 5-LO Gens umspannten einen Bereich von etwa 270 Basenpaaren Gesamtlänge, der sich 5'-upstream der Translationsstartstelle befindet und nochmals in Promoter-proximale und Kernpromoter-beinhaltende Sequenzen unterteilt wurde. Um potentiell vorhandene zellspezifische oder differenzierungsabhängige Unterschiede der Enzymregulation in verschiedenen Zelltypen und Zellstadien zu berücksichtigen, wurden Zelllinien myeloiden und lymphatischen Ursprungs zur Durchführung der DNA-Pulldowns

mit den Kernpromoter-enthaltenden Sequenzen verwendet. Die ausgewählten Zelllinien umfassten drei verschiedene Linien myeloider Abstammung in unterschiedlichen Reifegraden, die zudem in undifferenziertem, als auch differenziertem Zustand untersucht wurden. Zusätzlich wurden zwei weitere B-lymphozytäre Zelllinien verwendet. Die identifizierten DNA-Interaktoren für die verwendeten Zellen myeloider Abstammung in undifferenziertem und differenziertem Zustand und Zellen lymphatischen Ursprungs ähnelten sich in deutlichem Umfang hinsichtlich der Art und Anzahl ihrer identifizierten Proteine, als auch in ihrer jeweiligen Proteinzusammensetzung, sodass auf eine zellunspezifische und differenzierungsunabhängige Genregulation rückgeschlossen werden konnte. Zur weiteren Durchführung folgender DNA-Pulldowns der Promoter-proximalen Regionen, die entweder eine Konsensussequenz für den Vitamin D-Rezeptor, oder SMAD-bindende Elemente enthielten, wurde schlussendlich die promyeloische Modell-Zelllinie HL-60 in undifferenziertem und differenziertem Zustand gewählt, da hier in vorangegangenen Experimenten die meisten Proteine identifiziert werden konnten.

Die insgesamt als signifikant identifizierten interagierenden Proteine bestätigten bereits bekannte Transkriptionsfaktoren wie Sp1 und Sp3, lieferten allerdings zusätzlich eine Reihe von neuen potentiellen Regulatoren des proximalen ALOX5-Promoters. Einige der am häufigsten und wiederholt detektierten Proteine umfassten neben den bereits erwähnten Transkriptionsfaktoren der Sp-Familie weitere Faktoren der homologen KLF-Familie, die hierbei erstmals als Interaktoren der GC-reichen Promotersequenzen identifiziert wurden. Sp1 und Sp3 wurden als aktivierende, respektive mögliche inhibierende (Sp3) Interaktoren bestätigt, zudem konnte als weiterer neuer Faktor derselben Familie Sp2 mehrfach angereichert detektiert werden, was ein Zusammenspiel der Mitglieder der Transkriptionsfaktor-Familie nahelegt. Eine Interaktion mit den mittels Kernpromoter-enthaltenden Sequenz angereicherten Faktoren KLF13 und KLF16 scheint plausibel, da auch diese in der Lage sind, mit den Proteinen der Sp-Familie um deren Bindungsstelle zu konkurrieren und legt den Schluss nahe, dass beide Transkriptionsfaktoren im Bereich der einfachen und kumulativen GC-Boxen binden. Als weiteres Mitglied der KLF-Familie wurde in den Promoter-proximalen Regionen KLF5 identifiziert, dem besondere Bedeutung zukommt, da KLF5 in bereits veröffentlichten Studien als Effektorprotein im TGFβ-Signaltransduktionsweg charakterisiert wurde. KLF5 und KLF13 sind zudem bekanntermaßen an verschiedenen Mechanismen der Regulation inflammatorischer Prozesse beteiligt, was einen Zusammenhang im physiologischen Geschehen der 5-LO plausibel macht.

Eine weitere Transkriptionsfaktor-Familie von Interesse, deren Proteine ebenfalls in den Promoter-proximalen Sequenzbereichen als signifikante Interaktoren identifiziert wurden, findet sich in den CCAAT/enhancer binding proteins (C/EBPs), die im Entzündungsgeschehen gleichermaßen eine essentielle Rolle spielen. Hier wurden vor allem C/EBPα, C/EBPβ und C/EBPε in undifferenzierten und differenzierten HL-60 Zellen angereichert, wobei C/EBPα und C/EBPβ gleichzeitig nur in differenzierten HL-60 Zellen gefunden wurden. Das Heterodimer C/EBPα/β konnte in diesem Zusammenhang bereits als verantwortlicher Komplex der TGFβ/1,25$(OH)_2D_3$-induzierten Hochregulation der CD14-Transkription identifiziert werden, was mögliche Analogien der TGFβ/1,25$(OH)_2D_3$-abhängigen 5-LO Expression aufzeigt, die in kommenden Studien charakterisiert werden sollten. Unterstützende Evidenz der Bindung an den proximalen ALOX5 Promoter findet sich außerdem in publizierten Chip-Seq Datensätzen, was die Vermutung nahelegt, dass die Transkriptionsfaktoren der C/EBP-Familie einen potentiellen Anteil an der 5-LO Transkriptionsregulation aufweisen.

Zu den bereits aufgeführten Transkriptionsfaktoren konnten weitere transkriptionell aktive Proteine repetitiv identifiziert werden, die vor allem das Myc-assoziierte Zinkfingerprotein (MAZ), als auch die heterogenen nukleären Ribonukleoproteine (hnRNPs) hnRNP D und hnRNP K umschließen. MAZ stellt hierbei einen aktivierenden Transkriptionsfaktor mit Affinität zu Guanin-reichen Sequenzen dar, dessen Zusammenspiel mit Sp1 bereits in Studien gezeigt werden konnte. Zusätzlich sind sowohl MAZ, als auch beide hnRNPs an der Regulation G-Quadruplex-abhängiger Transkription verschiedener Gene beteiligt, was die Regulation der 5-LO-Expression erstmals mit sekundären DNA Strukturen in Verbindung bringt.

In einem zweiten Teil der Arbeit wurde aus diesem Grund die Möglichkeit der Ausbildung sekundärer DNA-Strukturen innerhalb des proximalen 5-LO-Promoters untersucht, die sich in erster Linie aus der Funktion einiger identifizierter Proteine ergab, die im Erkennen und Entwinden von G-Quadruplexen (G4-DNA Strukturen) eine Rolle innehaben. G-Quadruplexe können sich in Guanin-reichen DNA Sequenzen ausbilden und stellen potente Regulatoren der Transkription ihrer Target-Gene dar. Die meisten dieser Target-Gene sind Proto-Onkogene, weshalb G-Quadruplex Strukturen in jüngerer Zeit vermehrt als potentielle pharmazeutische Targets in Betracht gezogen werden. G-Quadruplexe finden sich außerdem in der Regulation von Telomeren.

Einige der angereicherten Proteine des Kernpromoter DNA-Pulldowns stehen in direktem funktionellen Zusammenhang der G4-DNA-vermittelten Genexpressionsregulation, wie

beispielsweise die Quadruplex-entwindenden Helikasen DHX36 und BLM, als auch die bereits erwähnten hnRNP D und K, sowie der Transkriptionsfaktor MAZ, welcher sowohl stabilisierende als auch entwindende Effekte auf G4-DNA aufweisen kann. Seine Quadruplex-entwindende Funktion konnte im Falle des onkogenen h-RAS-Promoters, der analoge Charakteristika zum ALOX5-Promoter aufweist, dessen Sp1-induzierte transkriptionelle Aktivierung vermitteln. Da der proximale 5-LO-Promoter in seinen fünffach-GC-Boxen potentielle Sequenzen zur Ausbildung von Quadruplex-Strukturen beinhaltet, wurde eine Oligonukleotidsequenz mit entsprechenden Tandem-GC-Boxen zur Untersuchung möglicher Sekundärstrukturen verwendet. Zur Analyse wurden sowohl immunologische (ELISA), als auch spektroskopische (UV-VIS, CD) Methoden gewählt. Die erhaltenen Ergebnisse konnten erstmalig die tatsächliche *in vitro* Ausbildung von G-Quadruplexen in der proximalen Sequenz des ALOX5-Promoters nachweisen. Um funktionelle Aussagen über mögliche regulatorische Einflüsse auf die 5-LO Expression vorliegender G4-DNA zu erhalten, wurden differenzierte HL-60 Zellen mit zwei unterschiedlichen G-Quadruplex stabilisierenden Agenzien behandelt. Das verwendete Porphyrin-Analogon (TMPyP4) zeigte hierbei nach Zellbehandlung weder Effekte auf die Expression der 5-LO-mRNA noch auf die Proteinexpression. Ein weiterer G4-DNA-Stabilisator, Pyridostatin, führte nach zellulärer Behandlung allerdings zur signifikanten Reduktion der Proteinexpression. Die gemischte Ergebnislage sollte als Anlass genommen werden, weitere Untersuchungen durchzuführen, um schlussendlich eine sichere Aussage über die potentielle Regulation der 5-LO Expression mittels G-Quadruplexen zu erhalten.

Abschließend liefert die vorliegende Arbeit grundlegende Vorarbeiten und einen Ausgangspunkt für weitere notwendige Studien, um sich mit der Charakterisierung der identifizierten Proteine in ihrer funktionellen Rolle in der transkriptionellen Regulation der 5-LO zu beschäftigen. Die angewandten Methoden verifizierten bereits bekannte ALOX5-Interaktoren, lieferten darüber hinaus jedoch auch eine hohe Zahl neuer Transkriptionsfaktoren, die die bisherige Kenntnis der Enzymexpression deutlich erweitern. Zusätzlich wurden G-Quadruplex als sekundäre DNA-Strukturen zum ersten Mal im proximalen Promoterbereich nachgewiesen und als möglicher Mechanismus der 5-LO Genexpression in Betracht gezogen.

8. References

[1] R. A. Lewis, K. F. Austen, and R. J. Soberman, 'Leukotrienes and Other Products of the 5-Lipoxygenase Pathway', *N. Engl. J. Med.*, vol. 323, no. 10, pp. 645–655, Sep. 1990.

[2] B. Samuelsson, 'Leukotrienes: mediators of immediate hypersensitivity reactions and inflammation', *Science*, vol. 220, no. 4597, pp. 568–575, May 1983.

[3] M. Peters-Golden, 'Cell biology of the 5-lipoxygenase pathway', *Am. J. Respir. Crit. Care Med.*, vol. 157, no. 6 Pt 2, pp. S227-231; discussion S231-232, S247-248, Jun. 1998.

[4] A. W. Ford-Hutchinson, M. Gresser, and R. N. Young, '5-Lipoxygenase', *Annu. Rev. Biochem.*, vol. 63, no. 1, pp. 383–417, 1994.

[5] C. N. Serhan, M. Hamberg, and B. Samuelsson, 'Lipoxins: novel series of biologically active compounds formed from arachidonic acid in human leukocytes', *Proc. Natl. Acad. Sci.*, vol. 81, no. 17, pp. 5335–5339, Sep. 1984.

[6] O. Rådmark, 'Arachidonate 5-lipoxygenase', *Prostaglandins Other Lipid Mediat.*, vol. 68–69, pp. 211–234, Aug. 2002.

[7] W. S. Powell and J. Rokach, 'Biochemistry, biology and chemistry of the 5-lipoxygenase product 5-oxo-ETE', *Prog. Lipid Res.*, vol. 44, no. 2, pp. 154–183, Mar. 2005.

[8] S. E. Dahlén *et al.*, 'Leukotrienes promote plasma leakage and leukocyte adhesion in postcapillary venules: in vivo effects with relevance to the acute inflammatory response', *Proc. Natl. Acad. Sci.*, vol. 78, no. 6, pp. 3887–3891, Jun. 1981.

[9] E. J. Goetzl and W. C. Pickett, 'The human PMN leukocyte chemotactic activity of complex hydroxy-eicosatetraenoic acids (HETEs).', *J. Immunol.*, vol. 125, no. 4, pp. 1789–1791, Oct. 1980.

[10] M. Peters-Golden and T. G. Brock, '5-Lipoxygenase and FLAP', *Prostaglandins Leukot. Essent. Fatty Acids*, vol. 69, no. 2, pp. 99–109, Aug. 2003.

[11] D. E. Sloniewsky *et al.*, 'Leukotriene D4 Activates Alveolar Epithelial Na,K-ATPase and Increases Alveolar Fluid Clearance', *Am. J. Respir. Crit. Care Med.*, vol. 169, no. 3, pp. 407–412, Feb. 2004.

[12] P. Rubin and K. W. Mollison, 'Pharmacotherapy of diseases mediated by 5-lipoxygenase pathway eicosanoids', *Prostaglandins Other Lipid Mediat.*, vol. 83, no. 3, pp. 188–197, May 2007.

[13] M. Peters-Golden and R. W. Mcnish, 'Redistribution of 5-Lipoxygenase and Cytosolic Phospholipase A2 to the Nuclear Fraction upon Macrophage Activation', *Biochem. Biophys. Res. Commun.*, vol. 196, no. 1, pp. 147–153, Oct. 1993.

[14] J. W. Woods *et al.*, '5-lipoxygenase and 5-lipoxygenase-activating protein are localized in the nuclear envelope of activated human leukocytes.', *J. Exp. Med.*, vol. 178, no. 6, pp. 1935–1946, Dec. 1993.

[15] O. Rådmark, O. Werz, D. Steinhilber, and B. Samuelsson, '5-Lipoxygenase, a key enzyme for leukotriene biosynthesis in health and disease', *Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids*, vol. 1851, no. 4, pp. 331–339, Apr. 2015.

[16] J. Fauler, C. Neumann, D. Tsikas, and J. C. Frölich, 'Enhanced synthesis of cysteinyl leukotrienes in atopic dermatitis', *Br. J. Dermatol.*, vol. 128, no. 6, pp. 627–630, 1993.

[17] E. M. Davidson, S. A. Rae, and M. J. Smith, 'Leukotriene B4, a mediator of inflammation present in synovial fluid in rheumatoid arthritis.', *Ann. Rheum. Dis.*, vol. 42, no. 6, pp. 677–679, Dec. 1983.

[18] J. Chu and D. Praticò, '5-lipoxygenase as an endogenous modulator of amyloid beta formation in vivo', *Ann. Neurol.*, vol. 69, no. 1, pp. 34–46, 2011.

[19] J. Ghosh and C. E. Myers, 'Inhibition of arachidonate 5-lipoxygenase triggers massive apoptosis in human prostate cancer cells', *Proc. Natl. Acad. Sci.*, vol. 95, no. 22, pp. 13182–13187, Oct. 1998.

[20] A. Hoque *et al.*, 'Increased 5-lipoxygenase expression and induction of apoptosis by its inhibitors in esophageal cancer: a potential target for prevention', *Carcinogenesis*, vol. 26, no. 4, pp. 785–791, Apr. 2005.

[21] L. G. Melstrom *et al.*, 'Overexpression of 5-Lipoxygenase in Colon Polyps and Cancer and the Effect of 5-LOX Inhibitors In vitro and in a Murine Model', *Clin. Cancer Res.*, vol. 14, no. 20, pp. 6525–6530, Oct. 2008.

[22] X.-Z. Ding, P. Iversen, M. W. Cluck, J. A. Knezetic, and T. E. Adrian, 'Lipoxygenase Inhibitors Abolish Proliferation of Human Pancreatic Cancer Cells', *Biochem. Biophys. Res. Commun.*, vol. 261, no. 1, pp. 218–223, Jul. 1999.

[23] O. Rådmark, O. Werz, D. Steinhilber, and B. Samuelsson, '5-Lipoxygenase: regulation of expression and enzyme activity', *Trends Biochem. Sci.*, vol. 32, no. 7, pp. 332–341, Jul. 2007.

[24] M. Brungs, O. Radmark, B. Samuelsson, and D. Steinhilber, 'On the Induction of 5-Lipoxygenase Expression and Activity in HL-60 Cells: Effects of Vitamin D3, Retinoic Acid, DMSO and TGFβ', *Biochem. Biophys. Res. Commun.*, vol. 205, no. 3, pp. 1572–1580, Dec. 1994.

[25] O. Rådmark and B. Samuelsson, 'Regulation of the activity of 5-lipoxygenase, a key enzyme in leukotriene biosynthesis', *Biochem. Biophys. Res. Commun.*, vol. 396, no. 1, pp. 105–110, May 2010.

[26] K. Ochi, T. Yoshimoto, S. Yamamoto, K. Taniguchi, and T. Miyamoto, 'Arachidonate 5-lipoxygenase of guinea pig peritoneal polymorphonuclear leukocytes. Activation by adenosine 5'-triphosphate.', *J. Biol. Chem.*, vol. 258, no. 9, pp. 5754–5758, Oct. 1983.

[27] C. A. Rouzer and B. Samuelsson, 'On the nature of the 5-lipoxygenase reaction in human leukocytes: enzyme purification and requirement for multiple stimulatory factors', *Proc. Natl. Acad. Sci.*, vol. 82, no. 18, pp. 6040–6044, Sep. 1985.

[28] R. a. F. Dixon *et al.*, 'Requirement of a 5-lipoxygenase-activating protein for leukotriene synthesis', *Nature*, vol. 343, no. 6255, p. 282, Jan. 1990.

[29] M. Luo, S. M. Jones, N. Flamand, D. M. Aronoff, M. Peters-Golden, and T. G. Brock, 'Phosphorylation by Protein Kinase A Inhibits Nuclear Import of 5-Lipoxygenase', *J. Biol. Chem.*, vol. 280, no. 49, pp. 40609–40616, Sep. 2005.

[30] O. Werz *et al.*, 'Extracellular signal-regulated kinases phosphorylate 5-lipoxygenase and stimulate 5-lipoxygenase product formation in leukocytes', *FASEB J.*, vol. 16, no. 11, pp. 1441–1443, Jul. 2002.

[31] O. Werz, J. Klemm, B. Samuelsson, and O. Rådmark, '5-Lipoxygenase is phosphorylated by p38 kinase-dependent MAPKAP kinases', *Proc. Natl. Acad. Sci.*, vol. 97, no. 10, pp. 5261–5266, May 2000.

[32] S. Markoutsa, D. Sürün, M. Karas, B. Hofmann, D. Steinhilber, and B. L. Sorg, 'Analysis of 5-lipoxygenase phosphorylation on molecular level by MALDI-MS', *FEBS J.*, pp. 1931–1947, 2014.

[33] C. D. Funk, S. Hoshiko, T. Matsumoto, O. Rdmark, and B. Samuelsson, 'Characterization of the human 5-lipoxygenase gene', *Proc. Natl. Acad. Sci.*, vol. 86, no. 8, pp. 2587–2591, Apr. 1989.

[34] S. Hoshiko, O. Rådmark, and B. Samuelsson, 'Characterization of the human 5-lipoxygenase gene promoter.', *Proc. Natl. Acad. Sci.*, vol. 87, no. 23, pp. 9073–9077, Dec. 1990.

[35] W. S. Dynan, 'Promoters for housekeeping genes', *Trends Genet.*, vol. 2, pp. 196–197, Jan. 1986.

[36] A.-K. Häfner, A. S. Kahnt, and D. Steinhilber, 'Beyond leukotriene formation—The noncanonical functions of 5-lipoxygenase', *Prostaglandins Other Lipid Mediat.*, vol. 142, pp. 24–32, Jun. 2019.

[37] E. S. Silverman *et al.*, 'Egr-1 and Sp1 Interact Functionally with the 5-Lipoxygenase Promoter and Its Naturally Occurring Mutants', *Am. J. Respir. Cell Mol. Biol.*, vol. 19, no. 2, pp. 316–323, Aug. 1998.

[38] G. Suske, E. Bruford, and S. Philipsen, 'Mammalian SP/KLF transcription factors: Bring in the family', *Genomics*, vol. 85, no. 5, pp. 551–556, May 2005.

[39] G. Suske, 'The Sp-family of transcription factors', *Gene*, vol. 238, no. 2, pp. 291–300, Oct. 1999.

[40] J. T. Kadonaga, K. R. Carner, F. R. Masiarz, and R. Tjian, 'Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain', *Cell*, vol. 51, no. 6, pp. 1079–1090, Dec. 1987.

[41] S. Chu and T. J. Ferro, 'Sp1: Regulation of gene expression by phosphorylation', *Gene*, vol. 348, pp. 1–11, Mar. 2005.

[42] S. P. Jackson and R. Tjian, 'O-glycosylation of eukaryotic transcription factors: Implications for mechanisms of transcriptional regulation', *Cell*, vol. 55, no. 1, pp. 125–133, Oct. 1988.
[43] K. K. Resendes and A. G. Rosmarin, 'Sp1 Control of Gene Expression in Myeloid Cells', *Crit. Rev. Eukaryot. Gene Expr.*, vol. 14, no. 3, 2004.
[44] J. T. Kadonaga, K. A. Jones, and R. Tjian, 'Promoter-specific activation of RNA polymerase II transcription by Sp1', *Trends Biochem. Sci.*, vol. 11, no. 1, pp. 20–23, Jan. 1986.
[45] D. Dishart *et al.*, 'GC-rich sequences in the 5-lipoxygenase gene promoter are required for expression in Mono Mac 6 cells, characterization of a novel Sp1 binding site', *Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids*, vol. 1738, no. 1, pp. 37–47, Dec. 2005.
[46] K. H. In *et al.*, 'Naturally occurring mutations in the human 5-lipoxygenase gene promoter that modify transcription factor binding and reporter gene transcription.', *J. Clin. Invest.*, vol. 99, no. 5, pp. 1130–1137, Mar. 1997.
[47] J. M. Drazen *et al.*, 'Pharmacogenetic association between *ALOX5* promoter genotype and the response to anti-asthma treatment', *Nat. Genet.*, vol. 22, no. 2, pp. 168–170, Jun. 1999.
[48] J. H. Dwyer *et al.*, 'Arachidonate 5-Lipoxygenase Promoter Genotype, Dietary Arachidonic Acid, and Atherosclerosis', *N. Engl. J. Med.*, vol. 350, no. 1, pp. 29–37, Jan. 2004.
[49] J. Uhl, N. Klan, M. Rose, K.-D. Entian, O. Werz, and D. Steinhilber, 'The 5-Lipoxygenase Promoter Is Regulated by DNA Methylation', *J. Biol. Chem.*, vol. 277, no. 6, pp. 4374–4379, Aug. 2002.
[50] A. P. Bird and A. P. Wolffe, 'Methylation-Induced Repression— Belts, Braces, and Chromatin', *Cell*, vol. 99, no. 5, pp. 451–454, Nov. 1999.
[51] C. Katryniok *et al.*, 'Role of DNA methylation and methyl-DNA binding proteins in the repression of 5-lipoxygenase promoter activity', *Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids*, vol. 1801, no. 1, pp. 49–57, Jan. 2010.
[52] N. Klan, S. Seuter, N. Schnur, M. Jung, and D. Steinhilber, 'Trichostatin A and Structurally Related Histone Deacetylase Inhibitors Induce 5-Lipoxygenase Promoter Activity', *Biol. Chem.*, vol. 384, no. 5, pp. 777–785, 2005.
[53] N. Schnur, S. Seuter, C. Katryniok, O. Rådmark, and D. Steinhilber, 'The histone deacetylase inhibitor trichostatin A mediates upregulation of 5-lipoxygenase promoter activity by recruitment of Sp1 to distinct GC-boxes', *Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids*, vol. 1771, no. 10, pp. 1271–1282, Oct. 2007.
[54] K. Ahmad *et al.*, 'Inhibition of class I HDACs abrogates the dominant effect of MLL-AF4 by activation of wild-type MLL', *Oncogenesis*, vol. 3, no. 11, p. e127, Nov. 2014.
[55] D. Härle, O. Rådmark, B. Samuelsson, and D. Steinhilber, 'Calcitriol and transforming growth factor-β upregulate 5-lipoxygenase mRNA expression by increasing gene transcription and mRNA maturation', *Eur. J. Biochem.*, vol. 254, no. 2, pp. 275–281, 1998.
[56] B. L. Sorg *et al.*, 'Analysis of the 5-lipoxygenase promoter and characterization of a vitamin D receptor binding site', *Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids*, vol. 1761, no. 7, pp. 686–697, Jul. 2006.
[57] S. Seuter, B. L. Sorg, and D. Steinhilber, 'The coding sequence mediates induction of 5-lipoxygenase expression by Smads3/4', *Biochem. Biophys. Res. Commun.*, vol. 348, no. 4, pp. 1403–1410, Oct. 2006.
[58] K. L. Stoffers, B. L. Sorg, S. Seuter, O. Rau, O. Rådmark, and D. Steinhilber, 'Calcitriol Upregulates Open Chromatin and Elongation Markers at Functional Vitamin D Response Elements in the Distal Part of the 5-Lipoxygenase Gene', *J. Mol. Biol.*, vol. 395, no. 4, pp. 884–896, Jan. 2010.
[59] S. Seuter, S. Väisänen, O. Rådmark, C. Carlberg, and D. Steinhilber, 'Functional characterization of vitamin D responding regions in the human 5-Lipoxygenase gene', *Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids*, vol. 1771, no. 7, pp. 864–872, Jul. 2007.
[60] K. Ahmad *et al.*, 'AF4 and AF4-MLL mediate transcriptional elongation of 5-lipoxygenase mRNA by 1, 25-dihydroxyvitamin D3', *Oncotarget*, vol. 6, no. 28, pp. 25784–25800, Jul. 2015.
[61] M. J. Saul *et al.*, 'TGFβ/SMAD signalling modulates MLL and MLL-AF4 mediated 5-lipoxygenase promoter activation', *Prostaglandins Other Lipid Mediat.*, vol. 133, pp. 60–67, Nov. 2017.

[62] B. Gilbert *et al.*, '5-Lipoxygenase is a direct p53 target gene in humans', *Biochim. Biophys. Acta BBA - Gene Regul. Mech.*, vol. 1849, no. 8, pp. 1003–1016, Aug. 2015.
[63] J. D. Watson and F. H. Crick, 'Molecular structure of nucleic acids', *Nature*, vol. 171, no. 4356, pp. 737–738, 1953.
[64] International Human Genome Sequencing Consortium, 'Finishing the euchromatic sequence of the human genome', *Nature*, vol. 431, no. 7011, pp. 931–945, Oct. 2004.
[65] Y. Qin and L. H. Hurley, 'Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions', *Biochimie*, vol. 90, no. 8, pp. 1149–1171, Aug. 2008.
[66] G. W. Collie and G. N. Parkinson, 'The application of DNA and RNA G-quadruplexes to therapeutic medicines', *Chem. Soc. Rev.*, vol. 40, no. 12, pp. 5867–5892, 2011.
[67] R. H. Shafer and I. Smirnov, 'Biological aspects of DNA/RNA quadruplexes', *Biopolymers*, vol. 56, no. 3, pp. 209–227, 2000.
[68] G. Yuan, Q. Zhang, J. Zhou, and H. Li, 'Mass spectrometry of G-quadruplex DNA: Formation, recognition, property, conversion, and conformation', *Mass Spectrom. Rev.*, vol. 30, no. 6, pp. 1121–1142, 2011.
[69] M.-Y. Kim, M. Gleason-Guzman, E. Izbicka, D. Nishioka, and L. H. Hurley, 'The Different Biological Effects of Telomestatin and TMPyP4 Can Be Attributed to Their Selectivity for Interaction with Intramolecular or Intermolecular G-Quadruplex Structures', *Cancer Res.*, vol. 63, no. 12, pp. 3247–3256, Jun. 2003.
[70] T. Ou, Y. Lu, J. Tan, Z. Huang, K.-Y. Wong, and L. Gu, 'G-Quadruplexes: Targets in Anticancer Drug Design', *ChemMedChem*, vol. 3, no. 5, pp. 690–713, 2008.
[71] C. Sissi, B. Gatto, and M. Palumbo, 'The evolving world of protein-G-quadruplex recognition: A medicinal chemist's perspective', *Biochimie*, vol. 93, no. 8, pp. 1219–1230, Aug. 2011.
[72] E. H. Blackburn and J. W. Szostak, 'The molecular structure of centromeres and telomeres', *Annu. Rev. Biochem.*, vol. 53, no. 1, pp. 163–194, 1984.
[73] W. I. Sundquist and A. Klug, 'Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops', *Nature*, vol. 342, no. 6251, p. 825, Dec. 1989.
[74] Y. Wang and D. J. Patel, 'Guanine residues in d (T2AG3) and d (T2G4) form parallel-stranded potassium cation stabilized G-quadruplexes with anti glycosidic torsion angles in solution', *Biochemistry*, vol. 31, no. 35, pp. 8112–8119, 1992.
[75] J.-L. Mergny and C. Hélène, 'G-quadruplex DNA: A target for drug design', *Nat. Med.*, vol. 4, no. 12, pp. 1366–1367, Dec. 1998.
[76] D. Sun *et al.*, 'Inhibition of Human Telomerase by a G-Quadruplex-Interactive Compound', *J. Med. Chem.*, vol. 40, no. 14, pp. 2113–2116, Jul. 1997.
[77] R. T. Wheelhouse, D. Sun, H. Han, F. X. Han, and L. H. Hurley, 'Cationic porphyrins as telomerase inhibitors: the interaction of tetra-(N-methyl-4-pyridyl) porphine with quadruplex DNA', *J. Am. Chem. Soc.*, vol. 120, no. 13, pp. 3261–3262, 1998.
[78] A. Siddiqui-Jain, C. L. Grand, D. J. Bearss, and L. H. Hurley, 'Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription', *Proc. Natl. Acad. Sci.*, vol. 99, no. 18, pp. 11593–11598, Sep. 2002.
[79] J. Dai *et al.*, 'An Intramolecular G-Quadruplex Structure with Mixed Parallel/Antiparallel G-Strands Formed in the Human BCL-2 Promoter Region in Solution', *J. Am. Chem. Soc.*, vol. 128, no. 4, pp. 1096–1098, Feb. 2006.
[80] S. Cogoi and L. E. Xodo, 'G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription', *Nucleic Acids Res.*, vol. 34, no. 9, pp. 2536–2549, Jan. 2006.
[81] S. L. Palumbo, R. M. Memmott, D. J. Uribe, Y. Krotova-Khan, L. H. Hurley, and S. W. Ebbinghaus, 'A novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulator of promoter activity', *Nucleic Acids Res.*, vol. 36, no. 6, pp. 1755–1769, Apr. 2008.

[82] A. Membrino, S. Cogoi, E. B. Pedersen, and L. E. Xodo, 'G4-DNA Formation in the HRAS Promoter and Rational Design of Decoy Oligonucleotides for Cancer Therapy', *PLOS ONE*, vol. 6, no. 9, p. e24421, Sep. 2011.

[83] D. Sun *et al.*, 'The proximal promoter region of the human vascular endothelial growth factor gene has a G-quadruplex structure that can be targeted by G-quadruplex–interactive agents', *Mol. Cancer Ther.*, vol. 7, no. 4, pp. 880–889, Apr. 2008.

[84] K. W. Lim, L. Lacroix, D. J. E. Yue, J. K. C. Lim, J. M. W. Lim, and A. T. Phan, 'Coexistence of Two Distinct G-Quadruplex Conformations in the hTERT Promoter', *J. Am. Chem. Soc.*, vol. 132, no. 35, pp. 12331–12342, Sep. 2010.

[85] S. I. Chanu and S. Sarkar, 'The Paradox of c-Myc Proto-oncogene and its Diverse Functions', *Cell Dev. Biol.*, vol. 3, no. 3, pp. 1–3, 2014.

[86] C. V. Dang *et al.*, 'Function of the c-Myc Oncogenic Transcription Factor', *Exp. Cell Res.*, vol. 253, no. 1, pp. 63–77, Nov. 1999.

[87] T. Simonsson, M. Kubista, and P. Pecinka, 'DNA tetraplex formation in the control region of c-myc', *Nucleic Acids Res.*, vol. 26, no. 5, pp. 1167–1172, Mar. 1998.

[88] J. M. Adams and S. Cory, 'The Bcl-2 Protein Family: Arbiters of Cell Survival', *Science*, vol. 281, no. 5381, pp. 1322–1326, Aug. 1998.

[89] G. B. Baretton *et al.*, 'Apoptosis and immunohistochemical bcl-2 expression in colorectal adenomas and carcinomas: Aspects of carcinogenesis and prognostic significance', *Cancer*, vol. 77, no. 2, pp. 255–264, 1996.

[90] N. Nagesh, R. Buscaglia, J. M. Dettler, and E. A. Lewis, 'Studies on the Site and Mode of TMPyP4 Interactions with Bcl-2 Promoter Sequence G-Quadruplexes', *Biophys. J.*, vol. 98, no. 11, pp. 2628–2633, Jun. 2010.

[91] X.-D. Wang *et al.*, 'Turning off Transcription of the bcl-2 Gene by Stabilizing the bcl-2 Promoter Quadruplex with Quindoline Derivatives', *J. Med. Chem.*, vol. 53, no. 11, pp. 4390–4398, Jun. 2010.

[92] Y. Feng *et al.*, 'Stabilization of G-quadruplex DNA and inhibition of Bcl-2 expression by a pyridostatin analog', *Bioorg. Med. Chem. Lett.*, vol. 26, no. 7, pp. 1660–1663, Apr. 2016.

[93] D. R. Lowy and B. M. Willumsen, 'Function and regulation of ras', *Annu. Rev. Biochem.*, vol. 62, no. 1, pp. 851–891, 1993.

[94] S. Cogoi, M. Paramasivam, A. Membrino, K. K. Yokoyama, and L. E. Xodo, 'The KRAS Promoter Responds to Myc-associated Zinc Finger and Poly(ADP-ribose) Polymerase 1 Proteins, Which Recognize a Critical Quadruplex-forming GA-element', *J. Biol. Chem.*, vol. 285, no. 29, pp. 22003–22016, Jul. 2010.

[95] C. B. Thompson, P. B. Challoner, P. E. Neiman, and M. Groudine, 'Expression of the c-myb proto-oncogene during cellular proliferation', *Nature*, vol. 319, no. 6052, p. 374, Jan. 1986.

[96] D. J. Slamon *et al.*, 'Studies of the human c-myb gene and its product in human acute leukemias', *Science*, vol. 233, no. 4761, pp. 347–351, Jul. 1986.

[97] C. A. Griffin and S. B. Baylin, 'Expression of the c-myb Oncogene in Human Small Cell Lung Carcinoma', *Cancer Res.*, vol. 45, no. 1, pp. 272–275, Jan. 1985.

[98] D. Sun, K. Guo, J. J. Rusche, and L. H. Hurley, 'Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents', *Nucleic Acids Res.*, vol. 33, no. 18, pp. 6070–6080, Jan. 2005.

[99] M. C. U. Hammond-Kosack, B. Dobrinski, R. Lurz, K. Docherty, and M. W. Kilpatrick, 'The human insulin gene linked polymorphic region exhibits an altered DNA structure', *Nucleic Acids Res.*, vol. 20, no. 2, pp. 231–236, Jan. 1992.

[100] P. Catasti, X. Chen, R. K. Moyzis, E. M. Bradbury, and G. Gupta, 'Structure–Function Correlations of the Insulin-linked Polymorphic Region', *J. Mol. Biol.*, vol. 264, no. 3, pp. 534–545, Dec. 1996.

[101] Z. Yu *et al.*, 'ILPR G-Quadruplexes Formed in Seconds Demonstrate High Mechanical Stabilities', *J. Am. Chem. Soc.*, vol. 131, no. 5, pp. 1876–1882, Feb. 2009.

[102] R. De Armond, S. Wood, D. Sun, L. H. Hurley, and S. W. Ebbinghaus, 'Evidence for the Presence of a Guanine Quadruplex Forming Region within a Polypurine Tract of the Hypoxia Inducible Factor 1α Promoter', *Biochemistry*, vol. 44, no. 49, pp. 16341–16350, Dec. 2005.

[103] S. J. Welsh *et al.*, 'Inhibition of the hypoxia-inducible factor pathway by a G-quadruplex binding small molecule', *Sci. Rep.*, vol. 3, p. 2799, Sep. 2013.

[104] A. T. Phan, V. Kuryavyi, S. Burge, S. Neidle, and D. J. Patel, 'Structure of an Unprecedented G-Quadruplex Scaffold in the Human c-kit Promoter', *J. Am. Chem. Soc.*, vol. 129, no. 14, pp. 4386–4392, Apr. 2007.

[105] M. Bejugam, S. Sewitz, P. S. Shirude, R. Rodriguez, R. Shahid, and S. Balasubramanian, 'Trisubstituted Isoalloxazines as a New Class of G-Quadruplex Binding Ligands: Small Molecule Regulation of c-kit Oncogene Expression', *J. Am. Chem. Soc.*, vol. 129, no. 43, pp. 12926–12927, Oct. 2007.

[106] Y. Qin, E. M. Rezler, V. Gokhale, D. Sun, and L. H. Hurley, 'Characterization of the G-quadruplexes in the duplex nuclease hypersensitive element of the PDGF-A promoter and modulation of PDGF-A promoter activity by TMPyP4', *Nucleic Acids Res.*, vol. 35, no. 22, pp. 7698–7713, Dec. 2007.

[107] S. L. Palumbo, S. W. Ebbinghaus, and L. H. Hurley, 'Formation of a Unique End-to-End Stacked Pair of G-Quadruplexes in the hTERT Core Promoter with Implications for Inhibition of Telomerase by G-Quadruplex-Interactive Ligands', *J. Am. Chem. Soc.*, vol. 131, no. 31, pp. 10878–10891, Aug. 2009.

[108] J. Shklover, P. Weisman-Shomer, A. Yafe, and M. Fry, 'Quadruplex structures of muscle gene promoter sequences enhance in vivo MyoD-dependent gene expression', *Nucleic Acids Res.*, vol. 38, no. 7, pp. 2369–2377, Apr. 2010.

[109] R. Rodriguez *et al.*, 'Small-molecule–induced DNA damage identifies alternative DNA structures in human genes', *Nat. Chem. Biol.*, vol. 8, no. 3, pp. 301–310, Mar. 2012.

[110] S. Amrane, A. Kerkour, A. Bedrat, B. Vialet, M.-L. Andreola, and J.-L. Mergny, 'Topology of a DNA G-Quadruplex Structure Formed in the HIV-1 Promoter: A Potential Target for Anti-HIV Drug Development', *J. Am. Chem. Soc.*, vol. 136, no. 14, pp. 5249–5252, Apr. 2014.

[111] R. Perrone *et al.*, 'A Dynamic G-Quadruplex Region Regulates the HIV-1 Long Terminal Repeat Promoter', *J. Med. Chem.*, vol. 56, no. 16, pp. 6521–6530, Aug. 2013.

[112] A. Sengar *et al.*, 'Structure of a (3+1) hybrid G-quadruplex in the PARP1 promoter.', Dec. 2018.

[113] S. G. Zidanloo, A. Hosseinzadeh Colagar, H. Ayatollahi, and J.-B. Raoof, 'Downregulation of the WT1 gene expression via TMPyP4 stabilization of promoter G-quadruplexes in leukemia cells', *Tumor Biol.*, vol. 37, no. 7, pp. 9967–9977, Jul. 2016.

[114] P. B. Arimondo *et al.*, 'Interaction of human DNA topoisomerase I with G-quartet structures', *Nucleic Acids Res.*, vol. 28, no. 24, pp. 4832–4838, Dec. 2000.

[115] Y. Enokizono *et al.*, 'Structure of hnRNP D Complexed with Single-stranded Telomere DNA and Unfolding of the Quadruplex by Heterogeneous Nuclear Ribonucleoprotein D', *J. Biol. Chem.*, vol. 280, no. 19, pp. 18862–18870, May 2005.

[116] Q.-S. Zhang, L. Manche, R.-M. Xu, and A. R. Krainer, 'hnRNP A1 associates with telomere ends and stimulates telomerase activity', *RNA*, vol. 12, no. 6, pp. 1116–1128, Jan. 2006.

[117] Y. Enokizono *et al.*, 'Destruction of quadruplex by proteins, and its biological implications in replication and telomere maintenance', *Nucleic Acids Symp. Ser.*, vol. 3, no. 1, pp. 231–232, Sep. 2003.

[118] S. Cogoi, M. Paramasivam, B. Spolaore, and L. E. Xodo, 'Structural polymorphism within a regulatory element of the human KRAS promoter: formation of G4-DNA recognized by nuclear proteins', *Nucleic Acids Res.*, vol. 36, no. 11, pp. 3765–3780, Jun. 2008.

[119] L. Xodo, M. Paramasivam, A. Membrino, and S. Cogoi, 'Protein hnRNPA1 binds to a critical G-rich element of KRAS and unwinds G-quadruplex structures: implications in transcription', *Nucleic Acids Symp. Ser.*, vol. 52, no. 1, pp. 159–160, Sep. 2008.

[120] A. Brys and N. Maizels, 'LR1 regulates c-myc transcription in B-cell lymphomas', *Proc. Natl. Acad. Sci.*, vol. 91, no. 11, pp. 4915–4919, May 1994.

[121] V. González, K. Guo, L. Hurley, and D. Sun, 'Identification and Characterization of Nucleolin as a c-myc G-quadruplex-binding Protein', *J. Biol. Chem.*, vol. 284, no. 35, pp. 23622–23635, Aug. 2009.

[122] D. J. Uribe, K. Guo, Y.-J. Shin, and D. Sun, 'Heterogeneous Nuclear Ribonucleoprotein K and Nucleolin as Transcriptional Activators of the Vascular Endothelial Growth Factor Promoter through Interaction with Secondary DNA Structures', *Biochemistry*, vol. 50, no. 18, pp. 3796–3806, May 2011.

[123] V. S. Chambers, G. Marsico, J. M. Boutell, M. Di Antonio, G. P. Smith, and S. Balasubramanian, 'High-throughput sequencing of DNA G-quadruplex structures in the human genome', *Nat. Biotechnol.*, vol. 33, no. 8, pp. 877–881, Aug. 2015.

[124] J. L. Huppert, A. Bugaut, S. Kumari, and S. Balasubramanian, 'G-quadruplexes: the beginning and end of UTRs', *Nucleic Acids Res.*, vol. 36, no. 19, pp. 6260–6268, Nov. 2008.

[125] A. K. Todd and S. Neidle, 'The relationship of potential G-quadruplex sequences in cis -upstream regions of the human genome to SP1-binding elements', *Nucleic Acids Res.*, vol. 36, no. 8, pp. 2700–2704, May 2008.

[126] P. Kumar, V. K. Yadav, A. Baral, P. Kumar, D. Saha, and S. Chowdhury, 'Zinc-finger transcription factors are associated with guanine quadruplex motifs in human, chimpanzee, mouse and rat promoters genome-wide', *Nucleic Acids Res.*, vol. 39, no. 18, pp. 8005–8016, Oct. 2011.

[127] A. Lew, W. J. Rutter, and G. C. Kennedy, 'Unusual DNA structure of the diabetes susceptibility locus IDDM2 and its effect on transcription by the insulin promoter factor Pur-1/MAZ', *Proc. Natl. Acad. Sci.*, vol. 97, no. 23, pp. 12508–12512, Nov. 2000.

[128] J. Álvaro-Blanco *et al.*, 'MAZ induces MYB expression during the exit from quiescence via the E2F site in the MYB promoter', *Nucleic Acids Res.*, vol. 45, no. 17, pp. 9960–9975, Sep. 2017.

[129] C. Leroy, D. Manen, R. Rizzoli, M. Lombes, and C. Silve, 'Functional importance of Myc-associated zinc finger protein for the human parathyroid hormone (PTH)/PTH-related peptide receptor-1 P2 promoter constitutive activity', *J. Mol. Endocrinol.*, vol. 32, no. 1, pp. 99–114, 2004.

[130] J. Song, H. Ugai, I. Kanazawa, K. Sun, and K. K. Yokoyama, 'Independent Repression of a GC-rich Housekeeping Gene by Sp1 and MAZ Involves the Same cis-Elements', *J. Biol. Chem.*, vol. 276, no. 23, pp. 19897–19904, Aug. 2001.

[131] H. Sun, J. K. Karow, I. D. Hickson, and N. Maizels, 'The Bloom's Syndrome Helicase Unwinds G4 DNA', *J. Biol. Chem.*, vol. 273, no. 42, pp. 27587–27592, Oct. 1998.

[132] A. S. Kamath-Loeb, J.-C. Shen, L. A. Loeb, and M. Fry, 'Werner Syndrome Protein II. CHARACTERIZATION OF THE INTEGRAL 3' → 5' DNA EXONUCLEASE', *J. Biol. Chem.*, vol. 273, no. 51, pp. 34145–34150, Dec. 1998.

[133] M. Fry and L. A. Loeb, 'Human Werner Syndrome DNA Helicase Unwinds Tetrahelical Structures of the Fragile X Syndrome Repeat Sequence d(CGG) n', *J. Biol. Chem.*, vol. 274, no. 18, pp. 12797–12802, Apr. 1999.

[134] P. L. Opresko *et al.*, 'POT1 Stimulates RecQ Helicases WRN and BLM to Unwind Telomeric DNA Substrates', *J. Biol. Chem.*, vol. 280, no. 37, pp. 32069–32080, Sep. 2005.

[135] J.-L. Li *et al.*, 'Inhibition of the Bloom's and Werner's Syndrome Helicases by G-Quadruplex Interacting Ligands', *Biochemistry*, vol. 40, no. 50, pp. 15194–15202, Dec. 2001.

[136] J. P. Vaughn *et al.*, '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.*, vol. 280, no. 46, pp. 38117–38120, Nov. 2005.

[137] S. D. Creacy, E. D. Routh, F. Iwamoto, Y. Nagamine, S. A. Akman, and J. P. Vaughn, '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.*, vol. 283, no. 50, pp. 34626–34634, Dec. 2008.

[138] W. Huang *et al.*, 'Yin Yang 1 contains G-quadruplex structures in its promoter and 5'-UTR and its expression is modulated by G4 resolvase 1', *Nucleic Acids Res.*, vol. 40, no. 3, pp. 1033–1049, Feb. 2012.

[139] E. P. Booy *et al.*, 'The RNA helicase RHAU (DHX36) suppresses expression of the transcription factor PITX1', *Nucleic Acids Res.*, vol. 42, no. 5, pp. 3346–3361, Mar. 2014.
[140] R. Tippana, H. Hwang, P. L. Opresko, V. A. Bohr, and S. Myong, 'Single-molecule imaging reveals a common mechanism shared by G-quadruplex–resolving helicases', *Proc. Natl. Acad. Sci.*, vol. 113, no. 30, pp. 8448–8453, Jul. 2016.
[141] J. Cuesta, M. A. Read, and S. Neidle, 'The design of G-quadruplex ligands as telomerase inhibitors', *Mini Rev. Med. Chem.*, vol. 3, no. 1, pp. 11–21, 2003.
[142] E. Izbicka *et al.*, 'Effects of Cationic Porphyrins as G-Quadruplex Interactive Agents in Human Tumor Cells', *Cancer Res.*, vol. 59, no. 3, pp. 639–644, Feb. 1999.
[143] Y. Mikami-Terao, M. Akiyama, Y. Yuza, T. Yanagisawa, O. Yamada, and H. Yamada, 'Antitumor activity of G-quadruplex–interactive agent TMPyP4 in K562 leukemic cells', *Cancer Lett.*, vol. 261, no. 2, pp. 226–234, Mar. 2008.
[144] C. Wei, G. Jia, J. Yuan, Z. Feng, and C. Li, 'A Spectroscopic Study on the Interactions of Porphyrin with G-Quadruplex DNAs', *Biochemistry*, vol. 45, no. 21, pp. 6681–6691, May 2006.
[145] A. Arora and S. Maiti, 'Effect of Loop Orientation on Quadruplex−TMPyP4 Interaction', *J. Phys. Chem. B*, vol. 112, no. 27, pp. 8151–8159, Jul. 2008.
[146] A. Bugaut and S. Balasubramanian, '5'-UTR RNA G-quadruplexes: translation regulation and targeting', *Nucleic Acids Res.*, vol. 40, no. 11, pp. 4727–4741, Jun. 2012.
[147] R. Rodriguez, S. Müller, J. A. Yeoman, C. Trentesaux, J.-F. Riou, and S. Balasubramanian, 'A Novel Small Molecule That Alters Shelterin Integrity and Triggers a DNA-Damage Response at Telomeres', *J. Am. Chem. Soc.*, vol. 130, no. 47, pp. 15758–15759, Nov. 2008.
[148] J. Husby, A. K. Todd, J. A. Platts, and S. Neidle, 'Small-molecule G-quadruplex interactions: Systematic exploration of conformational space using multiple molecular dynamics', *Biopolymers*, vol. 99, no. 12, pp. 989–1005, 2013.
[149] S. Müller *et al.*, 'Pyridostatin analogues promote telomere dysfunction and long-term growth inhibition in human cancer cells', *Org. Biomol. Chem.*, vol. 10, no. 32, pp. 6537–6546, 2012.
[150] K. Shin-ya *et al.*, 'Telomestatin, a Novel Telomerase Inhibitor from Streptomyces anulatus', *J. Am. Chem. Soc.*, vol. 123, no. 6, pp. 1262–1263, Feb. 2001.
[151] W. Li, P. Wu, T. Ohmichi, and N. Sugimoto, 'Characterization and thermodynamic properties of quadruplex/duplex competition', *FEBS Lett.*, vol. 526, no. 1–3, pp. 77–81, 2002.
[152] G. Biffi, D. Tannahill, J. McCafferty, and S. Balasubramanian, 'Quantitative visualization of DNA G-quadruplex structures in human cells', *Nat. Chem.*, vol. 5, no. 3, pp. 182–186, Mar. 2013.
[153] F. Rosu, V. Gabelica, C. Houssier, P. Colson, and E. D. Pauw, 'Triplex and quadruplex DNA structures studied by electrospray mass spectrometry', *Rapid Commun. Mass Spectrom.*, vol. 16, no. 18, pp. 1729–1736, 2002.
[154] E. S. Baker, S. L. Bernstein, V. Gabelica, E. De Pauw, and M. T. Bowers, 'G-quadruplexes in telomeric repeats are conserved in a solvent-free environment', *Int. J. Mass Spectrom.*, vol. 253, no. 3, pp. 225–237, Jul. 2006.
[155] R. Ferreira, A. Marchand, and V. Gabelica, 'Mass spectrometry and ion mobility spectrometry of G-quadruplexes. A study of solvent effects on dimer formation and structural transitions in the telomeric DNA sequence d(TAGGGTTAGGGT)', *Methods*, vol. 57, no. 1, pp. 56–63, May 2012.
[156] J. C. Venter *et al.*, 'The Sequence of the Human Genome', *Science*, vol. 291, no. 5507, pp. 1304–1351, Feb. 2001.
[157] B. T. Chait, 'Mass Spectrometry: Bottom-Up or Top-Down?', *Science*, vol. 314, no. 5796, pp. 65–66, Oct. 2006.
[158] N. L. Kelleher, H. Y. Lin, G. A. Valaskovic, D. J. Aaserud, E. K. Fridriksson, and F. W. McLafferty, 'Top Down versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry', *J. Am. Chem. Soc.*, vol. 121, no. 4, pp. 806–812, Feb. 1999.
[159] B. Bogdanov and R. D. Smith, 'Proteomics by FTICR mass spectrometry: Top down and bottom up', *Mass Spectrom. Rev.*, vol. 24, no. 2, pp. 168–200, 2005.
[160] X. Han, A. Aslanian, and J. R. Yates, 'Mass spectrometry for proteomics', *Curr. Opin. Chem. Biol.*, vol. 12, no. 5, pp. 483–490, Oct. 2008.

[161] A. Moradian, A. Kalli, M. J. Sweredoski, and S. Hess, 'The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications', *PROTEOMICS*, vol. 14, no. 4–5, pp. 489–497, 2014.
[162] S. Brunet, P. Thibault, E. Gagnon, P. Kearney, J. J. M. Bergeron, and M. Desjardins, 'Organelle proteomics: looking at less to see more', *Trends Cell Biol.*, vol. 13, no. 12, pp. 629–638, Dec. 2003.
[163] L. Tuli and H. W. Ressom, 'LC–MS Based Detection of Differential Protein Expression', *J. Proteomics Bioinform.*, vol. 2, pp. 416–438, Oct. 2009.
[164] T. M. Annesley, 'Ion Suppression in Mass Spectrometry', *Clin. Chem.*, vol. 49, no. 7, pp. 1041–1044, Jul. 2003.
[165] P. Fang *et al.*, 'Controlling nonspecific trypsin cleavages in LC-MS/MS-based shotgun proteomics using optimized experimental conditions', *Analyst*, vol. 140, no. 22, pp. 7613–7621, 2015.
[166] P. Giansanti, L. Tsiatsiani, T. Y. Low, and A. J. R. Heck, 'Six alternative proteases for mass spectrometry–based proteomics beyond trypsin', *Nat. Protoc.*, vol. 11, no. 5, pp. 993–1006, May 2016.
[167] J. S. Cottrell, 'Protein identification by peptide mass fingerprinting.', *Pept. Res.*, vol. 7, no. 3, pp. 115–124, 1994.
[168] W. J. Henzel, C. Watanabe, and J. T. Stults, 'Protein identification: the origins of peptide mass fingerprinting', *J. Am. Soc. Mass Spectrom.*, vol. 14, no. 9, pp. 931–942, Sep. 2003.
[169] D. J. C. Pappin, P. Hojrup, and A. J. Bleasby, 'Rapid identification of proteins by peptide-mass fingerprinting', *Curr. Biol.*, vol. 3, no. 6, pp. 327–332, Jun. 1993.
[170] M. Mann, R. C. Hendrickson, and A. Pandey, 'Analysis of Proteins and Proteomes by Mass Spectrometry', *Annu. Rev. Biochem.*, vol. 70, no. 1, pp. 437–473, 2001.
[171] M. Barthelery, U. Salli, and K. E. Vrana, 'Nuclear Proteomics and Directed Differentiation of Embryonic Stem Cells', *Stem Cells Dev.*, vol. 16, no. 6, pp. 905–920, Nov. 2007.
[172] H. Gadgil, L. A. Jurado, and H. W. Jarrett, 'DNA Affinity Chromatography of Transcription Factors', *Anal. Biochem.*, vol. 290, no. 2, pp. 147–178, Mar. 2001.
[173] J. Simicevic and B. Deplancke, 'Transcription factor proteomics—Tools, applications, and challenges', *PROTEOMICS*, vol. 17, no. 3–4, p. 1600317, 2017.
[174] A. Tacheny, M. Dieu, T. Arnould, and P. Renard, 'Mass spectrometry-based identification of proteins interacting with nucleic acids', *J. Proteomics*, vol. 94, pp. 89–109, Dec. 2013.
[175] B. Dey *et al.*, 'DNA–protein interactions: methods for detection and analysis', *Mol. Cell. Biochem.*, vol. 365, no. 1, pp. 279–299, Jun. 2012.
[176] L. M. Hellman and M. G. Fried, 'Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions', *Nat. Protoc.*, vol. 2, no. 8, pp. 1849–1861, Aug. 2007.
[177] D. Jing, J. Agnew, W. F. Patton, J. Hendrickson, and J. M. Beechem, 'A sensitive two-color electrophoretic mobility shift assay for detecting both nucleic acids and protein in gels', *PROTEOMICS*, vol. 3, no. 7, pp. 1172–1180, 2003.
[178] A. M. Khoury Christianson and F. C. Kafatos, 'Antibody detection of protein complexes bound to DNA.', *Nucleic Acids Res.*, vol. 21, no. 18, pp. 4416–4417, Sep. 1993.
[179] J. T. Kadonaga and R. Tjian, 'Affinity purification of sequence-specific DNA binding proteins', *Proc. Natl. Acad. Sci.*, vol. 83, no. 16, pp. 5889–5893, Aug. 1986.
[180] H. Im, J. A. Grass, K. D. Johnson, M. E. Boyer, J. Wu, and E. H. Bresnick, 'Measurement of Protein-DNA Interactions In Vivo by Chromatin Immunoprecipitation', in *Signal Transduction Protocols*, R. C. Dickson and M. D. Mendenhall, Eds. Totowa, NJ: Humana Press, 2004, pp. 129–146.
[181] R. Aebersold and M. Mann, 'Mass spectrometry-based proteomics', *Nature*, vol. 422, pp. 198–207, Mar. 2003.
[182] F. Butter *et al.*, 'Proteome-Wide Analysis of Disease-Associated SNPs That Show Allele-Specific Transcription Factor Binding', *PLOS Genet.*, vol. 8, no. 9, p. e1002982, Sep. 2012.

[183] N. C. Hubner, L. N. Nguyen, N. C. Hornig, and H. G. Stunnenberg, 'A Quantitative Proteomics Tool To Identify DNA–Protein Interactions in Primary Cells or Blood', *J. Proteome Res.*, vol. 14, no. 2, pp. 1315–1329, Feb. 2015.

[184] M. Yaneva and P. Tempst, 'Affinity Capture of Specific DNA-Binding Proteins for Mass Spectrometric Identification', *Anal. Chem.*, vol. 75, no. 23, pp. 6437–6448, Dec. 2003.

[185] B. L. Jutras, A. Verma, and B. Stevenson, 'Identification of Novel DNA-Binding Proteins Using DNA-Affinity Chromatography/Pull Down', *Curr. Protoc. Microbiol.*, vol. 24, no. 1, pp. 1F.1.1-1F.1.13, 2012.

[186] L. Trinkle-Mulcahy *et al.*, 'Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes', *J. Cell Biol.*, vol. 183, no. 2, pp. 223–239, Oct. 2008.

[187] L. Trinkle-Mulcahy, 'Resolving protein interactions and complexes by affinity purification followed by label-based quantitative mass spectrometry', *PROTEOMICS*, vol. 12, no. 10, pp. 1623–1638, 2012.

[188] M. Vermeulen, N. C. Hubner, and M. Mann, 'High confidence determination of specific protein–protein interactions using quantitative mass spectrometry', *Curr. Opin. Biotechnol.*, vol. 19, no. 4, pp. 331–337, Aug. 2008.

[189] J. Cox and M. Mann, 'Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology', *Annu. Rev. Biochem.*, vol. 80, no. 1, pp. 273–299, 2011.

[190] H. C. Eberl, M. Mann, and M. Vermeulen, 'Quantitative Proteomics for Epigenetics', *ChemBioChem*, vol. 12, no. 2, pp. 224–234, 2011.

[191] M. Mann, 'Innovations: Functional and quantitative proteomics using SILAC', *Nat. Rev. Mol. Cell Biol.*, vol. 7, no. 12, pp. 952–958, Dec. 2006.

[192] A. I. Nesvizhskii, 'Computational and informatics strategies for identification of specific protein interaction partners in affinity purification mass spectrometry experiments', *PROTEOMICS*, vol. 12, no. 10, pp. 1639–1655, May 2012.

[193] S.-E. Ong and M. Mann, 'Mass spectrometry–based proteomics turns quantitative', *Nat. Chem. Biol.*, vol. 1, no. 5, pp. 252–262, Oct. 2005.

[194] J. S. Andersen and M. Mann, 'Organellar proteomics: turning inventories into insights', *EMBO Rep.*, vol. 7, no. 9, pp. 874–879, Sep. 2006.

[195] D. K. Han, J. Eng, H. Zhou, and R. Aebersold, 'Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry', *Nat. Biotechnol.*, vol. 19, no. 10, pp. 946–951, Oct. 2001.

[196] X. Chen, S. Wei, Y. Ji, X. Guo, and F. Yang, 'Quantitative proteomics using SILAC: Principles, applications, and developments', *PROTEOMICS*, vol. 15, no. 18, pp. 3175–3192, 2015.

[197] M. D. Filiou, D. Martins-de-Souza, P. C. Guest, S. Bahn, and C. W. Turck, 'To label or not to label: Applications of quantitative proteomics in neuroscience research', *PROTEOMICS*, vol. 12, no. 4–5, pp. 736–747, 2012.

[198] Z. Li, R. M. Adams, K. Chourey, G. B. Hurst, R. L. Hettich, and C. Pan, 'Systematic Comparison of Label-Free, Metabolic Labeling, and Isobaric Chemical Labeling for Quantitative Proteomics on LTQ Orbitrap Velos', *J. Proteome Res.*, vol. 11, no. 3, pp. 1582–1590, Mar. 2012.

[199] Y. Shiio and R. Aebersold, 'Quantitative proteome analysis using isotope-coded affinity tags and mass spectrometry', *Nat. Protoc.*, vol. 1, no. 1, p. 139, Jun. 2006.

[200] A. Otto, D. Becher, and F. Schmidt, 'Quantitative proteomics in the field of microbiology', *PROTEOMICS*, vol. 14, no. 4–5, pp. 547–565, 2014.

[201] S.-E. Ong *et al.*, 'Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics', *Mol. Cell. Proteomics*, vol. 1, no. 5, pp. 376–386, May 2002.

[202] T. Geiger, J. Cox, P. Ostasiewicz, J. R. Wisniewski, and M. Mann, 'Super-SILAC mix for quantitative proteomics of human tumor tissue', *Nat. Methods*, vol. 7, no. 5, p. 383, 2010.

[203] A. Shenoy and T. Geiger, 'Super-SILAC: current trends and future perspectives', *Expert Rev. Proteomics*, vol. 12, no. 1, pp. 13–19, 2015.

[204] S.-E. Ong, 'The expanding field of SILAC', *Anal. Bioanal. Chem.*, vol. 404, no. 4, pp. 967–976, 2012.
[205] P. A. Everley, J. Krijgsveld, B. R. Zetter, and S. P. Gygi, 'Quantitative cancer proteomics: stable isotope labeling with amino acids in cell culture (SILAC) as a tool for prostate cancer research', *Mol. Cell. Proteomics*, vol. 3, no. 7, pp. 729–735, 2004.
[206] R. Barderas *et al.*, 'In-depth Characterization of the Secretome of Colorectal Cancer Metastatic Cells Identifies Key Proteins in Cell Adhesion, Migration, and Invasion', *Mol. Cell. Proteomics*, vol. 12, no. 6, pp. 1602–1620, Jun. 2013.
[207] J. P. Lim, S. Nair, S. Shyamasundar, J. Gunaratne, and B. H. Bay, 'Abstract 1061: Advanced quantitative mass-spectrometry-based SILAC proteome profiling of Y-box binding protein-1 (YB-1) overexpressing breast cancer cells unravels proteins involved in metastasis', *Cancer Res.*, vol. 78, no. 13 Supplement, pp. 1061–1061, Jul. 2018.
[208] G. R. Williams, J. R. Bethard, M. N. Berkaw, A. K. Nagel, L. M. Luttrell, and L. E. Ball, 'Exploring G protein-coupled receptor signaling networks using SILAC-based phosphoproteomics', *Methods*, vol. 92, pp. 36–50, 2016.
[209] T. S. Batth, M. Papetti, A. Pfeiffer, M. A. X. Tollenaere, C. Francavilla, and J. V. Olsen, 'Large-Scale Phosphoproteomics Reveals Shp-2 Phosphatase-Dependent Regulators of Pdgf Receptor Signaling', *Cell Rep.*, vol. 22, no. 10, pp. 2784–2796, Mar. 2018.
[210] A. Carpy *et al.*, 'Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC)-Based Quantitative Proteomics and Phosphoproteomics in Fission Yeast', *Cold Spring Harb. Protoc.*, vol. 2017, no. 6, p. pdb.prot091686, Jan. 2017.
[211] A. Cuomo, M. Soldi, and T. Bonaldi, 'SILAC-Based Quantitative Strategies for Accurate Histone Posttranslational Modification Profiling Across Multiple Biological Samples', in *Histones: Methods and Protocols*, B. Guillemette and L. R. Gaudreau, Eds. New York, NY: Springer New York, 2017, pp. 97–119.
[212] R. Noberini and T. Bonaldi, 'Chapter Sixteen - A Super-SILAC Strategy for the Accurate and Multiplexed Profiling of Histone Posttranslational Modifications', in *Methods in Enzymology*, vol. 586, A. K. Shukla, Ed. Academic Press, 2017, pp. 311–332.
[213] T. Zhang *et al.*, 'ING5 differentially regulates protein lysine acetylation and promotes p300 autoacetylation', *Oncotarget*, vol. 9, no. 2, pp. 1617–1629, Oct. 2017.
[214] B. D. Bryson and F. M. White, 'Quantitative Profiling of Lysine Acetylation Reveals Dynamic Crosstalk between Receptor Tyrosine Kinases and Lysine Acetylation', *PLOS ONE*, vol. 10, no. 5, p. e0126242, May 2015.
[215] M. Hilger and M. Mann, 'Triple SILAC to Determine Stimulus Specific Interactions in the Wnt Pathway', *J. Proteome Res.*, vol. 11, no. 2, pp. 982–994, Feb. 2012.
[216] M. Selbach and M. Mann, 'Protein interaction screening by quantitative immunoprecipitation combined with knockdown (QUICK)', *Nat. Methods*, vol. 3, no. 12, pp. 981–983, Dec. 2006.
[217] F. E. Paul, F. Hosp, and M. Selbach, 'Analyzing protein–protein interactions by quantitative mass spectrometry', *Methods*, vol. 54, no. 4, pp. 387–395, Aug. 2011.
[218] G. Mittler, F. Butter, and M. Mann, 'A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements', *Genome Res.*, vol. 19, no. 2, pp. 284–293, Jan. 2009.
[219] T. Viturawong, F. Meissner, F. Butter, and M. Mann, 'A DNA-Centric Protein Interaction Map of Ultraconserved Elements Reveals Contribution of Transcription Factor Binding Hubs to Conservation', *Cell Rep.*, vol. 5, no. 2, pp. 531–545, Oct. 2013.
[220] S. P. Gygi, B. Rist, S. A. Gerber, F. Turecek, M. H. Gelb, and R. Aebersold, 'Quantitative analysis of complex protein mixtures using isotope-coded affinity tags', *Nat. Biotechnol.*, vol. 17, no. 10, pp. 994–999, Oct. 1999.
[221] S. P. Gygi, B. Rist, T. J. Griffin, J. Eng, and R. Aebersold, 'Proteome Analysis of Low-Abundance Proteins Using Multidimensional Chromatography and Isotope-Coded Affinity Tags', *J. Proteome Res.*, vol. 1, no. 1, pp. 47–54, Feb. 2002.

[222] M. B. Goshe and R. D. Smith, 'Stable isotope-coded proteomic mass spectrometry', *Curr. Opin. Biotechnol.*, vol. 14, no. 1, pp. 101–109, Feb. 2003.
[223] J. C. Y. Chan, L. Zhou, and E. C. Y. Chan, 'The Isotope-Coded Affinity Tag Method for Quantitative Protein Profile Comparison and Relative Quantitation of Cysteine Redox Modifications', *Curr. Protoc. Protein Sci.*, vol. 82, no. 1, pp. 23.2.1-23.2.19, 2015.
[224] J. Yang, K. S. Carroll, and D. C. Liebler, 'The Expanding Landscape of the Thiol Redox Proteome', *Mol. Cell. Proteomics*, vol. 15, no. 1, pp. 1–11, Jan. 2016.
[225] J. D. Matthews *et al.*, 'Proteomic analysis of microbial induced redox-dependent intestinal signaling', *Redox Biol.*, vol. 20, pp. 526–532, Jan. 2019.
[226] J. Matthews, A. Reedy, T. Darby, R. Jones, and A. Neish, 'P-170 Proteomic Analysis of Redox-Dependent Intestinal Protein Thiol Modification by Isotope Coded Affinity-Tagged Labeling', *Inflamm. Bowel Dis.*, vol. 22, Mar. 2016.
[227] K. Xie, C. Bunse, K. Marcus, and L. I. Leichert, 'Quantifying changes in the bacterial thiol redox proteome during host-pathogen interaction', *Redox Biol.*, vol. 21, p. 101087, Feb. 2019.
[228] P. L. Ross *et al.*, 'Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents', *Mol. Cell. Proteomics*, vol. 3, no. 12, pp. 1154–1169, Dec. 2004.
[229] S. Wiese, K. A. Reidegeld, H. E. Meyer, and B. Warscheid, 'Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research', *PROTEOMICS*, vol. 7, no. 3, pp. 340–350, 2007.
[230] K. Aggarwal, L. H. Choe, and K. H. Lee, 'Shotgun proteomics using the iTRAQ isobaric tags', *Brief. Funct. Genomics*, vol. 5, no. 2, pp. 112–120, Jun. 2006.
[231] R. Moulder, S. D. Bhosale, D. R. Goodlett, and R. Lahesmaa, 'Analysis of the plasma proteome using iTRAQ and TMT-based Isobaric labeling', *Mass Spectrom. Rev.*, vol. 37, no. 5, pp. 583–606, 2018.
[232] L. DeSouza *et al.*, 'Search for Cancer Markers from Endometrial Tissues Using Differentially Labeled Tags iTRAQ and cICAT with Multidimensional Liquid Chromatography and Tandem Mass Spectrometry', *J. Proteome Res.*, vol. 4, no. 2, pp. 377–386, Apr. 2005.
[233] N. A. Karp, W. Huber, P. G. Sadowski, P. D. Charles, S. V. Hester, and K. S. Lilley, 'Addressing Accuracy and Precision Issues in iTRAQ Quantitation', *Mol. Cell. Proteomics*, vol. 9, no. 9, pp. 1885–1897, Sep. 2010.
[234] M. Bantscheff, M. Boesche, D. Eberhard, T. Matthieson, G. Sweetman, and B. Kuster, 'Robust and Sensitive iTRAQ Quantification on an LTQ Orbitrap Mass Spectrometer', *Mol. Cell. Proteomics*, vol. 7, no. 9, pp. 1702–1713, Sep. 2008.
[235] F. Delolme *et al.*, 'Proteolytic control of TGF-β co-receptor activity by BMP-1/tolloid-like proteases revealed by quantitative iTRAQ proteomics', *Cell. Mol. Life Sci.*, vol. 72, no. 5, pp. 1009–1027, Mar. 2015.
[236] P. P. Handakumbura, K. K. Hixson, S. O. Purvine, C. Jansson, and L. Paša-Tolić, 'Plant iTRAQ-Based Proteomics', *Curr. Protoc. Plant Biol.*, vol. 2, no. 2, pp. 158–172, 2017.
[237] J.-L. Hsu, S.-Y. Huang, N.-H. Chow, and S.-H. Chen, 'Stable-Isotope Dimethyl Labeling for Quantitative Proteomics', *Anal. Chem.*, vol. 75, no. 24, pp. 6843–6852, Dec. 2003.
[238] P. J. Boersema, R. Raijmakers, S. Lemeer, S. Mohammed, and A. J. R. Heck, 'Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics', *Nat. Protoc.*, vol. 4, no. 4, pp. 484–494, Apr. 2009.
[239] D. Kovanich, S. Cappadona, R. Raijmakers, S. Mohammed, A. Scholten, and A. J. R. Heck, 'Applications of stable isotope dimethyl labeling in quantitative proteomics', *Anal. Bioanal. Chem.*, vol. 404, no. 4, pp. 991–1009, Sep. 2012.
[240] A. F. M. Altelaar *et al.*, 'Benchmarking stable isotope labeling based quantitative proteomics', *J. Proteomics*, vol. 88, pp. 14–26, Aug. 2013.
[241] H.-T. Lau, H. W. Suh, M. Golkowski, and S.-E. Ong, 'Comparing SILAC- and Stable Isotope Dimethyl-Labeling Approaches for Quantitative Proteomics', *J. Proteome Res.*, vol. 13, no. 9, pp. 4164–4174, Sep. 2014.

[242] J.-L. Hsu, S.-Y. Huang, J.-T. Shiea, W.-Y. Huang, and S.-H. Chen, 'Beyond Quantitative Proteomics: Signal Enhancement of the a1 Ion as a Mass Tag for Peptide Sequencing Using Dimethyl Labeling', *J. Proteome Res.*, vol. 4, no. 1, pp. 101–108, Feb. 2005.

[243] C. Ji, N. Guo, and L. Li, 'Differential Dimethyl Labeling of N-Termini of Peptides after Guanidination for Proteome Analysis', *J. Proteome Res.*, vol. 4, no. 6, pp. 2099–2108, Dec. 2005.

[244] H. L. F. Swa, A. A. Shaik, L. H. K. Lim, and J. Gunaratne, 'Mass spectrometry based quantitative proteomics and integrative network analysis accentuates modulating roles of annexin-1 in mammary tumorigenesis', *PROTEOMICS*, vol. 15, no. 2–3, pp. 408–418, 2015.

[245] C.-J. Wu, Y.-W. Chen, J.-H. Tai, and S.-H. Chen, 'Quantitative Phosphoproteomics Studies Using Stable Isotope Dimethyl Labeling Coupled with IMAC-HILIC-nanoLC−MS/MS for Estrogen-Induced Transcriptional Regulation', *J. Proteome Res.*, vol. 10, no. 3, pp. 1088–1097, Mar. 2011.

[246] P. J. Boersema *et al.*, 'In-depth Qualitative and Quantitative Profiling of Tyrosine Phosphorylation Using a Combination of Phosphopeptide Immunoaffinity Purification and Stable Isotope Dimethyl Labeling', *Mol. Cell. Proteomics*, vol. 9, no. 1, pp. 84–99, Jan. 2010.

[247] S. Sadhukhan *et al.*, 'Metabolomics-assisted proteomics identifies succinylation and SIRT5 as important regulators of cardiac function', *Proc. Natl. Acad. Sci.*, vol. 113, no. 16, pp. 4320–4325, Apr. 2016.

[248] N. Khidekel *et al.*, 'Probing the dynamics of *O*-GlcNAc glycosylation in the brain using quantitative proteomics', *Nat. Chem. Biol.*, vol. 3, no. 6, pp. 339–348, Jun. 2007.

[249] Z. Sun *et al.*, 'Capture and Dimethyl Labeling of Glycopeptides on Hydrazide Beads for Quantitative Glycoproteomics Analysis', *Anal. Chem.*, vol. 84, no. 20, pp. 8452–8456, Oct. 2012.

[250] M. M. Makowski *et al.*, 'An interaction proteomics survey of transcription factor binding at recurrent TERT promoter mutations', *PROTEOMICS*, vol. 16, no. 3, pp. 417–426, 2016.

[251] K. A. Neilson *et al.*, 'Less label, more free: Approaches in label-free quantitative mass spectrometry', *PROTEOMICS*, vol. 11, no. 4, pp. 535–553, 2011.

[252] B. Zybailov, A. L. Mosley, M. E. Sardiu, M. K. Coleman, L. Florens, and M. P. Washburn, 'Statistical Analysis of Membrane Proteome Expression Changes in Saccharomyces cerevisiae', *J. Proteome Res.*, vol. 5, no. 9, pp. 2339–2347, Sep. 2006.

[253] N. M. Griffin *et al.*, 'Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis', *Nat. Biotechnol.*, vol. 28, no. 1, pp. 83–89, Jan. 2010.

[254] W. M. Old *et al.*, 'Comparison of Label-free Methods for Quantifying Human Proteins by Shotgun Proteomics', *Mol. Cell. Proteomics*, vol. 4, no. 10, pp. 1487–1502, Oct. 2005.

[255] J. A. Bubis, L. I. Levitsky, M. V. Ivanov, I. A. Tarasova, and M. V. Gorshkov, 'Comparative evaluation of label-free quantification methods for shotgun proteomics', *Rapid Commun. Mass Spectrom.*, vol. 31, no. 7, pp. 606–612, 2017.

[256] J. Cox, M. Y. Hein, C. A. Luber, I. Paron, N. Nagaraj, and M. Mann, 'Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ', *Mol. Cell. Proteomics*, vol. 13, no. 9, pp. 2513–2526, Sep. 2014.

[257] D. A. Megger *et al.*, 'Comparison of label-free and label-based strategies for proteome analysis of hepatoma cell lines', *Biochim. Biophys. Acta BBA - Proteins Proteomics*, vol. 1844, no. 5, pp. 967–976, May 2014.

[258] L. J. E. Goeminne, A. Argentini, L. Martens, and L. Clement, 'Summarization vs Peptide-Based Models in Label-Free Quantitative Proteomics: Performance, Pitfalls, and Data Analysis Guidelines', *J. Proteome Res.*, vol. 14, no. 6, pp. 2457–2465, Jun. 2015.

[259] L. N. Mueller *et al.*, 'SuperHirn – a novel tool for high resolution LC-MS-based peptide/protein profiling', *PROTEOMICS*, vol. 7, no. 19, pp. 3470–3480, 2007.

[260] J. Cox and M. Mann, 'MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification', *Nat. Biotechnol.*, vol. 26, no. 12, pp. 1367–1372, Dec. 2008.

[261] J. Merl, M. Ueffing, S. M. Hauck, and C. von Toerne, 'Direct comparison of MS-based label-free and SILAC quantitative proteome profiling strategies in primary retinal Müller cells', *PROTEOMICS*, vol. 12, no. 12, pp. 1902–1911, 2012.

[262] V. J. Patel *et al.*, 'A Comparison of Labeling and Label-Free Mass Spectrometry-Based Proteomics Approaches', *J. Proteome Res.*, vol. 8, no. 7, pp. 3752–3759, Jul. 2009.
[263] V. Brun, C. Masselon, J. Garin, and A. Dupuis, 'Isotope dilution strategies for absolute quantitative proteomics', *J. Proteomics*, vol. 72, no. 5, pp. 740–749, Jul. 2009.
[264] O. Stemmann, H. Zou, S. A. Gerber, S. P. Gygi, and M. W. Kirschner, 'Dual Inhibition of Sister Chromatid Separation at Metaphase', *Cell*, vol. 107, no. 6, pp. 715–726, Dec. 2001.
[265] R. J. Beynon, M. K. Doherty, J. M. Pratt, and S. J. Gaskell, 'Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides', *Nat. Methods*, vol. 2, no. 8, pp. 587–589, Aug. 2005.
[266] V. Brun *et al.*, 'Isotope-labeled Protein Standards: Toward Absolute Quantitative Proteomics', *Mol. Cell. Proteomics*, vol. 6, no. 12, pp. 2139–2149, Dec. 2007.
[267] O. Heudi *et al.*, 'Towards Absolute Quantification of Therapeutic Monoclonal Antibody in Serum by LC−MS/MS Using Isotope-Labeled Antibody Standard and Protein Cleavage Isotope Dilution Mass Spectrometry', *Anal. Chem.*, vol. 80, no. 11, pp. 4200–4207, Jun. 2008.
[268] J. Rappsilber, U. Ryder, A. I. Lamond, and M. Mann, 'Large-Scale Proteomic Analysis of the Human Spliceosome', *Genome Res.*, vol. 12, no. 8, pp. 1231–1245, Jan. 2002.
[269] Y. Ishihama *et al.*, 'Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein', *Mol. Cell. Proteomics*, vol. 4, no. 9, pp. 1265–1272, Sep. 2005.
[270] B. Schwanhäusser *et al.*, 'Global quantification of mammalian gene expression control', *Nature*, vol. 473, no. 7347, pp. 337–342, May 2011.
[271] A. H. Smits, P. W. T. C. Jansen, I. Poser, A. A. Hyman, and M. Vermeulen, 'Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics', *Nucleic Acids Res.*, vol. 41, no. 1, pp. e28–e28, Jan. 2013.
[272] L. Arike, K. Valgepea, L. Peil, R. Nahku, K. Adamberg, and R. Vilu, 'Comparison and applications of label-free absolute proteome quantification methods on Escherichia coli', *J. Proteomics*, vol. 75, no. 17, pp. 5437–5448, Sep. 2012.
[273] M. Karas, D. Bachmann, and F. Hillenkamp, 'Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules', *Anal. Chem.*, vol. 57, no. 14, pp. 2935–2939, 1985.
[274] M. Karas, D. Bachmann, U. Bahr, and F. Hillenkamp, 'Matrix-assisted ultraviolet laser desorption of non-volatile compounds', *Int. J. Mass Spectrom. Ion Process.*, vol. 78, pp. 53–68, Sep. 1987.
[275] F. Hillenkamp, M. Karas, R. C. Beavis, and B. T. Chait, 'Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers', *Anal. Chem.*, vol. 63, no. 24, pp. 1193A–1203A, 1991.
[276] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse, 'Electrospray ionization for mass spectrometry of large biomolecules', *Science*, vol. 246, no. 4926, pp. 64–71, Oct. 1989.
[277] M. Karas, M. Glückmann, and J. Schäfer, 'Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors', *J. Mass Spectrom.*, vol. 35, no. 1, pp. 1–12, 2000.
[278] N. Medina, T. Huth-Fehre, A. Westman, and B. U. R. Sundqvist, 'Matrix-assisted laser desorption: Dependence of the threshold fluence on analyte concentration', *Org. Mass Spectrom.*, vol. 29, no. 4, pp. 207–209, 1994.
[279] M. Karas, U. Bahr, and U. Gießmann, 'Matrix-assisted laser desorption ionization mass spectrometry', *Mass Spectrom. Rev.*, vol. 10, no. 5, pp. 335–357, 1991.
[280] Taylor Geoffrey Ingram, 'Disintegration of water drops in an electric field', *Proc. R. Soc. Lond. Ser. Math. Phys. Sci.*, vol. 280, no. 1382, pp. 383–397, Jul. 1964.
[281] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse, 'Electrospray ionization–principles and practice', *Mass Spectrom. Rev.*, vol. 9, no. 1, pp. 37–70, 1990.
[282] S. Nguyen and J. B. Fenn, 'Gas-phase ions of solute species from charged droplets of solutions', *Proc. Natl. Acad. Sci.*, vol. 104, no. 4, pp. 1111–1117, Jan. 2007.

[283] J. V. Iribarne and B. A. Thomson, 'On the evaporation of small ions from charged droplets', *J. Chem. Phys.*, vol. 64, no. 6, pp. 2287–2294, Mar. 1976.
[284] R. D. Smith, J. A. Loo, R. R. O. Loo, M. Busman, and H. R. Udseth, 'Principles and practice of electrospray ionization—mass spectrometry for large polypeptides and proteins', *Mass Spectrom. Rev.*, vol. 10, no. 5, pp. 359–452, Sep. 1991.
[285] R. D. Smith, J. A. Loo, C. G. Edmonds, C. J. Barinaga, and H. R. Udseth, 'New developments in biochemical mass spectrometry: electrospray ionization', *Anal. Chem.*, vol. 62, no. 9, pp. 882–899, May 1990.
[286] J. A. Loo, H. R. Udseth, and R. D. Smith, 'Peptide and protein analysis by electrospray ionization-mass spectrometry and capillary electrophoresis-mass spectrometry', *Anal. Biochem.*, vol. 179, no. 2, pp. 404–412, Jun. 1989.
[287] B. A. Mamyrin, 'Time-of-flight mass spectrometry (concepts, achievements, and prospects)', *Int. J. Mass Spectrom.*, vol. 206, no. 3, pp. 251–266, Mar. 2001.
[288] W. C. Wiley and I. H. McLaren, 'Time-of-Flight Mass Spectrometer with Improved Resolution', *Rev. Sci. Instrum.*, vol. 26, no. 12, pp. 1150–1157, Dec. 1955.
[289] B. A. Mamyrin, V. I. Karataev, D. V. Shmikk, and V. A. Zagulin, 'The mass-reflectron. A new nonmagnetic time-of-flight high resolution mass-spectrometer', *Zhurnal Ehksperimentalnoj Teor. Fiz.*, vol. 64, no. 1, pp. 82–89, 1973.
[290] M. L. Vestal, P. Juhasz, and S. A. Martin, 'Delayed extraction matrix-assisted laser desorption time-of-flight mass spectrometry', *Rapid Commun. Mass Spectrom.*, vol. 9, no. 11, pp. 1044–1050, 1995.
[291] R. M. Whittal and Liang. Li, 'High-Resolution Matrix-Assisted Laser Desorption/Ionization in a Linear Time-of-Flight Mass Spectrometer', *Anal. Chem.*, vol. 67, no. 13, pp. 1950–1954, Jul. 1995.
[292] T. Bergmann, T. P. Martin, and H. Schaber, 'High-resolution time-of-flight mass spectrometer', *Rev. Sci. Instrum.*, vol. 60, no. 4, pp. 792–793, Apr. 1989.
[293] M. Karas, U. Bahr, A. Ingendoh, and F. Hillenkamp, 'Laser Desorption/Ionization Mass Spectrometry of Proteins of Mass 100 000 to 250 000 Dalton', *Angew. Chem. Int. Ed. Engl.*, vol. 28, no. 6, pp. 760–761, 1989.
[294] P. H. Dawson, 'Quadrupole mass spectrometer', US4189640A, 19-Feb-1980.
[295] P. H. Dawson, 'Quadrupole mass analyzers: Performance, design and some recent applications', *Mass Spectrom. Rev.*, vol. 5, no. 1, pp. 1–37, 1986.
[296] R. A. Yost and C. G. Enke, 'Triple quadrupole mass spectrometry for direct mixture analysis and structure elucidation', *Anal. Chem.*, vol. 51, no. 12, pp. 1251–1264, Oct. 1979.
[297] F. W. McLafferty, 'Tandem mass spectrometry', *Science*, vol. 214, no. 4518, pp. 280–287, Oct. 1981.
[298] W. M. Brubaker, 'Auxiliary electrodes for quadrupole mass filters', US3129327A, 14-Apr-1964.
[299] Thermo Fisher Scientific, 'Q Exactive™ Hybrid Quadrupole-Orbitrap™ Mass Spectrometer'.
[300] K. H. Kingdon, 'A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures', *Phys. Rev.*, vol. 21, no. 4, pp. 408–418, Apr. 1923.
[301] A. Makarov, 'Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis', *Anal. Chem.*, vol. 72, no. 6, pp. 1156–1162, Mar. 2000.
[302] M. Scigelova and A. Makarov, 'Orbitrap Mass Analyzer – Overview and Applications in Proteomics', *PROTEOMICS*, vol. 6, no. S2, pp. 16–21, 2006.
[303] Q. Hu, R. J. Noll, H. Li, A. Makarov, M. Hardman, and R. G. Cooks, 'The Orbitrap: a new mass spectrometer', *J. Mass Spectrom.*, vol. 40, no. 4, pp. 430–443, 2005.
[304] A. Makarov *et al.*, 'Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer', *Anal. Chem.*, vol. 78, no. 7, pp. 2113–2120, Apr. 2006.
[305] A. Michalski *et al.*, 'Mass Spectrometry-based Proteomics Using Q Exactive, a High-performance Benchtop Quadrupole Orbitrap Mass Spectrometer', *Mol. Cell. Proteomics*, vol. 10, no. 9, p. M111.011015, Sep. 2011.

[306] R. A. Zubarev and A. Makarov, 'Orbitrap Mass Spectrometry', *Anal. Chem.*, vol. 85, no. 11, pp. 5288–5296, Jun. 2013.
[307] S. L. Kloet *et al.*, 'Towards elucidating the stability, dynamics and architecture of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics', *FEBS J.*, vol. 282, no. 9, pp. 1774–1785, 2015.
[308] E. Y. N. Lam, D. Beraldi, D. Tannahill, and S. Balasubramanian, 'G-quadruplex structures are stable and detectable in human genomic DNA', *Nat. Commun.*, vol. 4, p. 1796, Apr. 2013.
[309] T. K. Wöbke, 'Regulation of CD69 expression in human monocytic cells', Dissertation, Goethe-University, Frankfurt, 2013.
[310] J. Fettel, 'Gene regulation of 5-lipoxygenase', Master Thesis, Goethe-University, Frankfurt, 2013.
[311] J. D. Dignam, R. M. Lebovitz, and R. G. Roeder, 'Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei', *Nucleic Acids Res.*, vol. 11, no. 5, pp. 1475–1489, Mar. 1983.
[312] C. Englert, 'WT1—more than a transcription factor?', *Trends Biochem. Sci.*, vol. 23, no. 10, pp. 389–393, Oct. 1998.
[313] L. Yang, Y. Han, F. Saurez Saiz, and M. D. Minden, 'A tumor suppressor and oncogene: the WT1 story', *Leukemia*, vol. 21, no. 5, pp. 868–876, May 2007.
[314] G. Candiano *et al.*, 'Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis', *Electrophoresis*, vol. 25, no. 9, pp. 1327–1333, May 2004.
[315] I. Miller, J. Crawford, and E. Gianazza, 'Protein stains for proteomic applications: which, when, why?', *Proteomics*, vol. 6, no. 20, pp. 5385–5408, Oct. 2006.
[316] V. Brázda, J. Červeň, M. Bartas, N. Mikysková, J. Coufal, and P. Pečinka, 'The Amino Acid Composition of Quadruplex Binding Proteins Reveals a Shared Motif and Predicts New Potential Quadruplex Interactors', *Molecules*, vol. 23, no. 9, p. 2341, Sep. 2018.
[317] N. A. Karp and K. S. Lilley, 'Design and Analysis Issues in Quantitative Proteomics Studies', *PROTEOMICS*, vol. 7, no. S1, pp. 42–50, 2007.
[318] W. Urfer, M. Grzegorczyk, and K. Jung, 'Statistics for Proteomics: A Review of Tools for Analyzing Experimental Data', *PROTEOMICS*, vol. 6, no. S2, pp. 48–55, 2006.
[319] S. J. Collins, 'The HL-60 promyelocytic leukemia cell line: proliferation, differentiation, and cellular oncogene expression', *Blood*, vol. 70, no. 5, pp. 1233–1244, Nov. 1987.
[320] K. Karimi, T. R. Gemmill, and M. R. Lennartz, 'Protein kinase C and a calcium-independent phospholipase are required for IgG-mediated phagocytosis by Mono-Mac-6 cells', *J. Leukoc. Biol.*, vol. 65, no. 6, pp. 854–862, Jun. 1999.
[321] J. Kaczynski *et al.*, 'The Sp1-like protein BTEB3 inhibits transcription via the basic transcription element box by interacting with mSin3A and HDAC-1 co-repressors and competing with Sp1', *J. Biol. Chem.*, vol. 276, no. 39, pp. 36749–36756, Sep. 2001.
[322] D. P. Ramji and P. Foka, 'CCAAT/enhancer-binding proteins: structure, function and regulation', *Biochem. J.*, vol. 365, no. 3, pp. 561–575, Aug. 2002.
[323] A. K. Byrd *et al.*, 'Evidence That G-quadruplex DNA Accumulates in the Cytoplasm and Participates in Stress Granule Assembly in Response to Oxidative Stress', *J. Biol. Chem.*, vol. 291, no. 34, pp. 18041–18057, Aug. 2016.
[324] L.-X. Wang *et al.*, 'Pyridostatins selectively recognize two different forms of the human telomeric G-quadruplex structures and their anti-tumor activities in vitro', *Tetrahedron*, vol. 71, no. 30, pp. 4982–4986, Jul. 2015.
[325] A. M. Deaton and A. Bird, 'CpG islands and the regulation of transcription', *Genes Dev.*, vol. 25, no. 10, pp. 1010–1022, May 2011.
[326] Y. Wang, J. Zhu, and H. F. DeLuca, 'Where is the vitamin D receptor?', *Arch. Biochem. Biophys.*, vol. 523, no. 1, pp. 123–133, Jul. 2012.
[327] J. W. Morgan, D. M. Morgan, S. R. Lasky, D. Ford, N. Kouttab, and A. L. Maizel, 'Requirements for induction of vitamin D-mediated gene regulation in normal human B lymphocytes.', *J. Immunol.*, vol. 157, no. 7, pp. 2900–2908, Oct. 1996.

[328] C. M. Veldman, M. T. Cantorna, and H. F. DeLuca, 'Expression of 1,25-Dihydroxyvitamin D3 Receptor in the Immune System', *Arch. Biochem. Biophys.*, vol. 374, no. 2, pp. 334–338, Feb. 2000.
[329] H. Defacque *et al.*, 'Expression of Retinoid X Receptor α Is Increased upon Monocytic Cell Differentiation', *Biochem. Biophys. Res. Commun.*, vol. 220, no. 2, pp. 315–322, Mar. 1996.
[330] J. Fritsche, T. J. Stonehouse, D. R. Katz, R. Andreesen, and M. Kreutz, 'Expression of Retinoid Receptors during Human Monocyte Differentiation in Vitro', *Biochem. Biophys. Res. Commun.*, vol. 270, no. 1, pp. 17–22, Apr. 2000.
[331] J. Lømo *et al.*, 'RAR-, not RXR, ligands inhibit cell activation and prevent apoptosis in B-lymphocytes', *J. Cell. Physiol.*, vol. 175, no. 1, pp. 68–77, 1998.
[332] D. A. Haber, R. L. Sohn, A. J. Buckler, J. Pelletier, K. M. Call, and D. E. Housman, 'Alternative splicing and genomic structure of the Wilms tumor gene WT1.', *Proc. Natl. Acad. Sci.*, vol. 88, no. 21, pp. 9618–9622, Nov. 1991.
[333] T. Rabilloud, 'Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains', *PROTEOMICS Int. Ed.*, vol. 2, no. 1, pp. 3–10, 2002.
[334] S. P. Gygi, G. L. Corthals, Y. Zhang, Y. Rochon, and R. Aebersold, 'Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology', *Proc. Natl. Acad. Sci.*, vol. 97, no. 17, pp. 9390–9395, 2000.
[335] H. J. Issaq, 'The role of separation science in proteomics research', *Electrophoresis*, vol. 22, no. 17, pp. 3629–3638, 2001.
[336] M. Gstaiger and R. Aebersold, 'Applying mass spectrometry-based proteomics to genetics, genomics and network biology', *Nat. Rev. Genet.*, vol. 10, no. 9, p. 617, 2009.
[337] W. M. Lauber, J. A. Carroll, D. R. Dufield, J. R. Kiesel, M. R. Radabaugh, and J. P. Malone, 'Mass spectrometry compatibility of two-dimensional gel protein stains', *ELECTROPHORESIS*, vol. 22, no. 5, pp. 906–918, 2001.
[338] K. Speicher, O. Kolbas, S. Harper, and D. Speicher, 'Systematic analysis of peptide recoveries from in-gel digestions for protein identifications in proteome studies', *J. Biomol. Tech. JBT*, vol. 11, no. 2, pp. 74–86, Jun. 2000.
[339] A. Pandey and M. Mann, 'Proteomics to study genes and genomes', *Nature*, vol. 405, no. 6788, p. 837, 2000.
[340] K. Meyer and M. Selbach, 'Quantitative affinity purification mass spectrometry: a versatile technology to study protein–protein interactions', *Front. Genet.*, vol. 6, 2015.
[341] P. Navarro *et al.*, 'A multicenter study benchmarks software tools for label-free proteome quantification', *Nat. Biotechnol.*, vol. 34, no. 11, pp. 1130–1136, Nov. 2016.
[342] P. Fanis *et al.*, 'Five Friends of Methylated Chromatin Target of Protein-Arginine-Methyltransferase[Prmt]-1 (Chtop), a Complex Linking Arginine Methylation to Desumoylation', *Mol. Cell. Proteomics*, vol. 11, no. 11, pp. 1263–1273, Nov. 2012.
[343] D. B. Bloch *et al.*, 'Sp110 Localizes to the PML-Sp100 Nuclear Body and May Function as a Nuclear Hormone Receptor Transcriptional Coactivator', *Mol. Cell. Biol.*, vol. 20, no. 16, pp. 6138–6146, Aug. 2000.
[344] B.-J. M. Webb-Robertson *et al.*, 'Review, Evaluation, and Discussion of the Challenges of Missing Value Imputation for Mass Spectrometry-Based Label-Free Global Proteomics', *J. Proteome Res.*, vol. 14, no. 5, pp. 1993–2001, May 2015.
[345] X. Zhao *et al.*, 'MicroRNA-125a contributes to elevated inflammatory chemokine RANTES levels via targeting KLF13 in systemic lupus erythematosus', *Arthritis Rheum.*, vol. 62, no. 11, pp. 3425–3435, 2010.
[346] S. Banerjee *et al.*, 'miR-125a-5p regulates differential activation of macrophages and inflammation', *J. Biol. Chem.*, vol. 288, no. 49, pp. 35428–35436, Dec. 2013.
[347] H.-M. Lee, T. S. Kim, and E.-K. Jo, 'MiR-146 and miR-125 in the regulation of innate immunity and inflammation', *BMB Rep.*, vol. 49, no. 6, pp. 311–318, Jun. 2016.
[348] S. Busch *et al.*, '5-lipoxygenase is a direct target of miR-19a-3p and miR-125b-5p', *J. Immunol. Baltim. Md 1950*, vol. 194, no. 4, pp. 1646–1653, Feb. 2015.

[349] M. M. Makowski, C. Gräwe, B. M. Foster, N. V. Nguyen, T. Bartke, and M. Vermeulen, 'Global profiling of protein–DNA and protein–nucleosome binding affinities using quantitative mass spectrometry', *Nat. Commun.*, vol. 9, no. 1, p. 1653, Apr. 2018.
[350] G. S. Daftary *et al.*, 'Detailed Structural-Functional Analysis of the Krüppel-like Factor 16 (KLF16) Transcription Factor Reveals Novel Mechanisms for Silencing Sp/KLF Sites Involved in Metabolism and Endocrinology', *J. Biol. Chem.*, vol. 287, no. 10, pp. 7010–7025, Mar. 2012.
[351] S. K. Swamynathan, 'Krüppel-like factors: Three fingers in control', *Hum. Genomics*, vol. 4, no. 4, pp. 263–270, Apr. 2010.
[352] J.-W. Xiong, A. Leahy, H.-H. Lee, and H. Stuhlmann, 'Vezf1:A Zn Finger Transcription Factor Restricted to Endothelial Cells and Their Precursors', *Dev. Biol.*, vol. 206, no. 2, pp. 123–141, Feb. 1999.
[353] H. Gowher, H. Stuhlmann, and G. Felsenfeld, 'Vezf1 regulates genomic DNA methylation through its effects on expression of DNA methyltransferase Dnmt3b', *Genes Dev.*, vol. 22, no. 15, pp. 2075–2084, Jan. 2008.
[354] H. Gowher, K. Brick, R. D. Camerini-Otero, and G. Felsenfeld, 'Vezf1 protein binding sites genome-wide are associated with pausing of elongating RNA polymerase II', *Proc. Natl. Acad. Sci.*, vol. 109, no. 7, pp. 2370–2375, Feb. 2012.
[355] M. Yonaha and N. J. Proudfoot, 'Transcriptional termination and coupled polyadenylation in vitro', *EMBO J.*, vol. 19, no. 14, pp. 3770–3777, Jul. 2000.
[356] B. K. Ray, R. Murphy, P. Ray, and A. Ray, 'SAF-2, a splice variant of SAF-1, acts as a negative regulator of transcription', *J. Biol. Chem.*, vol. 277, no. 48, pp. 46822–46830, Nov. 2002.
[357] N. Dhaouadi *et al.*, 'Computational identification of potential transcriptional regulators of TGF-ß1 in human atherosclerotic arteries', *Genomics*, vol. 103, no. 5, pp. 357–370, May 2014.
[358] M. Smits *et al.*, 'Myc-associated zinc finger protein (MAZ) is regulated by miR-125b and mediates VEGF-induced angiogenesis in glioblastoma', *FASEB J.*, Mar. 2012.
[359] J. Song *et al.*, 'Transcriptional regulation by zinc-finger proteins Sp1 and MAZ involves interactions with the same cis-elements (Review)', *Int. J. Mol. Med.*, vol. 11, no. 5, pp. 547–553, May 2003.
[360] C. L. Parks and T. Shenk, 'The Serotonin 1a Receptor Gene Contains a TATA-less Promoter that Responds to MAZ and Sp1', *J. Biol. Chem.*, vol. 271, no. 8, pp. 4417–4430, Feb. 1996.
[361] R. C. DeKelver *et al.*, 'Cooperation between RUNX1-ETO9a and Novel Transcriptional Partner KLF6 in Upregulation of Alox5 in Acute Myeloid Leukemia', *PLOS Genet.*, vol. 9, no. 10, p. e1003765, Oct. 2013.
[362] B.-N. Jeon, J.-Y. Yoo, W.-I. Choi, C.-E. Lee, H.-G. Yoon, and M.-W. Hur, 'Proto-oncogene FBI-1 (Pokemon/ZBTB7A) Represses Transcription of the Tumor Suppressor Rb Gene via Binding Competition with Sp1 and Recruitment of Co-repressors', *J. Biol. Chem.*, vol. 283, no. 48, pp. 33199–33210, Nov. 2008.
[363] N. Hojo *et al.*, 'A Zbtb7a proto-oncogene as a novel target for miR-125a', *Mol. Carcinog.*, vol. 55, no. 12, pp. 2001–2009, 2016.
[364] Y. Yang *et al.*, 'Pokemon (FBI-1) interacts with Smad4 to repress TGF-β-induced transcriptional responses', *Biochim. Biophys. Acta BBA - Gene Regul. Mech.*, vol. 1849, no. 3, pp. 270–281, Mar. 2015.
[365] J.-A. Park, T.-H. Kim, B. Lee, E. Kwon, and K.-C. Kim, 'Expression of PRDM10 in arthritic synovial derived tissues', *Genes Genomics*, vol. 35, no. 6, pp. 685–691, Dec. 2013.
[366] R. C. Borghaei and M. Chambers, 'Expression of transcription factor zinc-binding protein-89 (ZBP-89) is inhibited by inflammatory cytokines', *Pathol. Lab. Med. Int. PLMI*, vol. 1, pp. 7–12, Aug. 2009.
[367] L. Bai, C. Logsdon, and J. L. Merchant, 'Regulation of epithelial cell growth by ZBP-89', *Int. J. Gastrointest. Cancer*, vol. 31, no. 1, pp. 79–88, May 2002.
[368] S. Hahn and H. Hermeking, 'ZNF281/ZBP-99: a new player in epithelial–mesenchymal transition, stemness, and cancer', *J. Mol. Med.*, vol. 92, no. 6, pp. 571–581, Jun. 2014.

[369] M. Pierdomenico *et al.*, 'Transcription Factor ZNF281: A Novel Player in Intestinal Inflammation and Fibrosis', *Front. Immunol.*, vol. 9, 2018.
[370] H. Zhou *et al.*, 'ZNF281 enhances cardiac reprogramming by modulating cardiac and inflammatory gene expression', *Genes Dev.*, vol. 31, no. 17, pp. 1770–1783, Sep. 2017.
[371] J. Lekstrom-Himes and K. G. Xanthopoulos, 'Biological Role of the CCAAT/Enhancer-binding Protein Family of Transcription Factors', *J. Biol. Chem.*, vol. 273, no. 44, pp. 28545–28548, Oct. 1998.
[372] K. V. Reddy, K. J. Serio, C. R. Hodulik, and T. D. Bigby, '5-Lipoxygenase-activating Protein Gene Expression KEY ROLE OF CCAAT/ENHANCER-BINDING PROTEINS (C/EBP) IN CONSTITUTIVE AND TUMOR NECROSIS FACTOR (TNF) α-INDUCED EXPRESSION IN THP-1 CELLS', *J. Biol. Chem.*, vol. 278, no. 16, pp. 13810–13818, Apr. 2003.
[373] Z. Pan, C. J. Hetherington, and D.-E. Zhang, 'CCAAT/Enhancer-binding Protein Activates the CD14 Promoter and Mediates Transforming Growth Factor β Signaling in Monocyte Development', *J. Biol. Chem.*, vol. 274, no. 33, pp. 23242–23248, Aug. 1999.
[374] É. Lamarche, N. Lala-Tabbert, A. Gunanayagam, C. St-Louis, and N. Wiper-Bergeron, 'Retinoic acid promotes myogenesis in myoblasts by antagonizing transforming growth factor-beta signaling via C/EBPβ', *Skelet. Muscle*, vol. 5, no. 1, p. 8, Mar. 2015.
[375] M. Alberich-Jordà *et al.*, 'C/EBPγ deregulation results in differentiation arrest in acute myeloid leukemia', *J. Clin. Invest.*, vol. 122, no. 12, pp. 4490–4504, Dec. 2012.
[376] C. J. Huggins *et al.*, 'C/EBPγ Suppresses Senescence and Inflammatory Gene Expression by Heterodimerizing with C/EBPβ', *Mol. Cell. Biol.*, vol. 33, no. 16, pp. 3242–3258, Aug. 2013.
[377] R. Morosetti *et al.*, 'A Novel, Myeloid Transcription Factor, C/EBPε, Is Upregulated During Granulocytic, But Not Monocytic, Differentiation', *Blood*, vol. 90, no. 7, pp. 2591–2600, Oct. 1997.
[378] S. Halene *et al.*, 'C/EBPε directs granulocytic-vs-monocytic lineage determination and confers chemotactic function via Hlx', *Exp. Hematol.*, vol. 38, no. 2, pp. 90-103.e4, Feb. 2010.
[379] S. Garapaty, M. A. Mahajan, and H. H. Samuels, 'Components of the CCR4-NOT Complex Function as Nuclear Hormone Receptor Coactivators via Association with the NRC-interacting Factor NIF-1', *J. Biol. Chem.*, vol. 283, no. 11, pp. 6806–6816, Mar. 2008.
[380] Y. J. Yang *et al.*, 'Microcephaly gene links Trithorax and REST/NRSF to control neural stem cell proliferation and differentiation', *Cell*, vol. 151, no. 5, pp. 1097–1112, Nov. 2012.
[381] G. V. Rayasam *et al.*, 'NSD1 is essential for early post-implantation development and has a catalytically active SET domain', *EMBO J.*, vol. 22, no. 12, pp. 3153–3163, Jun. 2003.
[382] N. Huang *et al.*, 'Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators', *EMBO J.*, vol. 17, no. 12, pp. 3398–3412, Jun. 1998.
[383] S. K. Rhie *et al.*, 'ZFX acts as a transcriptional activator in multiple types of human tumors by binding downstream from transcription start sites at the majority of CpG island promoters', *Genome Res.*, vol. 28, no. 3, pp. 310–320, Jan. 2018.
[384] R.-F. Bao *et al.*, 'miR-101 targeting ZFX suppresses tumor proliferation and metastasis by regulating the MAPK/Erk and Smad pathways in gallbladder carcinoma', *Oncotarget*, vol. 7, no. 16, pp. 22339–22354, Mar. 2016.
[385] Q. Wu *et al.*, 'CINP is a novel cofactor of KLF5 required for its role in the promotion of cell proliferation, survival and tumor growth', *Int. J. Cancer*, vol. 144, no. 3, pp. 582–594, 2019.
[386] D. Ma *et al.*, '1, 25(OH)2D3-induced interaction of vitamin D receptor with p50 subunit of NF-κB suppresses the interaction between KLF5 and p50, contributing to inhibition of LPS-induced macrophage proliferation', *Biochem. Biophys. Res. Commun.*, vol. 482, no. 2, pp. 366–374, Jan. 2017.
[387] Z. Xie *et al.*, 'TGF-β synergizes with ML264 to block IL-1β-induced matrix degradation mediated by Krüppel-like factor 5 in the nucleus pulposus', *Biochim. Biophys. Acta BBA - Mol. Basis Dis.*, vol. 1864, no. 2, pp. 579–589, Feb. 2018.

[388] P. Guo *et al.*, 'Pro-proliferative Factor KLF5 Becomes Anti-proliferative in Epithelial Homeostasis upon Signaling-mediated Modification', *J. Biol. Chem.*, vol. 284, no. 10, pp. 6071–6078, Mar. 2009.

[389] P. Guo, K.-W. Zhao, X.-Y. Dong, X. Sun, and J.-T. Dong, 'Acetylation of KLF5 Alters the Assembly of p15 Transcription Factors in Transforming Growth Factor-β-mediated Induction in Epithelial Cells', *J. Biol. Chem.*, vol. 284, no. 27, pp. 18184–18193, Jul. 2009.

[390] P. Guo, X.-Y. Dong, K. Zhao, X. Sun, Q. Li, and J.-T. Dong, 'Opposing Effects of KLF5 on the Transcription of MYC in Epithelial Proliferation in the Context of Transforming Growth Factor β', *J. Biol. Chem.*, vol. 284, no. 41, pp. 28243–28252, Oct. 2009.

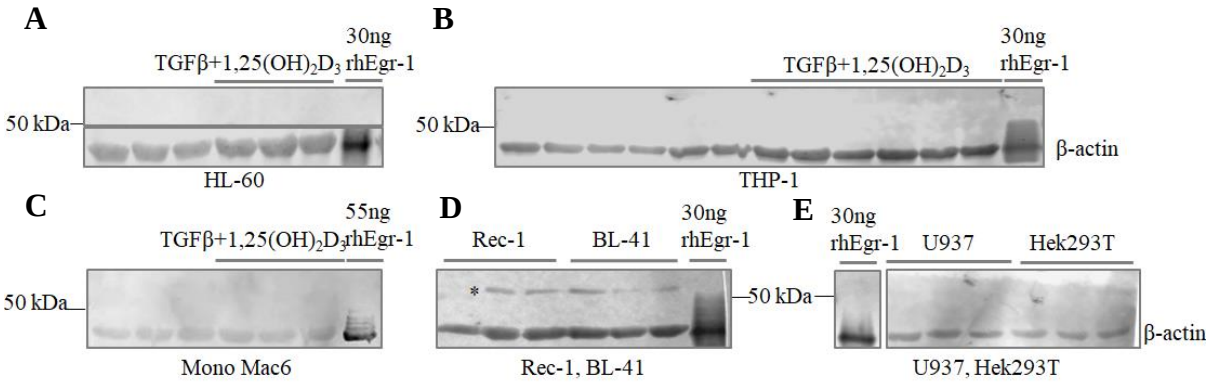
[391] A. Ponton, J.-P. Thirion, and P. Siroie, 'Repression of the 5-lipoxygenase gene by c-myb overexpression in differentiated HL-60 cells', *Prostaglandins*, vol. 53, no. 1, pp. 49–58, Jan. 1997.

[392] A. Membrino, M. Paramasivam, S. Cogoi, J. Alzeer, N. W. Luedtke, and L. E. Xodo, 'Cellular uptake and binding of guanidine -modified phthalocyanines to KRAS / HRAS G-quadruplexes', *Chem. Commun.*, vol. 46, no. 4, pp. 625–627, 2010.

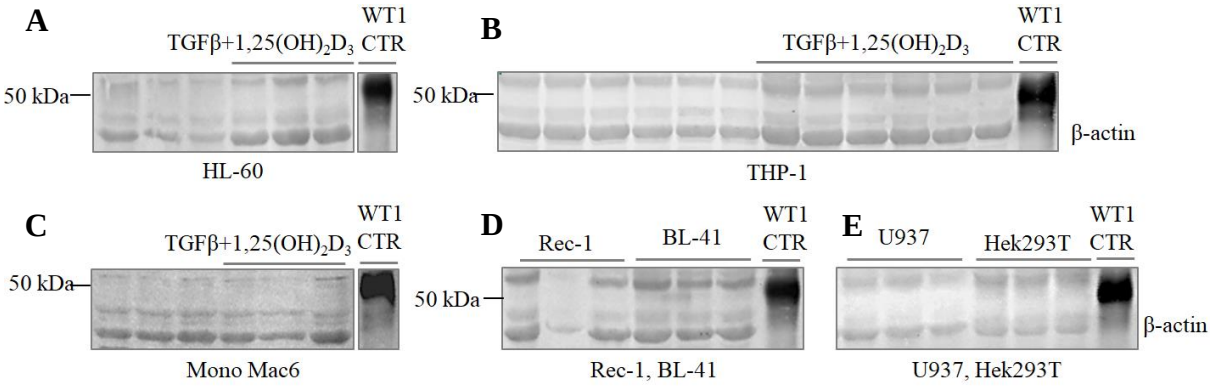
[393] S. Cogoi and L. E. Xodo, 'Enhanced G4-DNA binding of 5,10,15,20 (N-propyl-4-pyridyl) porphyrin (TPrPyP4): A comparative study with TMPyP4', *Chem. Commun.*, vol. 46, no. 39, pp. 7364–7366, 2010.

9. Appendix

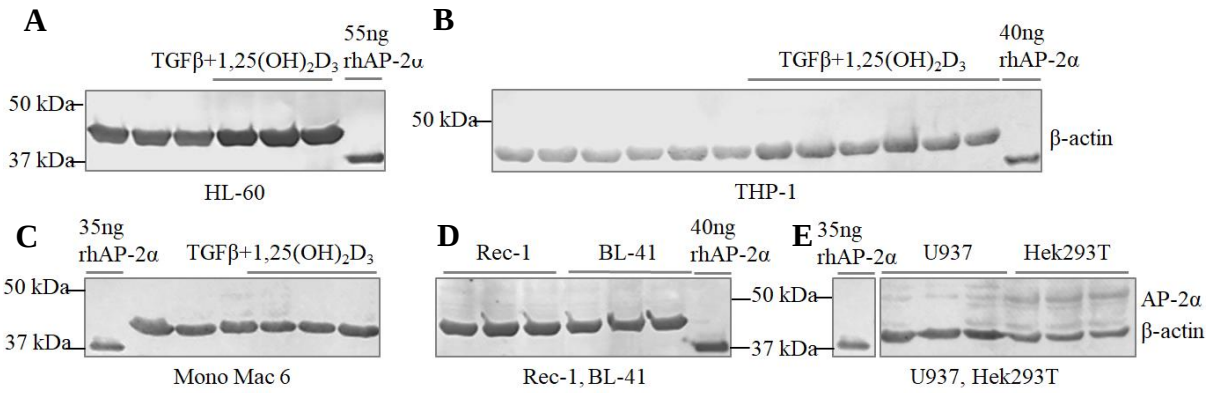
9.1 Supplementary figures



Supplementary figure 1: Immunoblotting of transcription factor Egr-1. 10-20 µl whole cell lysate containing 30 µg protein were blotted against β-actin as loading control (anti-Egr-1-ab 1:1000). Immunoblots comprised the depicted amounts of human recombinant Egr-1, serving as control. Blots are representative of three (six for THP-1) subsequent cell passages of (A) HL-60 in undifferentiated and differentiated (dHL-60), (B) THP-1 in undifferentiated and differentiated, (C) MM6 in undifferentiated and differentiated state, (D) Rec-1 and BL-41 and (E) 5-LO-negative controls U937 and Hek293T. Bands marked with an asterisk are supposedly artefacts, since the detected signals belong to the secondary antibody used.



Supplementary figure 2: Immunoblotting of transcription factor WT1. 10-20 µl whole cell lysate containing 30 µg protein were blotted against β-actin as loading control (anti-WT1-ab 1:1000). Immunoblots comprised a WT1-plasmid transfected control cell line. Blots are representative of three (six for THP-1) subsequent cell passages of (A) HL-60 in undifferentiated and differentiated (dHL-60), (B) THP-1 in undifferentiated and differentiated, (C) MM6 in undifferentiated and differentiated state, (D) Rec-1 and BL-41 and (E) 5-LO-negative controls U937 and Hek293T.



Supplementary figure 3: Immunoblotting of transcription factor AP-2α. 10-20 µl whole cell lysate containing 30 µg protein were blotted against β-actin as loading control (anti-TF-AP-2α-ab 1:1000). Immunoblots comprised the depicted amounts of recombinant human AP-2α as control. Blots are representative of three (six for THP-1) subsequent cell passages of (A) HL-60 in undifferentiated and differentiated (dHL-60), (B) THP-1 in undifferentiated and differentiated, (C) MM6 in undifferentiated and differentiated state, (D) Rec-1 and BL-41 and (E) 5-LO-negative controls U937 and Hek293T.

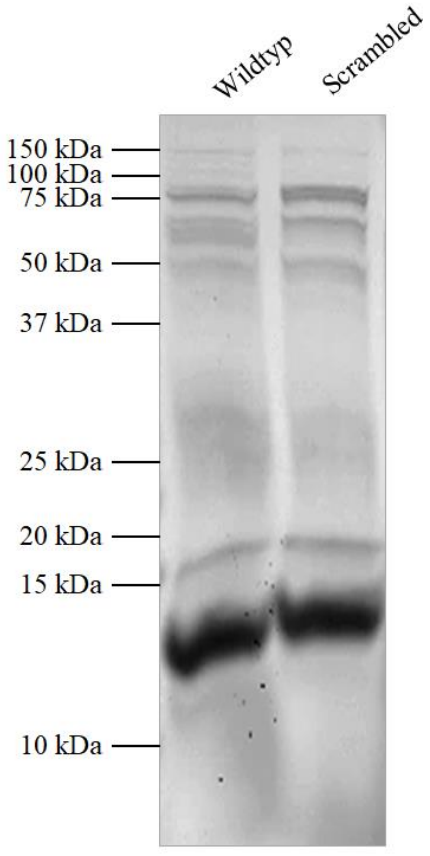
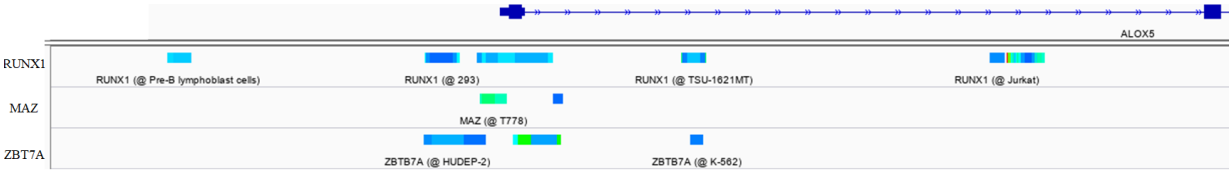
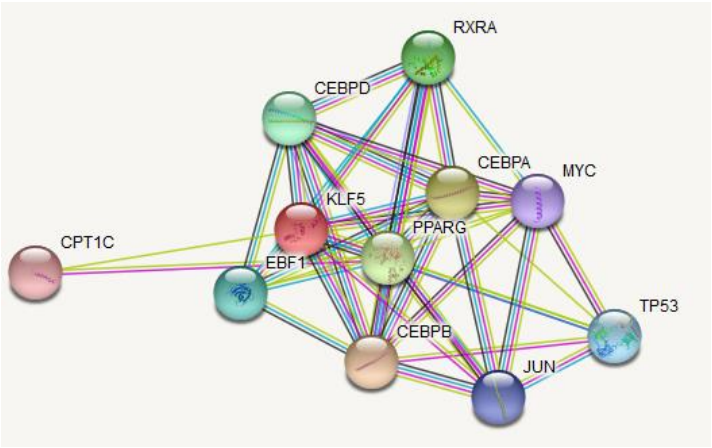


Figure 4: SDS-PAGE-based DNA pulldown with 800 µg of nuclear extract in each approach. NE was either incubated with WT or SCR ds-DNA and immobilized on magnetic beads. Proteins were separated by gel electrophoresis and subsequent in-gel tryptic digestion was applied with following MALDI-TOF-MS/MS for identifying proteins detected in both affinity enrichments.

Supplementary figure 5: Chip-Seq data (provided from IGV viewer) displaying binding sites for both RNA polymerase II and MAZ in the ALOX5 gene. MAZ possesses binding sites within the ALOX5 promoter, as well as in intron C and D, which correlate with the binding sites of RNA polymerase II, indicating a possible function in transcript elongation.



Supplementary figure 6: Chip-Seq data available for proving the binding of transcription factors RUNX1, MAZ and ZBT7A to the ALOX5 promoter (Data displayed by IGV viewer).



Supplementary figure 7: KLF5 interaction networks based on classification by STRING database.

9.2 Supplementary tables

protein	gene name	HL60	dHL60	THP1	dTHP1	MM6	dMM6	Rec-1	BL41
1. ABC3C	APOBEC3C				x				
2. BLM	BLM	x	x						
3. CIRBP	CIRBP							x	
4. CP7B1	CYP7B1			x					
5. CUL4A	CUL4A	x							
6. DDB2	DDB2		x						
7. DHX36	DHX36				x	x	x		
8. DPOLB	POLB		x						
9. ETV6	ETV6	x							
10. F120A	FAM120A							x	x
11. FAF1	FAF1						x		
12. FOSL2	FOSL2		x						
13. GANP	MCM3AP	x							
14. hnRNP D	HNRNPD	x		x		x			
15. hnRNP K	HNRNPK	x	x		x			x	x
16. hnRNP U	HNRNPU								x
17. Jun-D	JUND	x							
18. KLF13	KLF13	x						x	
19. KLF16	KLF16	x	x	x	x	x		x	
20. MAZ	MAZ	x			x	x			
21. MET17	METTL17			x					
22. MK	MDK					x			
23. MSD2	MSANTD2						x		
24. MSH2	MSH2	x							
25. MSH3	MSH3	x	x						
26. MTF2	MTF2	x							
27. NIP7	NIP7			x					
28. NU133	NUP133	x							
29. NU214	NUP214	x							
30. NUP53	NUP35	x							
31. NUP93	NUP93	x							
32. P52K	THAP12							x	
33. PCBP2	PCBP2		x						
34. PRD10	PRDM10		x		x				
35. RBM3	RBM3	x							
36. RBMS1	RBMS1						x		
37. RBP2	RANBP2	x							
38. RecQ1	RECQL	x							
39. RFA1	RPA1	x							
40. RFA2	RPA2	x					x		
41. RFA3	RPA3	x	x						

42. RFC1	RFC1	x	x						
43. RUNX1	RUNX1		x						
44. S35A5	SLC35A5			x					
45. SF3B1	SF3B1	x							
46. SMAL1	SMARCAL1		x						
47. Sp1	SP1	x	x	x	x	x		x	x
48. Sp2	SP2			x					x
49. Sp3	SP3	x			x	x	x		
50 Sp4	SP4				x				
51. SUZ12	SUZ12	x							
52. TCP4	SUB1	x							
53. TF3C5	GTF3C4	x							
54. TFAM	TFAM	x							
55. TM209	TMEM209	x							
56. VEZF1	VEZF1	x	x						
57. WDR76	WDR76	x							
58. YBOX1	YBX1		x						
59. ZBT7A	ZBTB7A	x	x		x				
60 ZF64B	ZFP64			x					
61. ZFP91	ZFP91					x			x
62. ZN148	ZNF148		x						
63. ZN281	ZNF281	x	x				x		
64. ZN316	ZNF316				x				
65. ZN444	ZNF444	x							
66. ZN579	ZNF579	x	x		x				

Supplementary table 1: Cumulatively identified proteins in all DNA pulldowns performed with core-promoter sequences of ALOX5 promoter.

protein	gene name	SBE		VDRE	
		HL-60	dHL-60	HL-60	dHL-60
1. BYST	BYSL	x			
2. CARTF	CARF	x		x	
3. CEBPA	CEBPA	x	x		
4. CEBPB	CEBPB		x		x
5. CEBPE	CEBPE		x	x	
6. CEBPG	CEBPG	x	x		
7. CHD1	CHD1	x			
8. CI114	SPOUT1		x		
9. DHX9	DHX9	x			
10. H15	HIST1H1B	x			
11. H2AY	H2AFY	x			
12. H4	H4C1	x			

13.	HIC2	HIC2		x	
14.	KLF12	KLF12		x	x
15.	KLF5	KLF5		x	
16.	LRRF2	LRRFIP2			x
17.	MAGBI	MAGEB18			x
18.	MCCA	MCCC1		x	x
19.	MCCB	MCCC2			x
20.	NFAC2	NFATC2	x		
21.	NR2C2	NR2C2		x	
22.	NSD1	NSD1		x	x
23.	ODB2	DBT		x	x
24.	PARP3	PARP3			x
25.	PERE	EPX	x		
26.	PON1	PON1			x
27.	PR38A	PRPF38A			x
28.	PRG2	PRG2	x		
29.	RL24	RPL24	x		
30.	RL27A	RPL27A	x		
31.	RL37A	RPL37A	x		
32.	RL5	RPL5	x		
33.	RL7	RPL7	x		
34.	RLF	RLF	x		
35.	RS13	RPS13	x		
36.	RS9	RPS9	x		
37.	SPTN1	SPTAN1			x
38.	TITIN	TTN	x		
39.	ZBT26	ZBTB26		x	
40.	ZFX	ZFX		x	x
41.	ZN148	ZNF148	x		
42.	ZN189	ZNF189		x	
43.	ZN212	ZNF212		x	
44.	ZN319	ZNF319		x	x
45.	ZN335	ZNF335		x	
46.	ZN668	ZNF668	x		
47.	ZN740	ZNF740	x	x	
48.	ZN775	ZNF775		x	
49.	ZNF24	ZNF24		x	
50.	ZSC25	ZSCAN25		x	

Supplementary table 2: Cumulatively identified proteins in DNA pulldowns with core-promoter proximal sequences of ALOX5 promoter

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides Statt, dass ich die vorgelegte Dissertation über

„Identification of proximal ALOX5 promoter interacting proteins by quantitative proteomics and evaluation of possible formation of secondary DNA structures“

selbstständig angefertigt und mich anderer Hilfsmittel als der in ihr angegeben nicht bedient habe, insbesondere, dass alle Entlehnungen aus anderen Schriften mit Angabe der betreffenden Schrift gekennzeichnet sind.

Immer, wenn eine Abbildung, Tabelle oder Textabschnitte identisch zu bereits veröffentlichen Publikationen sind, wurde es deutlich gekennzeichnet. Darüber hinaus ist hiermit angegeben, das Copyright Genehmigungen und/oder Ko-Autor Zustimmung eingeholt wurden.

Ich versichere, nicht die Hilfe einer kommerziellen Promotionsvermittlung in Anspruch genommen zu haben.

Frankfurt am Main, 29.05.2020

Katharina Schlag