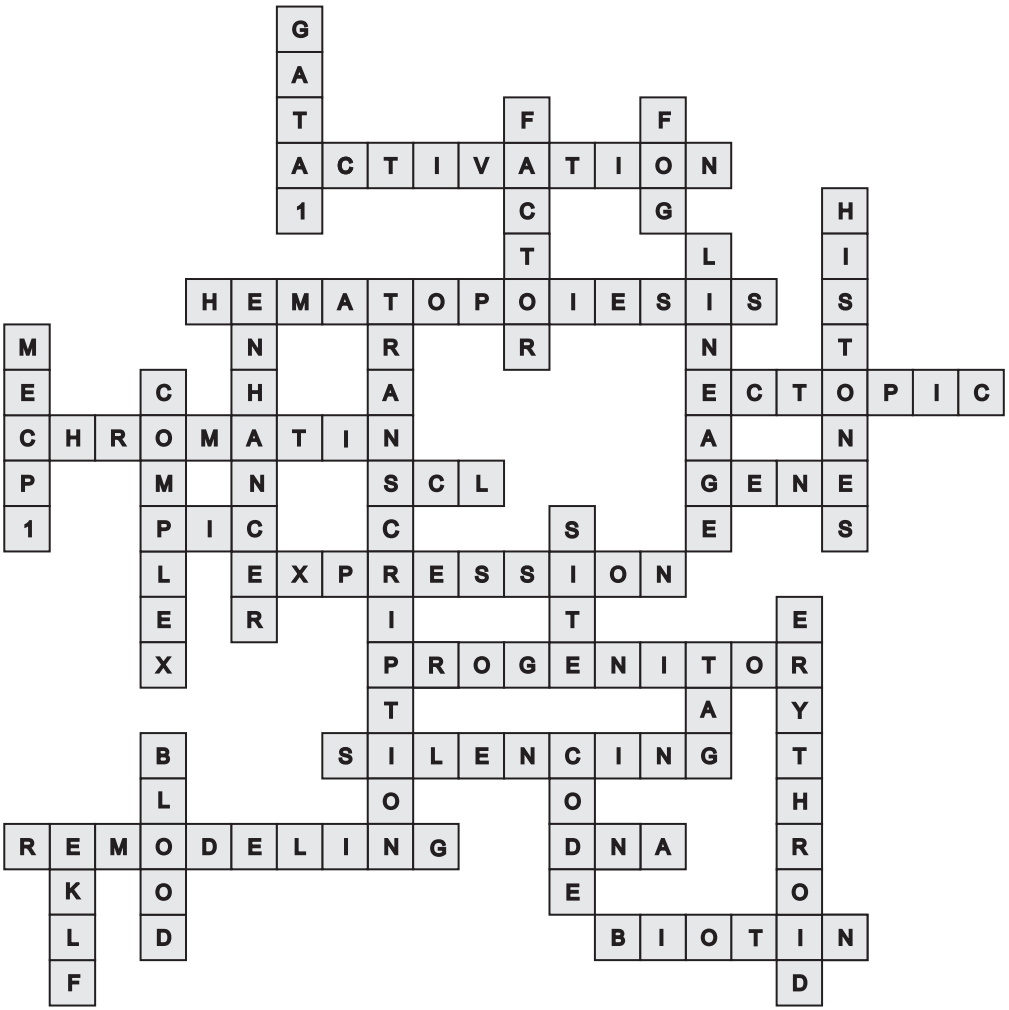
Characterization of Hematopoietic Transcription Factor Complexes in Erythroid Cells



Cover: Hematopoietic Scrabble.

The work presented in this thesis manuscript was performed in the Department of Cell Biology at the Erasmus MC Rotterdam, The Netherlands. The Department is a member of the Medisch Genetisch Centrum Zuid-West Nederland (MGC).
The research was supported financially by the EU. EC project title: Cellular memory: epigenetic regulation and inheritance of active and silent chromatin. EC project number: HPRN-CT-2000-00078. Printing of this thesis has been sponsored in part by Sigma and Tebu-Bio.

Characterization of Hematopoietic Transcription Factor Complexes in Erythroid Cells

Karakterising van hematopoietische transcriptie factor complexen in erytroïde cellen

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. S.W.J. Lamberts

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op
woensdag 11 januari 2006 om 15.45 uur.

door

Patrick Jean-François Rodriguez

geboren te Nîmes, Frankrijk

Promotiecomissie

Promotor:	Prof.dr. F.G. Grosveld
Overige leden:	Dr. J.N.J. Philipsen Prof.dr. E.A. Dzierzak Prof.dr. C.P. Verrijzer
Copromotor:	Dr. J. Strouboulis

A mes parents et à mon frère Christian,

- Table of Contents -

Abbreviations:

3C:	Chromosomal Conformation Capture
AGM:	Aorta-Gonad-Mesonephros
AMKL:	Acute MegaKaryoblastic Leukemia
ATP:	Adenosine Tri-Phosphate
BFU-E:	Burst Forming Unit-Erythroid
cDNA:	complementary DesoxyriboNucleic Acid
ChIP:	Chromatin ImmunoPrecipitation
CLP:	Common Lymphoid Progenitor
CMP:	Common Myeloid Progenitors
CTD:	C-TerminalDomain
E:	Embryonic Day
EMP:	Erythroid and Megakaryocytic Progenitors
ES:	Embryonic Stem
FDCP:	Factor Dependent Cell-Paterson mix cells
GM-CSF:	Granulocyte-macrophage colony-stimulating factor
GMP:	Granulocytes/monocytes Progenitors
GPA:	GlycoPhorin A
H:	Histones
HAT:	Histone AcetylTransferase
HDAC:	Histone DeACetylase activity
HLH:	Helix-Loop-Helix motif
HMT:	Histone-MethylTransferases
HS:	Hypersensitive Sites
HSC:	Hematopoietic Stem Cell
IL-3:	Interleukin 3
MAPK:	Mitogen-Activated Protein Kinase
MBP:	Major Basic Protein
MEL:	Mouse ErythroLeukemia
MLP:	MultiLineage Progenitors
PIC:	Pre-Initiation Complex
PML:	Promyelocytic Leukemia Protein
PRMT:	Protein Arginine MethyltTransferases
PTM:	PostTranslational Modifications
REDS:	Red Cell Differentiation Signal
RT-PCR:	Reverse Transcriptase PCR
SUMO:	Small Ubiquitin-Related Modifier
T-ALL:	T-cell Acute Lymphoblastic Leukemia
TMD:	Transient Myeloproliferative Disorder
YS:	Yolk Sac

Scope of the thesis

All blood cells are derived from a self-renewing population of pluripotent hematopoietic stem cells (HSCs). Each HSC has the potential to progressively differentiate into more committed progenitors to generate the mature cells. Hematopoietic differentiation is characterized by a concomitant activation and repression of specific transcriptional programs, accompanied by changes at the chromatin structure level. The cellular environment including signaling molecules and growth factors and, ultimately, transcription factors regulate the complex balance of proliferation of HSCs and progenitors versus terminal differentiation and growth arrest. Transcription factors bind directly to regulatory elements to modulate gene expression in combination with additional ubiquitous or tissue-restricted transcription factors, and/or by tethering co-factors such as chromatin remodeling and/or modifying enzymes. Gene targeting and ectopic expression of transcription factors have highlighted their function, however such approaches do not not directly address their molecular basis. Identification of transcription factor-interacting partners and their molecular function will therefore provide a detailed characterization of the protein regulatory network, essential for a better understanding of the molecular mechanisms underlying their function in hematopoiesis. This has been the focus of the work of my thesis.

Characterization of protein complexes from mammalian cells requires the availability of powerful technologies for protein purification. **Chapter 2** describes a new technique based on the single step, high affinity pull-down of an *in vivo* biotin-tagged protein using streptavidin beads. Using this technique, it is shown in **chapter 3** that the hematopoietic transcription factor GATA-1 is part of at least 5 distinct complexes in mature erythroid cells, including previously described as well as novel protein partners that are involved in gene activation and gene repression. In **Chapter 4,** chromatin immunoprecipitation (ChIP) analysis of gene targets shows that distinct GATA-1 complexes are stably bound to different subsets of target genes. We also identified a mechanism for GATA-1-mediated repression of an early hematopoietic program. Finally, **Chapter 5** describes the complementary characterization of protein complexes formed by Ldb1, a GATA-1 interacting partner, thus illustrating the importance of a cellular protein network. Our work provides evidence of a central role for GATA-1 in erythroid differentiation. GATA-1 through its interacting partners is linked on the one hand to repression of early or alternative hematopoietic programs and of cell proliferation related genes, and on the other hand to activation of erythroid specific genes.

Chapter 1:

Introduction

1. Hematopoiesis/Erythropoiesis

1.1 Introduction

Hematopoiesis is the generation of all the mature blood cells from a rare pool of hematopoietic stem cells (HSCs). Due to their limited life span, blood cells need to be continuously generated throughout life. Blood contains many types of cells executing different functions. These functions range from the transport of oxygen to all the cells, to the defence of the organism. Blood consists of red blood cells or erythrocytes, white blood cells, which include granulocytes (neutrophils, eosinophils and basophils), monocytes and lymphocytes, and platelets (Figure 1). Remarkably all classes of blood cell are derived from a single common multipotent progenitor, the so-called hematopoietic stem cell. In adults, HSCs are present primarily in the bone marrow at a very low incidence (1 to 10 cells in 100,000 cells). Specifically, HSCs have the capacity to self-renew and possess a high proliferative potential[1]. The hematopoietic system has been studied extensively and is a model of choice to study cellular and molecular mechanisms of differentiation. In addition, human diseases of the hematopoietic system, such as leukemias and hemoglobinopathies, are a serious health problem.

Differentiation during normal hematopoiesis occurs through a stepwise process, each step generating increasingly restricted hematopoietic cells. The first division of the pluripotent stem cell leads to one daughter cell keeping the stem cell characteristics, and the other daughter cell proceeding a differentiation pathway. The first progenitors to be generated from HSCs are the so-called multilineage progenitors (MLPs). These cells have the capacity to differentiate into any type of blood cell but have lost the capacity to self-renew. MLPs can differentiate into either common lymphoid progenitor (CLP) or common myeloid progenitors (CMP)[2]. Once the decision for a cell fate is made, progenitors display a progressively more lineage-restricted potential at each branch point. In order to produce a sufficient number of mature blood cells, every progenitor at each step of differentiation goes through many rounds of cell division generating identical cells.

During lineage differentiation, the progressive restriction multipotency is underscored by the expression of unique combinations of transcription factors. These transcription factors act together with chromatin-remodeling and modifying co-factors to establish a particular lineage-specific gene expression program[3,4]. Commitment of a precursor cell to a specific lineage is accompanied by expression of the lineage-specific genetic program, as well as the concomitant suppression of programs associated with early multipotential states and of alternative lineages. While it has been generally accepted that commitment is an irreversible process, recent evidence has suggested it can actually be reversed[5,6]. Transcriptome analysis of HSC populations revealed co-expression of specific lineage affiliated genes such as erythroid, megakaryocytic and myeloid programs[3], suggesting that these cells are “primed” for the different lineage-specific genetic programs. In addition, forced expression of a transcription factor such as GATA-1 in myeloid cells was shown to lead to the generation of erythroid cells, suggesting a reprogramming of the cells to a different hematopoietic program[7].

In addition to lineage switches within the hematopoietic system, several lines of evidence have suggested that HSCs possess a remarkable plasticity for non-hematopoietic cell differentiation, called transdifferentiation[8]. For example, HSCs from bone marrow were suggested to give rise to neuronal cells[9]. Such phenomena have extraordinary consequences on the understanding of cell differentiation and ultimately on developing therapeutic treatments.

However, the basis of the transdifferentiation potential of HSCs is poorly understood and very controversial[10].

1.2 Hematopoiesis during Development

In mammals, the yolk sac (YS) is the first extraembryonic hematopoietic tissue to be formed. At around day 7 of gestation (E7) in mouse, blood islands emerge from the mesoderm. Within these islands, blood cells arise in close juxtaposition to endothelial cells. Endothelial and hematopoietic cells are thought to arise from a common progenitor, the hemangioblast[11]. In the early stage embryo, hematopoietic cells differentiate to give rise to primitive erythroblasts. These cells enter the circulation of the embryo proper and continue to divide for several days. Upon differentiation, they express embryonic globin genes. Primitive erythrocytes are large and nucleated[12]. Hematopoiesis in the yolk sac occurs until around E13 in mouse[13].

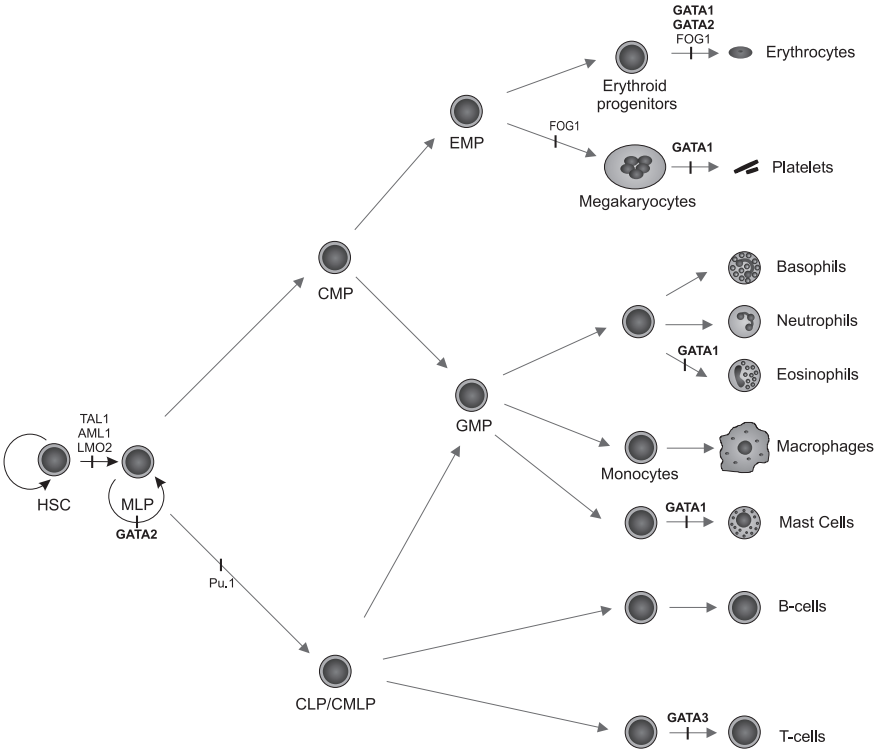


Figure 1: Schematic representation of the hematopoietic hierarchy

The Hematopoietic Stem Cell (HSC) gives rise to all mature blood cells. During hematopoietic differentiation, multipotent progenitors (MLP) have similar progenitor potential than the HSC, but have lost their ability to self-renew. They differentiate into Common Myeloid Progenitor (CMP) and Common Lymphoid Progenitor (CLP). CMP are at the origin of Erythroid/Megakaryocytes Progenitors (EMP) and the Granulocyte/monocyte Progenitor (GMP). Recently, it was shown that a progenitor called CMLP had the potential to give rise to CLP and also GMP[14]. EMP will give rise to erythroid and megakaryocytes. GMP will differentiate into granulocytes, including eosinophils, macrophages and mast cells. In lymphoid lineage, CLP give rise to B- cells and T-cells.

In contrast to primitive hematopoiesis, definitive hematopoiesis generates HSCs that give rise to all blood cells lineages, including the definitive erythrocytes. Definitive and primitive erythroid cells differ in that definitive red cells are smaller, enucleated cells that contain adult globin genes. However, transcription factors such as GATA-1, GATA-2, Lmo2 are expressed in both types of erythroblasts. The para-aortic splanchnopleura/aorta-gonad-mesonephros (PAS/AGM)[15,16] is the first hematopoietic tissue to arise within the body of the embryo proper. The PAS represents the precursor tissue to the AGM region that produces the first HSC at E10/11[17]. By E11.5, these cells migrate and colonize the fetal liver, which is considered as the principle hematopoietic tissue in the developing foetus[18,19]. At E12, the fetal liver exhibits a large increase of the number of HSCs, suggesting a role in the specific expansion of those cells[18,19]. As the embryo develops, new hematopoietic reservoirs are formed in tissues such as the thymus and the spleen. After birth, HSCs migrate from the fetal liver to the bone marrow, which constitutes the main hematopoietic site throughout life. Recently, another source of HSCs was identified in the placenta[20].

1.3 Regulation of Hematopoiesis

Hematopoiesis is often used as a model system to study the mechanisms involved in cell differentiation within a very tightly regulated environment. The production of all mature blood cells from one common progenitor is a stepwise process, influenced at different levels and at each branch point by the cellular microenvironment, signaling cues and the expression of a specific combination of tightly regulated transcription factors. The detailed roles of transcription factors during differentiation will be presented later in the introduction section. However the time, space and level of expression of such molecules are crucial factors in determining and executing cell fate. In addition to transcription factors, molecules from the environment such as cytokines and growth factors coupled to signaling pathways play a critical role, especially by regulating downstream transcription factors.

The molecular mechanisms by which cell fate is determined remains poorly understood. Two hypotheses have suggested that the decision can be either instructive or stochastic[21,22]. In the first case, the decision for a specific cell fate is directly determined by extrinsic signals. In opposition, the stochastic model predicts that commitment to a particular hematopoietic lineage is independent of the extrinsic signals. Instead, extrinsic signals are required for survival and differentiation of the committed cell[23]. For example, the targeted mutation of the erythropoietin gene or of its receptor leads to embryonic lethality because of reduced primitive erythropoiesis and an absence of definitive erythropoiesis *in vivo.* However, the presence of BFU-E and CFU-E progenitors in homozygous knockout mice, demonstrated that these molecules are not essential for the erythroid lineage commitment “per se” [24].

Cytokines are crucial for cellular survival and expansion. For example, interleukin 3 (IL-3) and Granulocyte-macrophage colony-stimulating factor (GM-CSF) stimulate proliferation of multipotential hematopoietic progenitors[25-27]. The latter also has a role in the proliferation of granulocytes and macrophages. Some other cytokines have a role in promoting cellular differentiation. For example, erythropoietin, IL-7 and IL-5 stimulate erythroblast, lymphoid and eosinophil differentiation, respectively[28,29]. In adult, most hematopoietic cytokines are produced by the bone marrow stroma cells[30].

Developmental signaling pathways such as the Notch and Wnt pathways also play critical roles in the regulation of hematopoiesis[31]. Moreover, cell adhesion molecules such as

collagens and laminin appear to act as regulators of hematopoiesis, but their exact function remains to be elucidated[32,33].

1.4 Definitive Erythropoiesis

With a lifespan estimated at 120 days in human, billions of red blood cells need to be produced each day. Erythrocytes, or red blood cells, are by far the most abundant cells in the blood comprising about 45 % of its volume. They are very rich in hemoglobin, the molecule responsible for binding and transporting oxygen. The expression of the globin genes is differentially regulated during development since the various globin genes provide specific physiological needs of the developing organism. Definitive erythrocytes are tiny biconcave discs, a shape that increases the efficiency of oxygen diffusion.

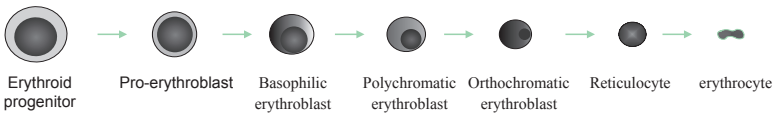


Figure 2: Schematic representation of different stages of erythropoiesis
The erythroid progenitor coming from the EMP differentiates into the first erythroid differentiated cell morphologically recognizable, the pro-erythroblast. Erythroid terminal differentiation leads to the production of mature red blood cells.

Generation of mature red blood cells is a tightly regulated multistep process. Erythroid cells derive from HSCs as mentioned above. Erythroid and megakaryocytic progenitors (EMPs) coming from the common myeloid progenitor (CMP) are restricted to generate only the erythroid or the megakaryocytic cells. The first erythroid specific progenitor to be formed is referred as the Burst Forming Unit-Erythroid (BFU-E). The process of maturation from BFU-E to active circulating erythrocytes, the circulating red blood cells, is characterized by a progressive differentiation of several progenitors (Figure 2), in which the first morphologically recognizable differentiated erythroid progenitor is the proerythroblastic cell. Proerythroblasts correspond to large cells of 16 μm in mean diameter with a nucleus that occupies 80 % of the cell. Differentiation from the proerythroblastic stage to enucleated cells is referred to as terminal erythroid differentiation. It occurs in a defined structure, the so-called erythroblastic island (Figure 3[33]), that was first identified by electron microscopy[35]. Terminal erythroid differentiation is characterized by the accumulation of high amounts of hemoglobin. In the final steps, the most differentiated erythroblastic cells, the orthochromatic erythroblasts, expel their nucleus to become reticulocytes, which represent the last stage of immature red cells that further lose specific organelles to generate the fully mature erythrocytes.

1.5 Hematopoiesis as a Biological Model

1.5.1 Multi-lineage priming

One of the main questions in the process of generating all hematopoietic lineages from a single HSC is to understand how transcription of a lineage-specific program is initiated. One hypothesis is that precursors cells are transcriptionally “primed”, meaning that they promiscuously express lineage-specific genes, albeit at low levels, prior to commitment and differentiation[36,37]. Expression analysis and microarray data from purified HSCs have

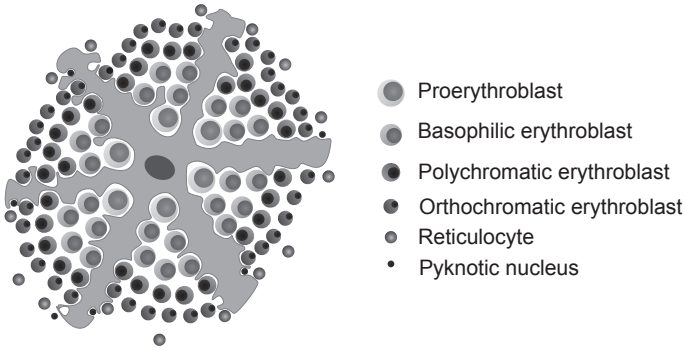


Figure 3: Erythroblastic island
The central macrophage is surrounded by erythroid cells. Distribution of erythroid cells depends on their stage of differentiation; the more immature cells are the closest to the centre. Adapted from reference 33.

demonstrated the presence of transcripts of multiple programs of hematopoietic lineages[3]. Analysis of MLPs shows co-expression of myeloid and lymphoid programs, and more restricted progenitors CMPs and CLPs express myeloid/erythroid genes and T and B lymphoid genes, respectively. Furthermore, expression analysis by Reverse Transcriptase PCR (RT-PCR) in a multipotential cell population has demonstrated that lineage-specific genes were co-expressed at the single cell level even before commitment[38]. For example, transcription factors such as GATA-1 and C/EBPα are expressed in the early progenitor cell CMP, GM and Meg/E related genes, respectively (Figure1). Presumably, this multilineage priming would allow progenitors to respond rapidly to changes in environmental signals and cues. An essential implication of such promiscuous expression is that chromatin would be in an “open structure”, thus allowing many key cis-regulatory elements of lineage-affiliated genes to be accessible for transcription prior to lineage commitment[37]. For example, erythroid specific DNase hypersensitive sites (HS) 1 and 2 of globin genes are sensitive in self-renewing multipotential FDCP cells (Factor Dependent Cell-Paterson mix cells)[39]. Similarly, in the same cells, enhancers of the T-lymphoid specific CD3 delta-chain and B-lymphoid specific immunoglobin heavy chain genes are also DNase I sensitive[40]. Collectively, these observations suggest that progenitors have a broadly open chromatin structure that allows transcription of several lineage specific programs supporting the model by which progenitors cells display promiscuous and simultaneous transcription of multiple hematopoietic differentiation programs. During hematopoietic differentiation, promiscuous expression of lineage-affiliated genes would become progressively more restricted as cells become more specialized.

Main questions arising from these observations are how one program is selected and what is the role of the transcription factors in that process?

1.5.2 Cell reprogramming

Cell commitment has been generally considered as an irreversible event. Instead, several recent lines of evidence have suggested that commitment of hematopoietic progenitors is more dynamic than previously thought. Cell reprogramming was shown to take place by the forced expression of oncogenes, cytokines and transcription factors in

multipotential cell lines[41-43]. The forced expression of oncogenes, for example v-raf, reprogrammed a lymphoid cell line into a myeloid cell fate[41]. Similarly, chicken erythroid/megakaryocytic progenitors transformed with the Myb-Ets encoded by the E26 leukemia virus (Myb-Ets progenitor, MEPs) could be differentiated into myeloblasts and eosinophils[8]. Remarkably, the ectopic expression in MEPs of lineage-specific transcription factors such as C/EBPα, a member of the CCAAT/enhancer-binding protein (C/EBP), or GATA-1 leads to the generation of eosinophils and megakaryocytic cells, respectively. Similar to expression of C/EBPα, the ectopic expression of PU.1 converted MEPs into eosinophils[44]. Interestingly, PU.1 expression leads to the repression of GATA-1 activity. Conversely, GATA-1 was also shown to antagonize PU.1 function in MEPs[45,46]. These data suggest hematopoietic cells remain able to give rise to lineages different than the one they were initially programmed for. At the molecular level, cell reprogramming involves activation of a specific lineage as well as repression of alternative lineages. Similar experiments have also been carried out with primary mammalian progenitor cells. HSCs, CMPs and CLPs were reprogrammed through the ectopic expression of lineage-restricted transcription factor GATA-1[47].

How GATA-1 is also able to reprogram lineage decisions will be detailed below.

2. Hematopoietic Transcription Factors

2.1 Transcription: General Principles

2.1.1 Cis-regulatory elements for gene regulation

Proper spatial and temporal gene expression is dependent on the presence of cis-regulatory DNA sequences, such as promoters, enhancer/silencers, Locus Control Regions and DNase I Hypersensitive sequences. These regions of DNA are characterized by their accessibility to regulatory factors. Such elements contain specific sequences that are recognized and bound by specific transcription factors that regulate gene expression.

2.1.2 Basal machinery

Transcription in eukaryotic cells is achieved by three RNA polymerases regulating three classes of genes. RNA pol I and III transcribe ribosomal RNA and small nuclear RNAs, respectively, whereas the vast majority of protein coding genes are transcribed by RNA pol II[48]. Typically, RNA pol II transcription begins with the binding of gene specific regulators at the promoter, near the site of transcription initiation. These factors modulate the activity of the transcription machinery activity either indirectly by recruiting chromatin remodeling/modifying enzymes and/or directly by interacting with components of the basal transcription machinery. Both mechanisms result in the recruitment of the transcription machinery to a core promoter. This positions the pre-initiation complex[49,50] (PIC), which consists of the RNA pol II enzyme and the well-characterized general transcription factor complex TFII (transcription factor for RNA pol II) consisting of the TFIIA to -H subunits, except the TFII E. However, transcription cannot yet start, as the machinery is still in an inactive state. The entry of TFII E into the complex causes a conformational change resulting in an open complex that allows transcription initiation to take place. After 30 bp of RNA synthesis, the RNA pol II C-terminal domain (CTD) is phosphorylated thus releasing the complex from the core promoter and allowing transcription elongation to proceed.

2.1.3 Transcription factors

Transcription factors are direct effectors as they recognize and bind to specific DNA sequence motifs located in regulatory sequences. Broadly speaking, tissue-specific transcription factors interact with ubiquitous transcription factors and (non DNA-binding) co-factors which can alter chromatin structure and/or tether the basal machinery to the gene, thus activating transcription. It is also important to note that such interactions can also result in repression through the recruitment of repressive complexes to gene targets acting at the chromatin level.

Transcription factors are classified into families based on their functional domains, such as the DNA binding domain. One of the first DNA binding domain to be discovered was the helix-loop-helix motif (HLH), composed of two α-helix domains separated by a stretch of amino acid forming a loop. Zinc-finger motifs represent another protein domain that can bind DNA with very high specificity and affinity. Additionally, the leucine zipper domain is another example of a DNA-binding motif[48]. This motif also mediates protein homodimerisation which inhibits the DNA binding activity. The specificity of DNA recognition and binding by a transcription factor is very high. Therefore, the smallest change in the DNA or the amino acid sequence of the DNA binding domain will modify the binding characteristic. Different outcome of target gene regulation resulting from such changes can lead to disease, for example.

These DNA binding domains are functionally critical as they mediate specific classes of proteins to interact and to form functional complexes that are important for proper gene regulation. Examples of such interactions include GATA-1 and Friend of GATA-1 (FOG-1) as well as GATA-1 and the pentameric complex formed between GATA-1 and the TAL-1/Ldb1/Lmo2/E2A transcription factors[51,52]. The main focus of this thesis revolves around tissue specific, i.e. erythroid transcription factors and is described in greater detail below.

2.2 Hematopoietic Transcription Factors

Once HSCs differentiate to a multipotential progenitor (MLP), the genetic programs that specify a defined lineage are characterized by changes in the nature and the level of the transcription factor complement being expressed. Much evidence, mostly from gene knockout experiments but also from the analysis of hematological disorders such as leukemias, has accumulated showing that hematopoietic restricted transcription factors play a very important role in these processes. In addition, microarray expression analysis have provided large amount of information on the combinations of genes being switched on and off during differentiation and development. However, details of the molecular mechanisms at the chromatin level of how the regulation of these genes occurs remain scant. For example, gene knockouts have demonstrated an essential function in HSCs for GATA-2, TAL-1/SCL, AML1, Lmo2[53-56] (review[57]). Other factors such as PU.1, C/EBPα, Ikaros, GATA-1, FOG-1 and Gfi-1b have been shown to be essential for the differentiation of myeloid, eosinophilic, lymphoid and Meg/E lineages[58-62].

The analysis of the gene knockout phenotypes of transcription factors relevant to the scope of this thesis will be presented in greater detail later. However, it is important to note here that ablation of seemingly unrelated transcription factors can result in a similar phenotype. For example, the FOG-1, Gfi-1b and GATA-1 knockouts have very similar phenotypes in the megakaryocytic and erythroid lineages. In addition, the conditional knockout of TAL-1 in erythroid cells leads to a phenotype similar to that of GATA-1. Not surprisingly, FOG-1 and TAL-1 are indeed interacting partners of GATA-1[51,63].

2.3 GATA Factors

The GATA transcription factors form a family of 6 evolutionarily conserved members, all of which contain two characteristic zinc finger domains. GATA factors bind the consensus DNA sequence A/TGATAA/G via a conserved zinc-finger Cys-X_2-Cys-X_{17}-Cys-X_2-Cys present in the carboxy terminal end of each protein (CF). A second GATA finger is more proximal to the amino-terminus of the protein (NF).

The family is divided into two subgroups based on their expression pattern. The hematopoietic subfamily is composed of GATA-1, -2 and -3, each being an essential transcription factor at different stages and lineages in the generation of blood cells (see below) as their gene knockouts result in early embryonic lethality.

GATA-2 is highly expressed in early hematopoietic progenitors but also in megakaryocytes and mast cells. GATA-2 null mice are characterized by a lack of definitive red blood cells and severely reduced yolk sac hematopoiesis that results in a severe anemia with embryonic lethality at E10-E11[53]. Furthermore, analysis of GATA-2 function has demonstrated GATA-2 as being essential in the expansion of multipotential hematopoietic progenitors[64]. It is also essential for expansion of HSCs in embryos and adults. Forced expression of GATA-2 in multipotential hematopoietic cells was shown to reduce erythroid differentiation[65]. These data suggest that during erythroid differentiation GATA-2 expression needs to be downregulated. In addition, GATA-1 null cells exhibit high level of GATA-2 expression, suggesting that GATA-1 represses GATA-2[66,67]. Recent data have demonstrated a direct repression of GATA-2 by GATA-1[68-70]. Taken together, these data suggest a close interplay between GATA-1 and GATA-2 during hematopoiesis.

GATA-3 is expressed in T lymphoid cells in the hematopoietic system, but also in a number of additional tissues such as the nervous system, skin, liver and kidney. Loss of GATA-3 function in mice shows a severe phenotype resulting in a early lethality due to internal bleeding and brain malformations[71]. GATA-3 is expressed in PSP and has been suggested to play a role in generation of the hematopoietic system in the embryo.

GATA-4, -5 and -6 are mainly expressed in other tissues such as the heart, lung and intestine[72,73]. The GATA-4 and GATA-6 knockouts result in embryonic lethality, due to defects in heart morphogenesis and endoderm lineage generation, respectively[74-76]. Among the GATA factors, only the GATA-5 knockout does not present any phenotype. Due to the topic of the thesis research, we will focus on GATA-1 protein and its partners.

2.3.1.1 GATA-1: gene and protein structure

GATA-1, the prototypic member of the GATA factor family, also known originally as NF-E1, NF-1, ery-1 and GF-1, is expressed in primitive and definitive erythroid[76], megakaryocytic[78,79], eosinophilic[80] and mast cells[78,81] and in the Sertoli cells of the testis[82]. GATA-1 was originally identified as a factor binding to conserved DNA sequences within the regulatory elements of globin genes[83,84]. It is now known that GATA-1 is a key regulator of gene expression of a very large number of genes in multiple cell types.

The GATA-1 transcriptional unit contains multiple cis-regulatory elements essential for its expression. The GATA-1 gene is X-linked and is composed of 7 exons, two promoters located in the two alternative non-coding first exons IE (erythroid specific) and IT (testis specific)[85] and five coding exons[86] (Figure 4). Both promoters lack a TATA box[86,87], but contain GATA binding sites which are required for the proper expression of the gene, thus suggesting an auto-

regulatory loop[85,86,88,89]. Since the translational start is in the exon II, the two promoters generate a unique GATA-1 protein. Disruption in mice of the specific-erythroid promoter (IE) leads to an arrest in primitive erythropoiesis while not affecting GATA-1 expression in Sertoli cells[90].

Transgenic mice expressing a reporter gene under the control of the different presumptive elements of GATA-1 and specific knockout mice have demonstrated that the different elements are required for spatial and temporal expression of GATA-1[88,89,91]. Importantly, the -3.9 kb upstream region of the IE promoter shown to be sufficient to recapitulate GATA-1 expression in primitive as well as definitive erythropoiesis contains a GATA site and an E-box[88,89].

The GATA-1 protein is composed of 413 amino acids with an apparent molecular weight of approximately 48 kDa (Figure 4). It contains at least three functional domains: the two characteristic GATA Zn fingers (NF= N-terminal zinc finger and CF= C-terminal zinc finger) as well as an activation domain specific to GATA-1 which encompasses the N-terminal 80 amino acids of the protein. The activation domain was originally defined in *in vitro* transfection experiments using fibroblasts. Deletion of this domain failed to confer transcriptional activation of a reporter gene[78]. However, the function of the activation domain in hematopoiesis remains unclear since deletion of the first 63 amino acids of the GATA-1 protein were dispensable for erythroid and megakaryocytic differentiation[92,93]. Nevertheless, it is important to note that erythropoiesis is impaired in transgenic mice expressing a GATA-1 mutant (NT) that is deleted for the first 83 amino acids of GATA-1[94]. Severity of the phenotype was directly linked to the level of expression. Mice expressing a comparable level of endogenous and NT mutant were impaired for definitive erythropoiesis. In contrast, when the NT mutant was expressed at higher level than the endogenous GATA-1, the phenotype was rescued, thus suggesting a role for the NT in hematopoiesis. Recently, the GATA-1 gene was engineered as to express GATA-1s (deletion of the first 63 aas). Expression of GATA-1s in mice led to hyperproliferation of a novel hematopoietic progenitor[95].

Of the GATA-1 zinc fingers, the CF is essential for recognizing and binding to the consensus GATA DNA motifs[78,96]. It is of note that virtually all erythroid genes contain such motifs[57]. The CF is also involved in interactions with other transcription factors such as EKLF, Sp1 and PU.1[97,98]. The NF does not bind directly to GATA DNA consensus sites. It is involved in the stabilization of DNA binding by the CF. However, it was shown in erythroid cells to bind to different consensus sequences than the GATA sequence recognized by the CF[99]. Trans-activation experiments in fibroblasts have shown that the NF can mediate DNA binding to the non-consensus binding site GATC thus activating transcription[100]. The NF is mainly involved in mediating interactions with a number of transcription factors and co-factors: FOG-1[51], EKLF[97], CBP/P300[101], TAL-1[63], Lmo-2[63] and also in mediating GATA-1 homodimerization or multimerization[102,103].

Evidence for the requirements of the GATA-1 zinc fingers in hematopoiesis has been obtained from rescue experiments of the GATA-1 knockout phenotypes. Early studies in the GATA-1 deficient proerythroblastic G1E cell line have demonstrated that both Zinc fingers are crucial GATA-1's ability to drive terminal erythroid differentiation[92]. More recently, *in vivo* evidence for the function(s) of each Zn finger in primitive and definitive erythropoiesis was obtained by expressing GATA-1 zinc finger mutants in GATA-1 knockdown mice, i.e. mice that expressed only a small percentage of wild type levels of GATA-1 protein[94]. In agreement with the data previously published, these experiments showed that both zinc fingers were essential for erythropoiesis since no live pups were obtained. However, deletion of the CF led to impairment

of both primitive and definitive erythropoiesis, whereas deletion of the NF led to normal primitive erythropoiesis only. These observations suggest differential requirements for the GATA-1 zinc fingers in primitive and definitive erythropoiesis[94].

Taken together, these data suggest a very important role for GATA-1 in definitive erythropoiesis and that all of its domains are required at this stage.

2.3.1.2 GATA-1 posttranslational modifications (PTMs)

Like many other factors, GATA-1 is posttranslationally modified in several ways. However, the functions of GATA-1 PTMs remain largely unclear.

GATA-1 can be acetylated *in vitro* and *in vivo* in erythroid cells by the two ubiquitously expressed histone acetyltransferases (HATs) P300/CPB[104,105]. Acetylation occurs on two highly conserved lysine residues present in the C-terminal part of each zinc-finger (Figure 4). GATA-1 acetylation in transfected fibroblasts was shown to increase its binding to DNA as well as GATA-1-mediated transcriptional activation of a reporter gene, when co-expressed with p300[104]. However, in another study in erythroid cells, GATA-1 acetylation did not result in any changes in DNA binding. Nevertheless, the same study using the GATA-1 null proerythroblastic G1E cells showed that GATA-1 acetylation was required for terminal erythroid differentiation[105].

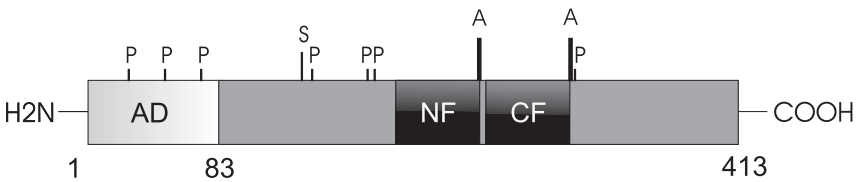


Figure 4: GATA-1 protein
The first 83 amino-acids correspond to the activation domain (AD). N-terminal zinc-finger (NF) is mainly involved in protein-protein interactions, whereas the C-terminal zinc finger mediates DNA binding to GATA sites.
On the top of GATA-1 protein are shown post-translational modifications. P: Phosphorylation; A: acetylation; S: Sumoylation.

Phosphorylation of GATA-1 can occur at seven serine residues[103]. Of these, six serines are phosphorylated in non-differentiated MEL cells. Phosphorylation of S310, which is in proximity to the DNA binding domain, appears only during the erythroid MEL differentiation. Phosphorylation of GATA-1 in K562 cells has been associated with an increase of its binding to DNA[106]. However, specific mutations of all serines to alanines of the murine GATA-1 did not influence GATA-1 binding or transcriptional activity *in vitro*[107]. Signaling pathways such as mitogen-Activated Protein Kinase (MAPK) activated by IL-3 or EGF, and phosphatidylinositol 3-kinase/Akt (PI3K) were shown to lead to specific GATA-1 phosphorylation[108-110]. In addition these reports show a function of phosphorylated GATA-1 in gene regulation. For example, transfection of GATA-1 in COS cells have shown that GATA-1 phosphorylation transactivates the survival gene E4bp4 and the tissue inhibitor of matrix metalloproteinase (TIMP-1)[109,110]. Mutation of serine 26 inhibits the induction of the anti-apoptotic gene Bcl-X_L[109]. Mutation of the two serines (Ser 26 and 178) in GATA-1 compromises interactions with its partner Lmo2 potentially affecting GATA-1 functions[108]. Despite these data, the precise function of GATA-1 phosphorylation in erythroid cells remains to be elucidated.

Recently, GATA-1 was also shown to be sumoylated *in vitro* and *in vivo*[111] (and personal unpublished observations). The small ubiquitin-related modifier (SUMO) is covalently ligated to a lysine residue within the consensus sequence LKXE (for review[112]). The modification process is similar to the ubiquitin modification pathway and includes three enzymatic steps. The GATA-1 consensus site for SUMOylation LKTE is located at position K137. Several functions have been associated with SUMOylation of transcription factors, such as gene repression or activation for Sp3 and p53, respectively[113,114], protein localization of the promyelocytic leukemia protein (PML)[115], protection from ubiquitin-mediated protein degradation[116]. However, any function(s) associated with GATA-1 sumoylation remain to be defined.

2.3.1.3 GATA-1 functions in hematopoiesis

In the hematopoietic system, GATA-1 is expressed at high levels in erythroid, megakaryocytic, eosinophilic and mast cells. GATA-1 expression is also detected in hematopoietic progenitors at lower levels. The levels of GATA-1 protein are dynamic during erythroid differentiation. GATA-1 is thought to be activated in progenitors, reaches maximal levels in erythroblasts and decreases prior to the last stage of terminal differentiation[117,118]. Thus, regulation of GATA-1 levels appears to be very tight.

GATA-1 is an essential transcription factor for normal erythropoiesis. Chimeric mice obtained from GATA-1 null embryonic stem (ES) cells are able to generate many hematopoietic lineages but not mature red blood cells[60]. More detailed genetic analysis has shown that GATA-1 null erythroid cells are not able to differentiate beyond the proerythroblastic stage due to apoptotic cell death[119,120]. Deletion of the GATA-1 gene in mice results in a lethal phenotype between E10.5-E11.5 due to severe anemia[121]. GATA-1 knockdown mice, in which the level of GATA-1 is 5 % of the physiological level (GATA-1.05 mice), lack primitive erythropoiesis and die from anemia between E11.5-E12.5[90]. These studies showed that a very low amount of GATA-1 in erythroid cells is sufficient to prevent apoptosis but is unable to promote erythroid differentiation[122].

Intriguingly, the overexpression of GATA-1 *in vitro* and *in vivo* also inhibits erythroid differentiation[123,124]. The block in terminal erythroid differentiation resulting from the overexpression of GATA-1 is different to that resulting from the GATA-1 knockout. Overexpressing erythroid cells differentiate past the proerythroblastic stage but fail to activate late differentiation markers, such as globins, and do not undergo cell cycle arrest at the G1 phase[123-125]. This leads to a failure in generating mature erythroid cells resulting in lethal anemia at E13.5 in mice[124]. The defect generated in erythroid cells by Gata1 overexpression is non-autonomous. It was shown that in the presence of wild type cells, Gata1 overexpressing erythroid cells are able to contribute to the erythroid differentiated pool of cells, as shown in heterozygous GATA-1 overexpressing females and chimeric mice[124]. This phenomenon involves a signaling molecule named Red Cell Differentiation Signal (REDS). REDS was shown to involve cells of the same type, characteristic of a homotypic mechanism, i.e. erythroid cells themselves communicate with each other[126]. REDS emerging from matured erythroid cells is thought to signal adjacent Gata1 overexpressing cells via a receptor-ligand interaction, leading to the degradation of proteins, including GATA-1 transcription factor. Thus, downregulation of GATA-1 allows the intrinsically GATA-1 overexpressing cells to differentiate properly. The degradation of GATA-1 through cleavage by caspases has been reported in primary erythroid cells[127]. Significantly, activation of caspases through death receptors blocked terminal erythroid differentiation at the basophilic erythroblast stage, concomitant with the down-regulation of GATA-1. These results

are not necessarily contradictory to the REDS model proposed for late stages of terminal differentiation, as GATA-1 degradation induced by caspases was studied in immature erythroid cells. This difference suggests that a high level of GATA-1 protein is required at earlier stages of erythroid maturation.

GATA-1 *in vivo* functions have also been addressed in other hematopoietic lineages. Generation of a specific megakaryocytic knockdown has underlined an essential role for GATA-1 in this lineage[128]. Mutant mice are characterized by an increase of the proliferation of the megakaryocytic progenitors and a defect in their maturation, thus leading to a dramatic reduction in the number of platelets[89].

GATA-1 is also expressed at high levels in eosinophils[80]. GATA-1 null cells fail to generate eosinophil progenitors in the fetal liver[129]. In addition, deletion of a positive regulatory element in the GATA-1 promoter blocks eosinophil development[130]. These two studies show that GATA-1 plays an essential role in eosinophilic differentiation. In mature cells, GATA-1 plays a role in the activation of eosinophilic genes, such as the granule major basic protein (MBP)[131].

Lastly, mast cell differentiation appears to be impaired in GATA-1^{low} mice. Even though mast cells were generated in these mice, an increase in the number of precursors was observed, in addition to an increase in apoptotic rates and defective maturation[132].

Taken together, these studies demonstrate an essential function of GATA-1 in several hematopoietic lineages. By contrast, the analysis of a testis-specific GATA-1 knockout has not demonstrated any apparent phenotype[133].

2.3.1.4 An instructive role for GATA-1 in Meg/E differentiation

In addition to gene targeting experiments that have documented an essential function for GATA-1 in several hematopoietic tissues, the enforced expression of GATA-1 in lineages where it is not normally expressed has provided evidence for an instructive role for GATA-1 in cell differentiation. An instructive role for GATA-1 can be defined as its ability to impose an expression program that promotes uni-lineage differentiation in a cell that had the potential to give rise to different cell types.

The first observations of GATA-1 being capable of driving a multipotent progenitor cell to express a unique transcription program was obtained using the early myeloid cell line 416B[43]. This cell line can differentiate into the megakaryocytic or granulocytic lineages. The ectopic expression of GATA-1 in these cells led to their differentiation specifically to megakaryocytes[43]. The expression of the unrelated SCL/TAL-1 transcription factor in the same cells did not result in differentiation. Along the same lines, ectopic GATA-1 expression in avian multipotential progenitors led to the transdifferentiation of myeloblasts into eosinophilic or megakaryocytic cells. The observed upregulation of megakaryocytic and the suppression of myelomonocytic markers suggests that the activation of specific transcription programs (e.g. megakaryocytic) occurs concomitantly with the suppression of alternative (e.g. myeloid) lineage programs[134]. Furthermore, the induced expression of GATA-1 in Granulocyte/Macrophage (GM) progenitors *in vitro* directs differentiation towards cells resembling erythroid and eosinophilic cells[7].

In addition to studies in the cell lines above, the instructive role of GATA-1 was also demonstrated in highly purified progenitor cells such as HSCs, CMP, CLP and Granulocytes/monocytes Progenitors (GMP)[47]. The enforced expression of GATA-1 “instructed” the differentiation of all these progenitors to the megakaryocytic and erythroid (Meg/E) lineages. Even progenitors committed to the lymphoid lineage (CLP) were reprogrammed into Meg/E fates by the ectopic expression of GATA-1, with the concomitant repression of

lymphoid-specific marker genes[7,47].

Together, all these observations suggest that GATA-1 is a key regulator in cell fate execution, with its functions involving both gene activation and repression. In Chapters 3 and 4 we provide molecular evidence of how GATA-1 can execute both repressive and activating functions.

2.3.1.5 GATA-1 protein interactions

Regulation of gene expression does not involve a single transcription factor at once, but rather a combination of transcription factors and co-factors, forming networks that change during cell specification. Accordingly, GATA-1 has been reported to undergo several interactions with other transcription factors and co-factors[57,135] (Figure 5). GATA-1 protein interactions are described below.

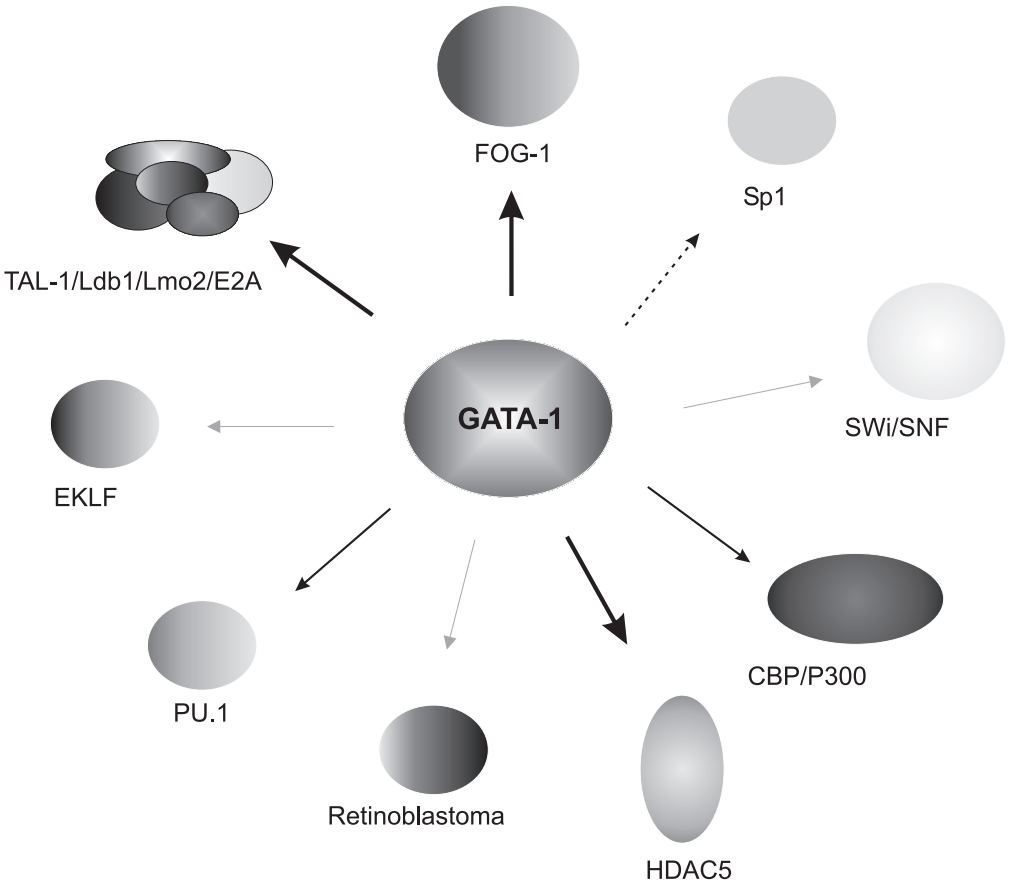


Figure 5: Protein partners of GATA-1
GATA-1 protein has been reported to bind to a number of transcription factors or co-factors. Evidence for interactions are based on different methods. Black arrows: immunoprecipitation; Grey arrow: GST pull-down; Dashed black arrow: EMSA.

FOG-1

The most prominent of the GATA-1 interacting proteins was identified from a MEL cell library using the yeast two-hybrid system. One of the isolated cDNA clones encodes a novel Zinc-finger protein of 998 amino-acids length, designated Friend of GATA (FOG-1)[51]. FOG-1 contains nine zinc-fingers, four of which can individually bind the N-finger of GATA-1[136]. Expression analysis of FOG-1 has revealed a tissue specific expression pattern similar to that of GATA-1. In hematopoiesis, FOG-1 is expressed strongly in erythroid and megakaryocytic cells[51]. However, FOG-1 expression could not be detected in eosinophils or in mast cells, both of which express GATA-1. FOG-1 is an essential hematopoietic factor, as the FOG-1 gene knockout is embryonic lethal at approximately E10.5–11.5 of gestation due to severe anemia. FOG-1 null embryos have a defect in primitive and definitive erythropoiesis with a block of cell maturation at the proerythroblastic stage, similar to that observed in the GATA-1 null mice[51]. FOG-1 mutants fail to express any globin genes and lack completely the megakaryocytic lineage. Analysis of megakaryocytic specific markers demonstrates an early block in megakaryocytic development[137]. Thus, FOG-1 plays an essential role in the megakaryocytic/erythroid lineages.

The phenotypic similarities observed in GATA-1 and FOG-1 null mice suggest that these two factors act in common pathways in the two hematopoietic cell lineages[137]. However, a more detailed analysis of the GATA-1 and FOG-1 knockout phenotypes revealed some differences. Whereas FOG null mice exhibit a complete failure of megakaryopoiesis, GATA-1 null mice have a block in megakaryocytic differentiation at mid-maturation, suggesting a broader, critical role for FOG-1 in the early stages of megakaryopoiesis. In the erythroid lineage, ablation of GATA-1 or FOG-1 leads to a differentiation block at the proerythroblast stage. Erythroid precursors survive longer in FOG-1 null mice than in GATA-1 null mice, suggesting that some GATA-1 functions are FOG-1-independent.

Early clues for the requirement of a co-factor in GATA-1 function emerged from the indispensability of the N-terminal zinc finger for GATA-1 functions[92]. Co-expression of FOG-1 and GATA-1 in G1E cells increases the expression of globin genes[51]. Similarly, co-transfection of FOG-1 and GATA-1 in progenitor 416B cells favored the generation of megakaryocytic cells, compared to cells transfected only with GATA-1. These data suggest that FOG-1 enhances the ability of GATA-1 to induce erythroid terminal differentiation and megakaryocytic differentiation, respectively. Furthermore, to address to what extent the requirement of GATA-1 in erythroid development is dependent on its association with FOG-1, a GATA-1 mutant was created. This GATA-1 mutant bears a single V205M amino acid change in the N-terminal zinc finger which abolishes the specific interaction with FOG-1, while not affecting DNA binding. When expressed in the GATA-1 null proerythroblastic G1E cell line, the V205M GATA-1 mutant failed to rescue terminal differentiation, in contrast to expression of the wild-type GATA-1 protein[67,138]. These results show that FOG-1 is an essential co-factor of GATA-1 in the terminal differentiation of erythroid cells. GATA-1/FOG-1 interactions have been linked to activation of genes such as globin, DC11 and HD2. Recently, ChIP data have demonstrated that FOG-1 facilitates binding of GATA-1 to β-globin genes, but also to the activated erythroid specific EKLF gene[69]. Similar observations have been made in the α-globin locus[68]. Importantly, the FOG-1-dependent function of GATA-1 has also been linked to gene repression. Expression of the GATA-1V205M in G1E cells leads to a lack of the down-regulation of GATA-2 and c-myc genes expression normally observed in terminal erythroid differentiation[67,138]. Whereas FOG-1 is dispensable for GATA-1 occupancy of the GATA-2 gene locus, FOG-1 was shown to be required for GATA-2 repression[70,139].

The GATA-1/FOG-1 interaction has also been linked to hematopoietic disorders *in vivo*. GATA-1 deficient mice rescued with a GATA-1 mutant unable to bind FOG-1 (V205G) exhibit X-linked thrombocytopenia at the adult stage, suggesting an essential role of the interaction of GATA-1/FOG-1 in late differentiation of megakaryocytes[140].

Importantly, the enforced expression of FOG-1 in hematopoietic cells has highlighted some additional functions in the regulation of cell fate. Constitutive expression of FOG-1 in MLPs leads to a block of C/EBPβ-mediated eosinophilic differentiation[141]. Expression of FOG-1 in an avian eosinophilic cell line, where it is not normally expressed, resulted in the dedifferentiation of the cells to a multipotential state[141]. This reprogramming function of FOG-1 is characterized by the suppression of eosinophilic markers such as EOS47 (which was shown to be activated by GATA-1 in eosinophils), suggesting that FOG-1 is a repressor of the eosinophilic lineage. Furthermore, deletion of the N-terminal zinc finger of GATA-1 abrogates the FOG-1-mediated repressing function. Recruitment of FOG-1 by GATA-1 to the target gene is then essential. Taken together, these data suggest that GATA-1 acts in complex with FOG-1 to mediate gene repression of alternative lineage programs. In Chapters 3 and 4, we show that GATA-1 forms distinct activating and repressive complexes in erythroid cells. We provide a mechanism for the GATA-2 repression mediated by GATA-1 in a FOG-1 dependent manner. In addition, we show that GATA-1 and FOG-1 bind to a repressed eosinophilic gene in erythroid cells.

SCL/TAL-1

Many of the genes that encode critical hematopoietic transcription factors are also implicated in leukemias, supporting the hypothesis that the appropriate regulation of transcription factor networks is essential for maintaining proper tissue homeostasis within the hematopoietic compartment. Chromosomal rearrangements in leukemia are of particular interest. The gene encoding SCL/TAL-1 was first identified as a chromosomal translocation in T-cell acute lymphoblastic leukemia (T-ALL)[142-145]. The translocation fused the SCL/TAL-1 gene with the T-cell receptor δ-locus, thus causing the misregulated expression of an aberrant TAL-1 fusion protein in the T-cell lineage, where TAL-1 is not normally expressed.

The TAL-1 gene encodes for a transcription factor belonging to the basic helix-loop-helix (bHLH) family. The activity of TAL-1 relies on two important domains of the protein. First, the HLH domain promotes homo- or hetero-dimerization of the protein with the ubiquitously expressed bHLH E2A proteins (E47, E12 and HEB). Second, the basic domain mediates DNA binding to the CANNTG consensus sequence, referred to as the E-box.

During embryogenesis, TAL-1 is expressed in hemogenic sites (YS and AGM), as well as in neural tissues and the vascular system[146]. The gene knockout of TAL-1 has shown that it is essential for the development of all hematopoietic lineages[54]. TAL-1 mutant ES cells in chimeric mice do not contribute to the development of any hematopoietic lineage. Importantly, expression of TAL-1 cDNA in TAL-1$^{-/-}$ cells rescues the hematopoietic defects demonstrating that TAL-1 is essential in ES cells for hematopoietic development. Thus, these studies implicate TAL-1 as one of the earliest expressed transcription factors required for hematopoiesis.

Within the hematopoietic system, the expression pattern of TAL-1 is very similar to that of GATA-1, namely, it is expressed in erythroid, megakaryocytic and mast cells[147]. It is also highly expressed in HSCs. Shivadasani and colleagues have shown by targeted gene disruption that TAL-1 is necessary for the production of embryonic blood cells. Tal-1 null embryos die at E9-10.5 due to anemia[148]. A functional inter-connection between GATA-1 and TAL-1 in erythropoiesis is supported by the fact that GATA binding sites in the TAL-1 promoter

are required for its transcription activation[149]. More recently, the role of TAL-1 specifically in the erythroid and megakaryocytic lineages was addressed using a TAL-1 conditional knockout. The loss of TAL-1 in these animals resulted in defects in both lineages, with a loss of progenitors, leading to anemia and thrombocytopenia. However, myeloid progenitors remained unaffected[150]. These observations suggest a crucial role for TAL-1 in the erythroid and megakaryocytic lineages[150,151], and parallel those of the GATA-1 knockout which also results in defects in the Meg/E lineages[66,119,128].

TAL-1 has been shown to be part of a large, so-called pentameric, protein complex that includes E47, Lmo2, Ldb1 and GATA-1. *In vitro* this complex acts as a transcriptional activator[52]. Lmo2 contains a LIM-domain which is involved in protein-protein interactions by acting as the bridging molecule between DNA binding proteins. Lmo2 null mice are characterized by a complete absence of hematopoietic cells, a phenotype which is very similar to that of the TAL-1 knockout[55,152]. The TAL-1 partner LIM-domain binding protein (Ldb1) is a ubiquitously expressed protein which is extremely conserved between human and mouse[153]. Loss of function of Ldb1 results in a severe developmental patterning defect[154] (A. Hostert, unpublished data). Interestingly, the hematopoietic system was also impaired, characterized by an absence of yolk sac blood islands in Ldb1 null embryos. This observation supports a role of Ldb1 in erythropoiesis, probably through interaction with proteins such as GATA-1 and TAL-1. Interestingly, Lmo2 was shown by immunoprecipitations to interact directly with GATA-1 in MEL cells, and to serve as a bridging factor between GATA-1 and TAL-1[63]. Thus, Lmo2 could serve as a bridge between GATA-1 and TAL-1/Ldb1 in the pentameric complex[52].

The function of the pentameric complex has been linked to the regulation of a number of target genes The complex binds to DNA through GATA/E-box motifs, with both DNA binding motifs spaced within 9-11bp to each other. The GATA/E-box motif was identified in a number of genes by ChIP assays using TAL-1 specific antibodies in MEL cell chromatin[155], and in the DNase Hypersensitive Site 1 (HS I) of the GATA-1 gene[89]. The TAL-1 complex in association with SP1 binds to the promoter of the c-kit gene, which codes for a kinase essential for hematopoietic development[156]. In addition, the erythroid specific EKLF gene contains upstream of the promoter a GATA-E-box and a GATA motif. In transcriptional assays, these motifs were shown to be sufficient to drive a level of expression comparable to the minimal promoter[157]. Recently, TAL-1 was shown to form a complex including GATA-1, E2A, Lmo2 that binds and activates the promoter of the GlycophorinA (GPA) gene, coding for one of the most abundant erythroid surface proteins[158]. Transcriptional activation of the GPA gene was shown to be dependent on the GATA-E-box and Sp1 motifs.

Lastly, TAL-1 has also been shown to interact in erythroid cells with both activating and repressing co-factors. Using immunoprecipitation experiments, P300 was found to interact with TAL-1 in differentiated MEL cells, and co-transfection assays showed that this interaction led to gene activation[159]. In MEL cells, TAL-1 also co-immunoprecipitated with the Sin3A transcriptional co-repressor and with HDAC1[160]. These observations suggest that TAL-1 can be involved in gene repression as well as gene activation.

EKLF

EKLF was identified as an erythroid specific transcription factor[161]. It belongs to the SP/XKLF (specificity protein/krüppel-like factor) family based on its three conserved Cys2His2 type zinc-fingers. This domain, located at the C-terminal end of the protein, mediates DNA binding to GC-rich sequences, such as CACC boxes found in many erythroid genes[161,162]. The

zinc-fingers overlap with a nuclear localization domain. In the N-terminal of the protein, two proline-rich domains have been defined as transactivation domains since *in vitro* they were shown to mediate transcriptional activation[163,164].

EKLF expression is restricted to the erythroid lineage only[165]. Hence, it is not surprising to find a GATA functional binding site in the promoter of EKLF, suggesting that EKLF is downstream of GATA-1 in erythropoiesis[166]. The regulatory sequences located 1kb 5' to the start of transcription are sufficient to drive erythroid-specific expression in mice[157]. Analysis of this sequence revealed a GATA/E-box/GATA motif that was shown to be essential for EKLF expression *in vitro*. In addition, these sequences are essential in transgenic mice[167].

As shown in EKLF gene knockout studies, the complete lack of erythropoiesis results in embryonic lethality at around day 14. Heterozygous knockout mice do not present with any defects[168,169]. EKLF$^{-/-}$ embryos have normal primitive yolk sac hematopoiesis up to E12, and die from anemia in early fetal life following shift of after the switch of hematopoiesis to the fetal liver[168]. The EKLF phenotype can be related to its role in regulating globin gene expression. The β-globin gene contains a CACC box in the promoter to which EKLF can bind with a very high affinity[169-172]. When crossed with mice containing a human β-globin locus, EKLF$^{-/-}$ null embryos express no β-globin, however they do exhibit elevated expression of γ-globin[173]. In addition, DNase hypersensitivity of the β-globin Locus Control Region (LCR) in EKLF mutants showed that HS3 became less sensitive to DNase digestion, suggesting that loss of EKLF affects chromatin structure at the LCR[173]. More direct evidence in a recent study further reinforced the important role of EKLF in β-globin locus regulation by showing it is required for the formation of the three-dimensional Active Chromatin Hub (ACH) and for the expression of the β-globin gene[174]. In addition, recent microarray analysis revealed that the Alpha Hemoglobin Stabilizing Protein (AHSP) gene, a protein that stabilizes free α-globin chains, was downregulated in EKLF knockout erythroid cells[175]. Importantly, a number of other target genes were found to be commonly affected in both the GATA-1 and EKLF knockouts[175,176].

EKLF interacts with a number of other proteins in exerting its functions. EKLF binds through its DNA binding domain to the mSin3a and HDAC1 repressors, thus preventing DNA binding[177]. Whereas this interaction suggests a role for EKLF in gene repression, such a function has not yet been described *in vivo*. EKLF has also been shown to bind proteins involved in gene activation, such as CBP/P300, which acetylates EKLF on conserved lysine residues thus enhancing the activation of the β-globin promoter[178]. It was also shown to require a chromatin remodeling complex related to SWI/SNF to induce the active conformation of the β-globin gene[179]. In order to gain further insight into the functions of EKLF *in vivo*, a strategy to tag the protein is presented in Chapter 2.

PU.1

PU.1 is a key transcription factor in hematopoiesis[180]. It is required for myeloid lineage differentiation[58,181]. PU.1 overexpression in MEL cells blocks erythroid differentiation and this block can be relieved by overexpression of GATA-1[45,182]. The Ets DNA binding domain of PU.1 has been shown to physically interact with the C-terminal domain of GATA-1 thus inhibiting DNA binding and GATA-1-mediated transcriptional activation in erythroid cells[45,182]. Conversely, GATA-1 interacts with PU.1 in myeloblasts leading to PU.1-dependent transcription repression, without affecting the PU.1 gene expression[46]. Whereas the GATA-1 interaction with PU.1 prevents recruitment of c-jun by PU.1. for transcriptional activation, the molecular mechanism by which the PU.1 interaction with GATA-1 represses GATA-1 activity in erythroid cells remains

to be elucidated[183,184]. These observations suggest that the two hematopoietic transcription factors GATA-1 and PU.1 are functionally antagonistic in cell lineage decisions between the myeloid and erythroid lineages. In agreement with this model, recent knockdown experiments in zebrafish have shown that expression levels of PU.1 versus GATA-1 represent an essential aspect of the balance of myeloid versus erythroid commitment[185].

CBP/P300

The two highly related, ubiquitously expressed histone acetyltransferases (HAT), CBP and P300, have been reported to interact with a number of transcription factors (Reviewed by Goodman and Smolik[186]). Immunoprecipitation experiments in MEL cells have shown that CBP can bind to GATA-1 and acetylation of GATA-1 correlates with the enhancement of GATA-1 transcriptional activity in murine fibroblasts[101]. Since histone acetylation is essentially linked to gene activation, one model suggests that GATA-1 tethers such chromatin modifiers to target genes to facilitate transcription, as was described for example for the β-globin locus[69]. Another model would suggest that transcriptional activation may be mediated by acetylated GATA-1, since GATA-1 acetylation by P300 was linked to an increase of its transcriptional activity[104]. However, the role of GATA-1 acetylation in enhancing transcriptional activation remains a somewhat controversial issue as two reports previously published on GATA-1 acetylation do not agree on this aspect[104,105].

Rb

Retinoblastoma is a tumor suppressor with a critical function in cell cycle progression from G1 to S phase[187]. The Rb null mutation leads to embryonic lethality characterized by neuronal defects and anemia due to erythropoietic impairment[188,189]. Importantly, erythroid cells in GATA-1 overexpressing mice fail to arrest at the G1 phase of the cell cycle. This defect is characterized by a decrease of the active hypophosphorylated form of Rb that interacts with the transcription factor E2F to stop cell cycle progression[123]. In addition, GATA-1 has been shown to interact directly with Rb[123]. These observations suggest a common pathway between Rb and GATA-1 in erythropoiesis.

2.3.1.6 GATA-1 Target genes

GATA sequences occur frequently in the genome and virtually all erythroid genes contain GATA binding sites in their regulatory elements[57]. The number of GATA-1 target genes is ever increasing since GATA-1 is also involved in regulating gene expression in lineages other than the erythroid lineages (i.e eosinophils, megakaryocytes). GATA-1 has been reported to regulate genes involved in many pathways, such as anti-apoptotic regulation, hemoglobin synthesis, cell signaling and also the cell cycle[176]. Direct binding of GATA-1 to target genes has been addressed primarily by Electro Mobility Shift Assays (EMSA) and by Chromatin Immuno Precipitation assays (ChIP). The latter consists of using an antibody to precipitate the protein of interest from formaldehyde crosslinked chromatin and to assay the specific *in vivo* bound DNA sequences by Polymerase Chain Reaction (PCR).

Globin genes

The β-globin gene was the first GATA-1 target gene to be identified. GATA-1 was found to bind to the β-globin gene enhancer[83,84]. It can also bind the α-globin gene[78]. Recent studies of the α-globin gene cluster have shown that in committed erythroid cells, GATA-1 binds

to the promoter and to an upstream HS[68]. However, GATA-1 null mice can still produce some hemoglobin suggesting that GATA-1 might not be absolutely essential for globin expression. In addition, the GATA-1 null cells express GATA-2 at an abnormally high level, suggesting that GATA-2 can partially replace GATA-1 in activating globin expression[66].

ChIP-on-Chip analysis has identified GATA *in vivo* binding sites across the β-globin LCR[190]. Recently, using the 3C (Chromosomal Conformation Capture) approach in erythroid cells, GATA-1 and its FOG-1 cofactor were shown to mediate spatial interactions between β-globin gene and the LCR[191]. 3C technology enables the *in vivo* mapping of interactions between chromosomal regions independent of their transcriptional status[192].

Cell cycle

Cell cycle-related genes are affected in GATA-1 overexpressing erythroid cells. For example, the decrease in cyclin E expression normally observed during erythroid differentiation was not observed upon GATA-1 overexpression[123]. These observation suggested a role for GATA-1 in the regulation of the cell cycle, and more precisely in the G1 to S transition. Microarray data, obtained from the GATA-1 null G1E proerythroblastic cell line expressing an inducible form of GATA-1, showed that a number of cell cycle genes, such as $p27^{kip1}$, were activated, while some others, such as Cyclin D2, were repressed[193]. However, this study did not systematically investigate whether these genes were direct or indirect targets of GATA-1. Nevertheless, in the same study it was shown that GATA-1 could bind directly to the repressed promoter of c-myc, a known oncogene involved in cell proliferation[193]. These observations suggest that in differentiating erythroid cells, c-myc repression, and hence repression of cell proliferation, is GATA-1 dependent. As we mentioned, GATA-1 plays a role in cell cycle as it was shown to interact with the Rb protein[123]. Chapter 4 describes that GATA-1 binds to the c-myc promoter in association with the essential hematopoietic transcription factor Gfi-1b, which acts as a repressor.

Signaling pathways

Cell surface receptors are essential molecules that are part of the signaling pathways that respond to a stimuli at the cell/environment interface and trigger intracellular responses. During erythroid differentiation the activation of glycophorin A (GPA, a membrane protein associated with the erythroid marker Ter119) and the erythopoietin receptor (EpoR, receptor to the Epo growth factor that promotes differentiation of erythroid cells) were shown to be GATA-1-dependent[158,81]. EpoR is present at the surface of erythroid, megakaryocytes and mast cells. Binding of Epo to its receptor triggers signaling cascade, which leads to the proliferation, differentiation and survival of erythroid progenitors[194]. However, in GATA-1 null proerythroblasts EpoR transcription appeared unaffected, suggesting a potential redundancy with GATA-2[66].

Transcription factors

Transcription factors constitute an important group of GATA-1 target genes. Thus GATA-1 play a pivotal role in broad transcriptional regulation. As described earlier in the introduction, GATA-2 is an essential factor in early hematopoiesis. Erythroid cells lacking GATA-1 exhibit higher GATA-2 expression levels[66]. GATA-2 gene analysis has revealed multiple GATA binding sites in regulatory elements. Using ChIP assays, Bresnick and colleagues investigated GATA-1 binding to the GATA-2 locus[139,195]. In progenitor cells, the active GATA-2 locus is occupied by GATA-2 and CBP. Upon erythroid commitment, differentiation and

increased GATA-1 expression, GATA-1 replaces GATA-2 binding in the GATA-2 locus, CBP is displaced and a wave of deacetylation occurs concomitant with the GATA-2 repression. Using the G1E GATA-1 null cell line expressing an inducible form of GATA-1, it was clearly shown that GATA-2 repression required the expression and binding of GATA-1 to the locus. Data presented in Chapters 3 and 4 of this thesis provide further molecular insight as to how GATA-1 and FOG-1 repress the GATA-2 locus.

TAL-1 also plays an essential role during hematopoiesis where it is expressed early in the hematopoietic stem cells and later in the erythroid, megakaryocytic lineages and in mast cells[147]. Detailed studies of the TAL-1 promoter revealed the presence of GATA sites. GATA-1 was shown to bind directly to the TAL-1 promoter, in cooperation with the ubiquitously expressed SP1 transcription factor[149]. *In vitro* data suggested that GATA-1 promoted activation of the erythroid-specific expression of the TAL-1 gene.

EKLF, an erythroid specific transcription factor, has also been reported to be a direct target of GATA-1. The EKLF promoter region contains two GATA binding sites, and importantly, one was shown to be crucial for its transcriptional activation[107]. The motif GATA/E-box/GATA present in an upstream enhancer element was shown to be essential for *in vivo* erythroid expression of an EKLF transgene[167]. More recently, and consistent with previous data, the expression of GATA-1 in the GATA-1 null G1E cells was correlated with the upregulation of EKLF gene transcription[193]. In Chapter 4, we show GATA-1 binding in association with the TAL-1/Ldb1 complex to the upstream enhancer of the EKLF gene in differentiated erythroid cells[157].

In human eosinophilic cells, GATA-1 protein was also shown to regulate the expression of an eosinophil/granulocyte gene, the major basic protein (MBP)[131]. In the mouse hematopoietic system, Mbp gene is highly expressed in eosinophils and less so in basophils[196,197]. The MBP protein is one of the principal mediators in the inflammatory response. A more detailed analysis of the MBP gene has revealed functional binding sites for GATA-1 in the promoter and GATA-1 binding to the MBP promoter in eosinophils was associated with modulation of the gene's expression[198].

2.4 GATA-1 and Diseases

Down's syndrome patients present a high incidence of hematopoietic malignancies such as transient myeloproliferative disorder (TMD) or acute megakaryoblastic leukemia (AMKL). AMKL patients express a truncated GATA-1 protein in which the N-terminal 83 amino acid activation domain has been deleted because of a premature stop codon; the full-length GATA-1 protein is absent[199-201]. These observations suggest that the loss of full length GATA-1 is an essential step in leukemogenesis in these patients. But the mechanism by which the mutations contribute to leukemia remains unclear. Indeed, even in normal hematopoietic cells, a second isoform of GATA-1 is expressed from an alternative translational initiation site located at methionine 84, thus omitting the N-terminal activation domain. Despite a lower level of expression, the shorter isoform can be detected in erythroid cell lines such as the murine MEL and the human K562 erythroleukemic cells as well as in the mouse tissues. The shorter protein shows normal DNA binding capacity, as it contains the two characteristic zinc-fingers[202].

In addition, missense mutations have been identified in families presenting with anemia and thrombocytopenia, though these disorders remain rare. Strikingly, all mutations associated with these disorders reside in the N-terminal zinc finger of GATA-1. The first mutation that was found in patients suffering from severe anemia and thrombocytopenia, was a substitution of a

valine to a methionine at position 205 (V205M). This mutation leads to a dramatic decrease of GATA-1 binding to FOG-1, as was also shown in vitro[138,201]. Other mutations in the GATA-1 protein such as R216Q, were shown to be present in patients suffering from macrothrombocytopenia with β-thalassemia. This mutation in the N-terminal zinc finger does not affect the interaction with FOG-1, but affects the binding to palindromic GATA sequences[130].

3. Chromatin and Transcription

The length of the DNA contained in the nucleus when fully stretched is about 2 meters. To fit into the nucleus, DNA is compacted into chromatin. Despite the enormous degree of compaction that the DNA undergoes, the DNA in chromatin must be rapidly accessible for interactions with protein complex that regulate many DNA functions (replication, transcription, repair and recombination). The dynamic organization of DNA in chromatin is therefore an essential aspect of gene regulation.

3.1 Definition-Structure

To form the so-called chromatin, DNA is in complex with proteins, mostly histones. The fundamental unit of chromatin is the nucleosome. It is composed of a histone (H) octamer around which 146 bp of DNA are wrapped. The octamer consists of two copies of each of the histones H2A, H2B, H3 and H4. Structurally, histone proteins contain two domains: a globular domain involved in histone-histone and histone-DNA interactions, and a N-terminal tail of variable length that is the site of many posttranslational modifications, involved in activation, silencing, chromatin assembly and DNA repair[203]. Nucleosome arrays assembled on DNA are separated by short (10-80bp) segments of DNA (the linker DNA) to form a beads-on-a-string-like structure, which can fold into a helical compacted structure, the so-called 30 nm fibre. This structure is found in both interphase chromatin and mitotic chromosomes. Histone H1, a fifth class of histones, binds to the nucleosomes and to linker DNA thus compacting the nucleosomal array into forming higher order chromatin structures such as the 30nm fiber and beyond. The overall compaction ratio estimated in mitotic chromosomes is about 10,000[204]. However the precise structure of higher order chromatin in chromosomes *in vivo,* beyond the 30nm fiber, remains unclear.

In the nucleus, two forms of chromatin can be distinguished: euchromatin and heterochromatin. Heterochromatin represents around 10% of the genome and remains condensed throughout the cell cycle. It is primarily associated with the gene-poor telomeric and pericentromeric regions and is thus considered to be a transcriptionally inactive form of chromatin. Further studies have described two subtypes of heterochromatin. Constitutive heterochromatin contains very few genes and is composed of a high content of repetitive sequences located in large regions such as centromeres and telomeres. Facultative heterochromatin corresponds to regions that condense into heterochromatin-like structures as a result of developmental changes. One of the best-studied examples of facultative heterochromatin is the X chromosome of mammals, that is subject to compaction in the X-inactivation process.

In contrast, euchromatin decondenses after metaphase, is nuclease sensitive and contains a high density of genes. In addition, the distinctions between heterochromatin and euchromatin are also reflected in the types and combinations of posttranslational modifications in histone tails (see below).

3.2 Chromatin and Gene Expression

Nucleosomes associated with active genes were shown to be more accessible to nucleases than those associated with inactive genes[205]. Clues on the dynamic aspect of chromatin came from the isolation of proteins that are able to modify chromatin. In studying modification of chromatin structure, the best-studied level is that of the nucleosome. Generally, there are three ways in which chromatin structure can be modulated (i) by nucleosome remodeling catalyzed by specific large protein complexes containing specific enzymes that hydrolyze ATP, (ii) by covalent posttranslational modification of histone tails, (iii) by histone variants with specialized properties.

3.2.1 Histone modifications

The N-terminal tails are not required for the structural integrity of the nucleosome[206]. However, they are essential for the regulation of chromatin structure *in vivo* as they are the sites of a very complex combination of post-translational modifications. The size of the tails varies from 16 to 44 amino acids, depending on the histone. In addition to the modifications presented below, we can mention that other modifications such as ubiquitination, sumoylation and PARylation can also occur[203].

Histone acetylation/deacetylation

One of the best-described histone modifications is acetylation. Acetylation disrupts higher order chromatin folding and facilitates recruitment of transcription factors[207]. For these reasons, histone tail acetylation is generally correlated with transcription[208]. Histone tail acetylation occurs on specific lysine residues and is reversible, thus resulting in a dynamic state of acetylation and deacetylation. The acetylation status of histone tails is catalyzed by the enzymes Histone Acetyl Transferases (HAT) and Histone Deacetylases (HDAC).

CBP (CREB Binding Protein) and p300, two closely related mammalian proteins, were identified as co-activators of transcription[186,209] and possess HAT activity. These co-factors are thought to be tethered by specific transcription factors to gene targets, for example, by their interaction with transcription factors, such as GATA-1 or EKLF[101].

Interestingly, yeast Gcn5 and its human homologue p300/CBP-associated factor (PCAF) are transcription factors that carry their own intrinsic HAT activity. Human TAF250, which is part of the basal transcriptional machinery, has also been shown to possess intrinsic HAT activity[210]. Finally, transcriptional effectors such as PCAF bind preferentially to specific acetylated lysines on histone tails through the bromodomain[211].

HDACs work antagonistically to HATs in that they are associated with gene repression[212]. Based on their homology to yeast HDACs RPD3 and HDA1, mammalian HDACs are grouped respectively into class I and II. Activity of class I and II HDACs can be inhibited using the deacetylase inhibitor trichostatin A (TSA). HDAC1 and HDAC2 belong to the HDAC class I and share 84 % homology. Most HDACs do not contain DNA-binding activities, hence transcription factors recruit them to specific target genes. For example, the repression of the cyclinE expression resulting from Rb protein binding to E2F was shown to involve recruitment of HDAC1 protein binding to the promoter of cyclinE gene[213].

Importantly, HDAC inhibitors are important molecules in cancer research, as they are effectors of many pathways, such as growth arrest and apoptosis of transformed cells. They are already being used as therapeutic agents for the treatment of certain forms of cancer[214].

Histone methylation

Methylation of histones was first described in 1964[215], however, the histone-methyltransferases (HMT) responsible for catalyzing the histone methylation reaction were only recently discovered[216]. Since then, other HMTs have been described and classified based on the specificity of lysine residue methylation. Unlike histone acetylation, methylation can occur on lysine and arginine residues, catalyzed respectively by the SET domain-containing family of HMTs and the protein arginine methyltransferases (PRMTs). Methylation of histones plays an important role in chromatin regulation and is linked to both transcriptional activation and repression[217]. Methylation of histone H3 on lysine residue 9 (H3-K9) is associated with heterochromatin formation[218] and gene repression. In heterochromatic regions, the Suv39H HMT methylates H3-K9, which is then recognized by the heterochromatin-associated protein HP1[219]. In contrast, methylation of H3-K4 is linked to gene activation[220].

While regulation of histone acetylation is highly dynamic, histone methylation has been considered as very stable mark of the chromatin[221]. It is only very recently that an enzyme, referred to as LSD1 (Lysine Specific Demethylase1), has been identified that can demethylate histones. This enzyme has been previously shown to be part of repressor complexes, such as CtBP or Co-REST[222,223].

Histone phosphorylation

Phosphorylation of serines is not as well characterized as other histone modifications; nevertheless the function of phosphorylation is linked to chromatin regulation during the cell cycle. Specific phosphorylation of H3-S10 is required for chromosomal condensation, whereas the level of the phosphorylated form rapidly decreases after mitosis. Moreover, the Snf1 kinase that phosphorylates specifically H3-S10 cooperates with the activating co-factor GCN5 HAT, suggesting that phosphorylation of H3-S10 is associated with gene activation[224]. In addition, the level of phosphorylated H1 increases during the cell cycle from the G2 to the S phase[225].

The “histone code”

The number of reported histone modifications and their biological relevance is ever increasing. The brief introduction given in this thesis is highlighting some of the main concepts, but does not address them all exhaustively. Nonetheless, histones are modified in several ways leading to a concept of a complex interplay between all modifications, referred as the “histone code” [226]. Histone modifications on the one hand have a structural function that influences interactions with DNA, and on the other hand affect binding and recruitment of effectors molecules[227]. Bromo- and chromodomains have been demonstrated to bind with high specificity to specific acetylated and methylated lysine residues, respectively[228,229]. Furthermore, histone modifications were shown to prevent or facilitate binding of effectors. For example, acetylation of H3-K9/K14 abrogates affinity of the co-repressor complex INHAT[230]. In addition, the NuRD complex is displaced from histone H3 if the latter is methylated on K4[231]. The binding of HP1 to methylated H3 can recruit SUV39H, that can methylate H3 tail of neighboring histones, therefore spreading the modification and providing an example of how heterochromatin can spread[232]. In addition, specific histone modifications can modulate additional modifications in nearby residues. For example, methylation of H3-K9 for instance is inhibited if H3-S10 is phosphorylated[233], and by contrast, phosphorylation of H3-S10 facilitates the acetylation of H3-K14[234]. These observations suggest that histone modifications can influence each other.

3.2.2 Histone variants

In addition to the major histones contained in nucleosomes, some different histones have been described, the so-called histone "variants". They are associated with all histones except H4, and differ structurally from the main histones by changes in amino acid sequences. However, their functions are as yet undetermined but they are thought to be involved in epigenetics changes. They are assembled in alternative chromatin structures that can be transmitted to the next generation[235]. In chromatin, they replace corresponding major histones. The majority of nucleosome assembly involving canonical histones occurs during the replication phase. In contrast to the major class of histones, some histones variants, such as the H3.3 can be incorporated throughout the cell cycle[236]. Interestingly, H3.3 has also shown to play a role in transcriptionally active chromatin of the chromosome[236,237].

3.2.3 Chromatin remodeling complexes

Large ATP-dependent chromatin remodeling complexes increase accessibility of the nucleosomal DNA during regulation of transcription. Enzymes possessing ATPase activity constitute the "heart" of the remodeling complexes. Based on homology of sequences, three classes of ATP-dependent chromatin remodeling complexes are distinguished: the SWI2/SNF, the ISWI and Mi-2 families[238]. Recently, a fourth class of complexes, INO80/SWR, was described in *S. cerevisiae*[239] (Figure 6). In addition to the ATPase domain, these enzymes contain other functional domains such as bromodomains and PHD domains. Importantly, whereas SWI/SNF and ISWI complexes have a unique ATPase activity, the Mi-2 complex family has the particularity to possess ATPase and HDAC activities. These complexes are expressed ubiquitously.

SWI/SNF and ISWI

hSWI/SNF and ISWI, which are the two best-studied families of chromatin remodeling complexes, are conserved from yeast through to Drosophila and man. The first chromatin-remodeling complex to be discovered in yeast was a 2 MDa complex containing the SWI2/SNF2 subunit responsible for nucleosome disruption[240]. Similar complexes were also identified in Drosophila and in human. In mammals there are two SWI/SNF2 homologues, Brm and Brg1[241,242].

Whereas early evidence suggested that the SWI/SNF complex was recruited to promoters by RNA pol II, subsequent data suggested that this is not the case. Instead, SWI/WNF is tethered to DNA by gene-specific transcription factors, eventually recruiting RNA pol II for transcription[243,244]. In addition, recruitment of SWI/SNF by an activator on a gene target has been shown *in vivo* by ChIP assays[245]. In mammals, the SWI/SNF is recruited to the β-globin locus by the essential erythroid specific transcription factor EKLF[246], and to myeloid genes by C/EBPβ[247]. *In vitro* data have also indicated that GATA-zinc fingers are able to interact with SWI/SNF[248] but no *in vivo* function has been described. Targeted inactivation of the Brg1 gene in mice also resulted in an increase in expression of a number of genes, thus suggesting that the complex may also be linked to gene repression[249].

ISWI chromatin remodeling complexes contain SNF2h as the ATPase subunit. In human, SNF2h forms three distinct complexes: RSF, ACF/WCRF and CHRAC. The ACF/WCRF complex is composed of 2 subunits: SNF2h and ACF, while the CHRAC complex contains in addition two subunits P15 and P17.

Studies *in vitro* in Drosophila have linked ISWI to gene activation, but an increasing number of reports have also described a strong link between SNF2h and gene repression[250].

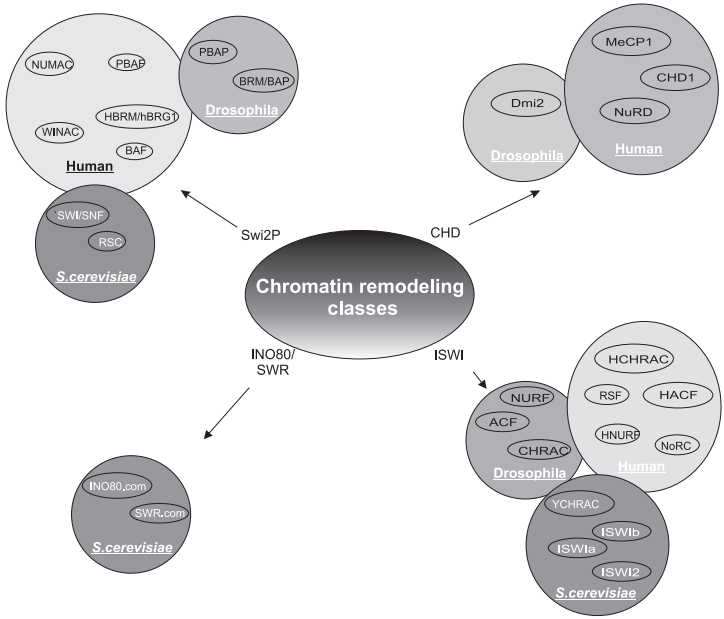


Figure 6: Chromatin remodeling complexes.
Chromatin remodeling complexes in different species are classified based on their ATPase domain: CHD, ISWI, INO80 and Swi2P. Adapted from reference 239.

NCoR, a SNF2h-containing chromatin remodeling complex, has also been shown to repress ribosomal gene transcription by recruiting histone deacetylases on rDNA promoters[251]. In addition, SNF2h was shown to localize with heterochromatin through its interaction with ACF1 during replication[252]. Gene targeting of Brg1 and SNF2h have revealed that they are both essential for embryonic development. Brg1$^{-/-}$ mice die very early in embryogenesis, before the preimplantation stage (E3.5-E6.5)[253]. The SNF2h null phenotype in mice is also embryonic lethal. However, SNF2h null embryos undergo implantation[254]. Importantly, SNF2h knockdown experiments in primary erythroid progenitors were shown to lead to a marked reduction of GlyA+ cells and a decrease of β-globin mRNA accumulation, suggesting an essential role in erythropoiesis[254]. In Chapter 3, we describe a specific interaction of GATA-1 with SNF2h in the context of the ACF/WCRF complex, suggesting a possible function in a common pathway in erythropoiesis.

At the molecular level, Brg1 functions on both naked and nucleosomal DNA, and SNF2h is stimulated preferentially by nucleosomal DNA[255,256]. All chromatin complexes are able to remodel nucleosomes according to the “sliding model”, in which the amount of DNA exposed and nucleosome spacing remain unchanged[238]. However, the remodeling function of SWI/SNF induces topological changes in the nucleosomal structure that lead to an increase in the amount of DNA being exposed, and is referred to as “conformational change”.

NuRD/MeCP1

Purification of the NuRD/Mi-2 complex from HeLa cells and *Xenopus* eggs led to the characterization of the only type of chromatin remodeling complex that contains both ATPase

and HDAC activities[257,258]. The purified MeCP1 protein complex is composed of 10 subunits, including all subunits of the NuRD complex and in addition the MBD2 subunit[259]. The latter was shown to bind methylated DNA[259,260]. This complex was shown to preferentially bind, remodel and deacetylate methylated nucleosomes over naked DNA[259].

In HeLa cells, characterization of the NuRD complex has revealed the presence of 7 components, including the ATPase/helicase Mi-2, HDAC1 and 2, Metastasis Associated Protein 2 (Mta2), Methyl Binding Protein 3 (MBD3), Retinoblastoma Associated Protein 46 and 48 (RbAp46/48). The core complex composed of HDAC1, HDAC2 and RbAp46/48 is shared between the NuRD complex and the chromatin-remodeling complex Sin3a. Within the NuRD complex, the MTA2 subunit plays an important role in modulating the activity of the HDACs[261]. The MBD3 subunit of the complex is a methyl-CpG-binding domain-containing protein[262]. Further studies have shown that the NuRD complex is itself part of a larger protein complex of 10 subunits, the so-called MeCP1 complex. It contains in addition to the NuRD complex, MBD2, P66 and P68[259,263,264]. MBD2 is a methyl DNA binding domain-containing protein that is related to MBD3. However, unlike MBD3, MBD2 can bind to methylated DNA *in vivo* and *in vitro* in mammalian cells.

The NuRD/MeCP1 complexes have been associated with gene repression during development[265-267]. As with other remodeling complexes, NuRD/MeCP1 are recruited to target genes through interactions with DNA binding transcription factors. For example, Ikaros, an essential transcription factor for B and T cell development[268], was shown to recruit NuRD complex to a heterochromatic region upon T-cell activation, suggesting that recruitment of NuRD is associated with alternative lineage suppression[269]. Interestingly, interaction of GATA-3 with NuRD in differentiating Th2 cells was shown to activate transcription of the IL-4 gene after displacing MBD2 from the chromatin[266]. Expression of Mi2 defective for the ATPase activity relieved transcriptional repression, suggesting that the Mi2 protein is essential in the repression activity of the complex[259].

In addition, the presence of the MBD2/3 subunits in the complex suggests that it might be recruited to methylated DNA, thus mediating transcriptional silencing[270]. Importantly, the function of MBD2 and MBD3 differs between organisms. Whereas mammalian MBD3, either by itself or interacting with NuRD, does not show any affinity for methylated DNA, recombinant MBD3 isolated from *Xenopus* binds methylated DNA[260,261]. Further, in mammalian cells, MBD2 can mediate the interaction between NuRD and methylated DNA *in vitro*[261].

Genetic studies of Mi2 have given some insight as to the function of the NuRD/MeCP1 complexes *in vivo*. The Mi-2β conditional knockout in thymocytes showed that Mi2 is required at different stages of T cell maturation[271]. Despite its interaction with the repressive NuRD complex, Mi-2β was unexpectedly shown to promote expression of CD4 by association with HAT[271]. Recently, the function of MBD3L2, a homologue of MBD2 and MBD3 lacking the methyl CpG binding domain, was shown to be interchangeable with MBD2 in the MeCP1 complex *in vitro*. MBD3L2 mediates gene repression, but could displace MeCP1 from methylated DNA[272]. These observations suggest that Mbd related proteins could regulate function of MeCP1.

In Chapter 3, we report that GATA-1 interacts in erythroid cells with MeCP1 through its interaction with FOG-1. This complex is stably bound to repressed genes during erythropoiesis. In addition, it was shown recently that FOG-1 mediates the interaction of NuRD with GATA-1 by directly binding to MTA proteins[273].

References

1. Lemischka, I., Raulet, D. & Mulligan, R. Developmental potential and dynamic behavior of hematopoietic stem cells. *Cell* **45**, 917-27 (1986).
2. Prohaska, S., Scherer, D., Weissman, I. & Kondo, M. Developmental plasticity of lymphoid progenitors. *Semin Immunol* **14**, 377-84 (2002).
3. Akashi, K. et al. Transcriptional accessibility for genes of multiple tissues and hematopoietic lineages is hierarchically controlled during early hematopoiesis. *Blood* **101**, 383-9 (2003).
4. Akashi, K., Traver, D., Miyamoto, T. & Weissman., I. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**, 193-7 (2000).
5. Nolta, J. & Jordan, C. Spotlight on Hematopoietic Stem Cells: Looking Beyond Dogma. *Leukemia* **15**, 1677-80 (2001).
6. Chirag, J. & Enver, T. Molecular complexities of stem cells. *Curr Opin Hematol.* **10**, 220-8 (2003).
7. Heyworth, C., Pearson, S., May, G. & Enver, T. Transcription factor-mediated lineage switching reveals plasticity in primary committed progenitor cells. *Embo J* **21**, 3770-81 (2002).
8. Graf, T. Differentiation plasticity of hematopoietic cells. *Blood* **99**, 3089-101 (2002).
9. Brazelton TR, R. F., Keshet GI, Blau HM. From marrow to brain: expression of neuronal phenotypes in adult mice. *Science* **290**, 1775-9 (2000).
10. Wurmser, A. & Gage, F. Cell fusion causes confusion. *Nature* **416**, 542-5 (2002).
11. Choi, K., Kennedy, M., Kazarov, A., Papadimitriou, J. & Keller, G. A common precursor for hematopoietic and endothelial cells. *Development* **125**, 725-32 (1998).
12. Wong, P., Chung, S.-H., Reicheld, S. & Chui, D. Hemoglobin switching during murine embryonic development: evidence for two populations of embryonic erythropoietic progenitor cells. *Blood*, 716-21 (1986).
13. Palis, J. & Yoder, M. Yolk sac hematopoiesis: the first blood cells of mouse and man. *Exp Hematol*, 926-37 (2001).
14. Lu, M., Kawamoto, H., Katsube, Y., Ikawa, T. & Katsura., Y. The Common Myelolymphoid Progenitor: A Key Intermediate Stage in Hemopoiesis Generating T and B Cells. *The Journal of immunology* **169**, 3519-25 (2002).
15. Godin, I., Garcia-Porrero, J., Coutinho, A., Dieterlen-Lievre, F. & ., M. M. Para-aortic splanchnopleura from early mouse embryos contains B1a cell progenitors. *Nature* **364**, 67-70 (1993).
16. Medvinsky, A., Muller, N. S. & ., E. D. An early pre-liver intraembryonic source of CFU-S in the developing mouse. *Nature* **364**, 64-7 (1993).
17. Medvinsky, A. & Dzierzack, E. Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell* **86**, 897-906 (1996).
18. Ema, H. & Nakauchi, H. Expansion of hematopoietic stem cells in the developing liver of a mouse embryo. *Blood* **95**, 2284-8 (2000).
19. Muller, A., Medvinsky, A., Strouboulis, J., Grosveld, F. & Dzierzak, E. Development of hematopoietic stem cell activity in the mouse embryo. *Immunity* **1**, 291-301 (1994).
20. Ottersbach, K. & Dzierzak, E. The murine placenta contains hematopoietic stem cells within the vascular labyrinth region. *Dev Cell* **8**, 377-87 (2005).
21. Borzillo, G., Ashmun, R. & Sherr, C. Macrophage lineage switching of murine early pre-B lymphoid cells expressing transduced fms genes. *Mol Cell Biol.* **10**, 2703-14 (1990).
22. Hu, M. et al. Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev* **11**, 774-85 (1997).
23. Socolovsky, M., Lodish, H. & Daley, C. Control of hematopoietic differentiation: lack of specificity in signaling by cytokine receptors. *Proc Natl Acad Sci U S A* **95**, 657-5 (1998).
24. Wu, H., Liu, X., Jaenisch, R. & Lodisch, H. Generation of committed erythroid BFU-E and CFU-E progenitors does not require erythropoietin or the erythropoietin receptor. *Cell* **83**, 59-67 (1995).
25. Crook, G., Hao, Q., Petersen, D., Barsky, L. & Bockstoce, D. IL-3 Increases Production of B Lymphoid Progenitors from Human CD34+CD38- Cells. *The Journal of immunology*, 2382-89 (2000).

26. Migliaccio, A., Bruno, M. & Migliaccio, G. Evidence for direct action of human biosynthetic (recombinant) GM-CSF on erythroid progenitors in serum-free culture. *Blood* **70**, 1867-71 (1987).
27. Migliaccio, G. et al. Stem Cell Factor Induces Proliferation and Differentiation of Highly Enriched Murine Hematopoietic Cells. *Proc Natl Acad Sci U S A* **88**, 7420-4.
28. Sudo, T. et al. Interleukin 7 production and function in stromal cell-dependent B cell development. *J Exp Med* **170**, 333-8 (1989).
29. Clutterbuck, E., Hirst, E. & CJSanderson. Human interleukin-5 (IL-5) regulates the production of eosinophils in human bone marrow cultures: comparison and interaction with IL-1, IL-3, IL-6, and GMCSF. *Blood* **73**, 1504-12 (1989).
30. Charbord, P. in *HEMATOLOGY* (ed. Zon, L.) (oxford, 2001).
31. Duncan, A. et al. Integration of Notch and Wnt signaling in hematopoietic stem cell maintenance. *Nat Immunology*, 314-22 (2005).
32. Liesveld, J. et al. Adhesive interactions of normal and leukemic human CD34+ myeloid progenitors: role of marrow stromal, fibroblast, and cytomatrix components. *Exp Hematol* **19**, 63-70 (1991).
33. Whyatt, D. and Grosveld, F. (2002) Cell-nonautonomous function of the retinoblastoma tumour suppressor protein: new interpretations of old phenotypes. EMBO rep, 3, 130-135.
34. Federico, M. et al. Increased expression of laminin receptors during myeloid differentiation. *Int J Cancer* **49**, 32-7 (1991).
35. Kashimura, M. Scanning electron microscopy studies of bone marrow. *Scan Electron Microsc*, 445-53 (1982).
36. Enver, T. & Greaves, M. Loops, lineage, and leukemia. *Cell* **94**, 9-12 (1998).
37. Cross, M. & T, E. The lineage commitment of haemopoietic progenitor cells. *Curr Opin Genet Dev.* **7**, 609-13 (1997).
38. Miyamoto, T. et al. Myeloid or lymphoid promiscuity as a critical step in hematopoietic lineage commitment. *Dev Cell* **3**, 137-47 (2002).
39. Jimenez, G., Griffith, S., Ford, A., Greaves, M. & T, E. Activation of the beta-globin locus control region precedes commitment to the erythroid lineage. *Proc Natl Acad Sci U S A* **89**, 10618-22 (1992).
40. Ford, A. et al. Immunoglobulin heavy-chain and CD3 delta-chain gene enhancers are DNase I-hypersensitive in hemopoietic progenitor cells. *Proc Natl Acad Sci U S A* **89**, 3424-8. (1992).
41. Klinken, S., Alexander, W. & Adams, J. Hemopoietic lineage switch: v-raf oncogene converts Emu-myc transgenic B cells into macrophages. *Cell* **53**, 857-67 (1988).
42. Kondo, M., Weissman, I. & Akashi, K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91**, 661-72 (1997).
43. Visvader, J., Elefanty, A., Strasser, A. & Adams, J. GATA-1 but not SCL induces megakaryocytic differentiation in an early myeloid line. *Embo J* **11**, 4557-64 (1992).
44. Nerlov, C., McNagny, K., Doderlein, G., Kowen-Leutz, E. & Graf, T. Distinct C/EBP functions are required for eosinophil lineage commitment and maturation. *Genes Dev* **12**, 2413-23 (1998).
45. Rekhtman, N., Radparvar, F., Evans, T. & Skoultchi, A. I. Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells. *Genes Dev* **13**, 1398-411 (1999).
46. Nerlov, C., Querfurth, E., Kulessa, H. & Graf, T. GATA-1 interacts with the myeloid PU.1 transcription factor and represses PU.1-dependent transcription. *Blood* **98**, 2543-51 (2000).
47. Iwasaki, H. et al. GATA-1 converts lymphoid and myelomonocytic progenitors into the megakaryocyte/erythrocyte lineages. *Immunity* **19**, 451-62 (2003).
48. Alberts, B. et al. *Molecular Biology of The Cell* (Garland Science, New York, 2002).
49. Buratowski, S. The basics of basal transcription by RNA polymerase II. *Cell* **77**, 1-3 (1994).
50. Roeder, R. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* **21**, 327-35 (1996).
51. Tsang, A. P. et al. FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell* **90**, 109-19 (1997).
52. Wadman, I. A. et al. The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *Embo J* **16**, 3145-57 (1997).

53. Tsai, T. et al. An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature* **15**, 221-6 (1994).
54. Porcher, C. et al. The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. *Cell* **86**, 47-57 (1996).
55. Warren, A. et al. The oncogenic cysteine-rich LIM domain protein Rbtn2 is essential for erythroid development. *Cell* **78**, 45-57 (1994).
56. Okuda, T., Deursen, J. v., Hlebert, S., Grosveld, F. & JR, D. AML1, the targer of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell* **84**, 321-330 (1996).
57. Cantor, A. B. & Orkin, S. H. Transcriptional regulation of erythropoiesis: an affair involving multiple partners. *Oncogene* **21**, 3368-76 (2002).
58. Scott, E., Simon, M., Anastasi, J. & Singh, H. Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science* **265**, 1573-7 (1994).
59. Georgopoulos, K. et al. The Ikaros gene is required for the development of all lymphoid lineages. *Cell* **79**, 143-56 (1994).
60. Pevny, L. et al. Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* **349**, 257-60 (1991).
61. Saleque, S., Cameron, S. & Orkin, S. H. The zinc-finger proto-oncogene Gfi-1b is essential for development of the erythroid and megakaryocytic lineages. *Genes Dev* **16**, 301-6 (2002).
62. Saleque, S., Cameron, S. & Orkin, S. H. The zinc-finger proto-oncogene Gfi-1b is essential for development of the erythroid and megakaryocytic lineages. *Genes Dev* **16**, 301-6 (2001).
63. Osada, H., Grutz, G., Axelson, H., Forster, A. & Rabbitts, T. H. Association of erythroid transcription factors: complexes involving the LIM protein RBTN2 and the zinc-finger protein GATA1. *Proc Natl Acad Sci U S A* **92**, 9585-9 (1995).
64. Tsai, F. & Orkin, S. Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. *Blood* **89**, 3636-43 (1997).
65. Heyworth, C., Gale, K., Dexter, M., May, G. & Enver., T. A GATA-2/estrogen receptor chimera functions as a ligand-dependent negative regulator of self-renewal. *Genes Dev* **13**, 1847-60 (1999).
66. Weiss, M., Keller, G. & Orkin, S. Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. *Genes Dev* **8**, 1184-97 (1994).
67. Crispino, J., Lodish, M., MacKay, J. & Orkin, S. Use of altered specificity mutants to probe a specific protein-protein interaction in differentiation: the GATA-1:FOG complex. *Mol Cell* **3**, 219-28 (1999).
68. Anguita, E. et al. Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. *Embo J* **23**, 2841-52 (2004).
69. Letting, D. L., Chen, Y. Y., Rakowski, C., Reedy, S. & Blobel, G. A. Context-dependent regulation of GATA-1 by friend of GATA-1. *Proc Natl Acad Sci U S A* **101**, 476-81 (2004).
70. Pal, S. et al. Coregulator-dependent facilitation of chromatin occupancy by GATA-1. *Proc Natl Acad Sci U S A* **101**, 980-5 (2004).
71. Pandolfi, P. et al. Targeted disruption of the GATA3 gene causes severe abnormalities in the nervous system and in fetal liver haematopoiesis. *Nat Genet* **11**, 40-4 (1995).
72. Peterkin, T., Gibson, A., Loose, M. & Patient, R. The roles of GATA-4, -5 and -6 in vertebrate heart development. *Semin Cell Dev Biol* **16**, 83-94 (2005).
73. Molkentin, J. The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression. *J Biol Chem* **275**, 38949-52 (2000).
74. Kuo, H. et al. Control of segmental expression of the cardiac-restricted ankyrin repeat protein gene by distinct regulatory pathways in murine cardiogenesis. *Development* **126**, 4223-34 (1999).
75. Molkentin, J., Lin, Q., Duncan, S. & Olson, E. Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes Dev* **11**, 1061-72 (1997).
76. Morrisey, E. et al. GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes Dev* **12**, 3579-90 (1998).
77. Tsai, S. F. et al. Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature* **339**, 446-51 (1989).

78. Martin, D., Zon, L., Mutter, G. & Orkin, S. Expression of an erythroid transcription factor in megakaryocytic and mast cell lineages. *1990* **344**, 444-7 (1990).
79. Romeo, P. et al. 1990. Megakaryocytic and erythrocytic lineages share specific transcription factors. *Nature* **344**, 447-9.
80. Zon, L. et al. Expression of mRNA for the GATA-binding proteins in human eosinophils and basophils: potential role in gene transcription. *Blood* **81**, 3234-41 (1993).
81. Zon, L. et al. GATA-binding transcription factors in mast cells regulate the promoter of the mast cell carboxypeptidase A gene. *J Biol Chem* **266**, 22948-53 (1991).
82. Yomogida, K. et al. Developmental stage- and spermatogenic cycle-specific expression of transcription factor GATA-1 in mouse Sertoli cells. *Development* **120**, 1759-66 (1994).
83. Evans, T. & Felsenfeld, G. The erythroid-specific transcription factor Eryf1: a new finger protein. *Cell* **58**, 877-85 (1989).
84. Wall, L., deBoer, E. & Grosveld, F. The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. *Genes Dev* **2**, 1089-100 (1988).
85. Ito, E. et al. Erythroid transcription factor GATA-1 is abundantly transcribed in mouse testis. *nature* **362**, 466-8 (1993).
86. Tsai, S., Strauss, E. & Orkin, S. Functional analysis and in vivo footprinting implicate the erythroid transcription factor GATA-1 as a positive regulator of its own promoter. *Genes Dev* **5**, 919-31 (1991).
87. Hannon, R., Evans, T., Felsenfeld, G. & Gould, H. Structure and promoter activity of the gene for the erythroid transcription factor GATA-1. *Proc Natl Acad Sci U S A* **88**, 3004-8 (1991).
88. Onodera, K. et al. GATA-1 transcription is controlled by distinct regulatory mechanisms during primitive and definitive erythropoiesis. *Proc Natl Acad Sci U S A* **94**, 4487-92 (1997).
89. Vyas, P. et al. Different sequence requirements for expression in erythroid and megakaryocytic cells within a regulatory element upstream of the GATA-1 gene. *Development* **126**, 2799-811 (1999).
90. Takahashi, S. et al. Arrest in primitive erythroid cell development caused by promoter-specific disruption of the GATA-1 gene. *J Biol Chem* **272**, 12611-5 (1997).
91. McDevitt, M., Fujiwara, Y., RA, R. S. & Orkin, S. An upstream, DNase I hypersensitive region of the hematopoietic-expressed transcription factor GATA-1 gene confers developmental specificity in transgenic mice. *Proc Natl Acad Sci U S A* **24**, 7976-81 (1997).
92. Weiss, M. J., Yu, C. & Orkin, S. H. Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. *Mol Cell Biol* **17**, 1642-51 (1997).
93. Visvader, J., Crossley, M., Hill, J., Orkin, S. & Adams, J. The C-terminal zinc finger of GATA-1 or GATA-2 is sufficient to induce megakaryocytic differentiation of an early myeloid cell line. *Mol Cell Biol* **15**, 634-41 (1995).
94. Shimizu, R., Takahashi, S., Ohneda, K., Engel, J. D. & Yamamoto, M. In vivo requirements for GATA-1 functional domains during primitive and definitive erythropoiesis. *Embo J* **20**, 5250-60 (2001).
95. Li, Z. et al. Developmental stage-selective effect of somatically mutated leukemogenic transcription factor GATA1. *Nat Genet* **37**, 613-9 (2005).
96. Yang, H. & Evans, T. Distinct roles for the two cGATA-1 finger domains. *Mol Cell Biol* **12**, 4562-70 (1992).
97. Merika, M. & Orkin, S. Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Kruppel family proteins Sp1 and EKLF. *Mol Cell Biol* **15**, 2437-47 (1995).
98. Rekhtman, N., Radparvar, F., Evans, T. & Skoultchi, A. Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells. *Genes Dev* **13**, 1398-411 (1999).
99. Whyatt, D. J., deBoer, E. & Grosveld, F. The two zinc finger-like domains of GATA-1 have different DNA binding specificities. *Embo J* **12**, 4993-5005 (1993).
100. Newton, A., Mackay, J. & Crossley, M. The N-terminal zinc finger of the erythroid transcription factor GATA-1 binds GATC motifs in DNA. *J Biol Chem* **276**, 35794-801 (2001).
101. Blobel, G. A., Nakajima, T., Eckner, R., Montminy, M. & Orkin, S. H. CREB-binding protein cooperates with transcription factor GATA-1 and is required for erythroid differentiation. *Proc Natl Acad Sci U S A* **95**, 2061-6 (1998).

102. Mackay, J. et al. Involvement of the N-finger in the self-association of GATA-1. *J Biol Chem* **273**, 30560-7 (1998).
103. Crossley, M., Merika, M. & Orkin, S. Self-association of the erythroid transcription factor GATA-1 mediated by its zinc finger domains. *Mol Cell Biol* **15**, 2448-56 (1995).
104. Boyes, J., Byfield, P., Nakatani, Y. & Ogryzko, V. Regulation of activity of the transcription factor GATA-1 by acetylation. *Nature* **396**, 594-8 (1998).
105. Hung, H. L., Lau, J., Kim, A. Y., Weiss, M. J. & Blobel, G. A. CREB-Binding protein acetylates hematopoietic transcription factor GATA-1 at functionally important sites. *Mol Cell Biol* **19**, 3496-505 (1999).
106. Partington, G. & Patient, R. Phosphorylation of GATA-1 increases its DNA-binding affinity and is correlated with induction of human K562 erythroleukaemia cells. *Nucleic Acids Res* **27**, 1168-75 (1999).
107. Crossley, M. & Orkin, S. H. Phosphorylation of the erythroid transcription factor GATA-1. *J Biol Chem* **269**, 16589-96 (1994).
108. Towatari, M., Ciro, M., Ottolenghi, S., Tzuzuki, S. & T, E. Involvement of mitogen-activated protein kinase in the cytokine-regulated phosphorylation of transcription factor GATA-1. *Hematol J* **5**, 262-72 (2004).
109. Yu, Y. et al. MAPK-mediated phosphorylation of GATA-1 promotes Bcl-XL expression and cell survival. *J Biol Chem* **280**, 29533-42 (2005).
110. Kadri, Z. et al. Phosphatidylinositol 3-kinase/Akt induced by erythropoietin renders the erythroid differentiation factor GATA-1 competent for TIMP-1 gene transactivation. *Mol Cell Biol* **25**, 7412-22 (2005).
111. Collavin, L. et al. Modification of the erythroid transcription factor GATA-1 by SUMO-1. *Proc Natl Acad Sci U S A* **101**, 8870-5 (2004).
112. Hay, R. SUMO:a history of modification. *Mol Cell* **18**, 1-12 (2005).
113. Sapetschnig, A. et al. Transcription factor Sp3 is silenced through SUMO modification by PIAS1. *Embo J* **21**, 5206-15 (2002).
114. Gostissa, M. et al. Activation of p53 by conjugation to the ubiquitin-like protein SUMO-1. *Embo J* **18**, 6462-71 (1999).
115. Ishov, A. et al. PML is critical for ND10 formation and recruits the PML-interacting protein daxx to this nuclear structure when modified by SUMO-1. *J Cell Biol* **147**, 221-34 (1999).
116. Desterro, J., Rodriguez, M. & Hay, R. SUMO-1 modification of IkappaBalpha inhibits NF-kappaB activation. *Mol Cell* **2**, 233-9 (1998).
117. Ohneda, K. & Yamamoto, M. Roles of hematopoietic transcription factors GATA-1 and GATA-2 in the development of red blood cell lineage. *Acta Haematol.* **108**, 237-45 (2002).
118. Gutierrez, L. et al. Homotypic signaling regulates Gata1 activity in the erythroblastic island. *Development* **131**, 3183-93 (2004).
119. Pevny, L. et al. Development of hematopoietic cells lacking transcription factor GATA-1. *Development* **121**, 163-72 (1995).
120. Weiss, M. J. & Orkin, S. H. Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis. *Proc Natl Acad Sci U S A* **92**, 9623-7 (1995).
121. Fujiwara, Y., Browne, C. P., Cunniff, K., Goff, S. C. & Orkin, S. H. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc Natl Acad Sci U S A* **93**, 12355-8 (1996).
122. Pan X, O. O., Ohneda K, Lindeboom F, Iwata F, Shimizu R, Nagano M, Suwabe N, Philipsen S, Lim KC, Engel JD, Yamamoto M. Graded levels of GATA-1 expression modulate survival, proliferation, and differentiation of erythroid progenitors. *J Biol Chem* **280**, 22385-94 (2005).
123. Whyatt, D. et al. The level of the tissue-specific factor GATA-1 affects the cell-cycle machinery. *Genes Funct* **1**, 11-24 (1997).
124. Whyatt, D. et al. An intrinsic but cell-nonautonomous defect in GATA-1-overexpressing mouse erythroid cells. *Nature* **406**, 519-24 (2000).
125. DJ, D. W. et al. The level of the tissue-specific factor GATA-1 affects the cell-cycle machinery. *Genes Funct* **1**, 11-24 (1997).
126. Gutierrez, L. et al. Homotypic signaling regulates Gata1 activity in the erythroblastic island. *Development* **131**, 3183-93 (2004).

127. Maria, R. D. et al. Negative regulation of erythropoiesis by caspase-mediated cleavage of GATA-1. *Nature* **401**, 489-93 (1999.).
128. Shivdasani, R. A., Fujiwara, Y., McDevitt, M. A. & Orkin, S. H. A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. *Embo J* **16**, 3965-73 (1997).
129. Hirasawa, R. et al. Essential and instructive roles of GATA factors in eosinophil development. *J Exp Med* **195**, 1379-86 (2002).
130. Yu, C. et al. Targeted deletion of a high-affinity GATA-binding site in the GATA-1 promoter leads to selective loss of the eosinophil lineage in vivo. *J Exp Med* **195**, 1387-95 (2002).
131. Yamaguchi, Y. et al. Mechanisms of transcription in eosinophils: GATA-1, but not GATA-2, transactivates the promoter of the eosinophil granule major basic protein gene. *Blood* **91**, 3447-58 (1998).
132. Harigae, H. et al. Differential roles of GATA-1 and GATA-2 in growth and differentiation of mast cells. *Genes Cells* **3**, 39-50 (1998).
133. Lindeboom, F. et al. A tissue-specific knockout reveals that Gata1 is not essential for Sertoli cell function in the mouse. *Nucleic Acids Res* **31**, 5405-12 (2003).
134. Kulessa, H., Frampton, J. & Graf, T. GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts. *Genes Dev* **9**, 1250-62 (1995).
135. Ferreira, R., Ohneda, K., Yamamoto, M. & Philipsen, S. GATA1 function, a paradigm for transcription factors in hematopoiesis. *Mol Cell Biol* **25**, 1215-27 (2005).
136. Fox, A. H. et al. Transcriptional cofactors of the FOG family interact with GATA proteins by means of multiple zinc fingers. *Embo J* **18**, 2812-22 (1999).
137. Tsang, A. P., Fujiwara, Y., Hom, D. B. & Orkin, S. H. Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG. *Genes Dev* **12**, 1176-88 (1998).
138. Nichols, K. E. et al. Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. *Nat Genet* **24**, 266-70 (2000).
139. Martowicz, M., Grass, J., Boyer, M., Guend, H. & Bresnick, E. Dynamic GATA factor interplay at a multicomponent regulatory region of the GATA-2 locus. *J Biol Chem* **280**, 1724-32 (2005).
140. Shimizu, R., Ohneda, K., Engel, J., Trainor, C. & Yamamoto, M. Transgenic rescue of GATA-1-deficient mice with GATA-1 lacking a FOG-1 association site phenocopies patients with X-linked thrombocytopenia. *Blood* **103**, 2560-7 (2004).
141. Querfurth, E. et al. Antagonism between C/EBPbeta and FOG in eosinophil lineage commitment of multipotent hematopoietic progenitors. *Genes Dev* **14**, 2515-25 (2000).
142. Begley, C. et al. Chromosomal translocation in a human leukemic stem-cell line disrupts the T-cell antigen receptor delta-chain diversity region and results in a previously unreported fusion transcript. *Proc Natl Acad Sci U S A* **86**, 2031-2035 (1989).
143. Begley, C. et al. The gene SCL is expressed during early hematopoiesis and encodes a differentiation-related DNA-binding motif. *Proc Natl Acad Sci U S A* **86**, 10128-32 (1989).
144. Finger, L. et al. Involvement of the TCL5 gene on human chromosome 1 in T-cell leukemia and melanoma. *Proc Natl Acad Sci U S A* **86**, 5039-43 (1989).
145. Chen, Q. et al. The tal gene undergoes chromosome translocation in T cell leukemia and potentially encodes a helix-loop-helix protein. *Embo J* **9**, 414-24 (1990).
146. Green, A. & Begley, C. SCL and related hemopoietic helix-loop-helix transcription factors. *Int J Cell Cloning.*. **10**, 269-76 (1992).
147. Mouthon, M. et al. Expression of tal-1 and GATA-binding proteins during human hematopoiesis. *Blood* **81**, 647-55 (1993).
148. Shivdasani, R., Mayer, E. & Orkin, S. Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. *Nature* **373**, 432-4 (1995).
149. Lecointe, N. et al. GATA-and SP1-binding sites are required for the full activity of the tissue-specific promoter of the tal-1 gene. *Oncogene* **9**, 2623-32 (1994).
150. Hall, M. A. et al. The critical regulator of embryonic hematopoiesis, SCL, is vital in the adult for megakaryopoiesis, erythropoiesis, and lineage choice in CFU-S12. *Proc Natl Acad Sci U S A* **100**, 992-7 (2003).

151. Mikkola, H. K. et al. Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene. *Nature* **421**, 547-51 (2003).
152. Yamada, Y. et al. The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis. *Proc Natl Acad Sci U S A* **95**, 3890-5 (1998).
153. Drechsler, M. et al. Genomic structure, alternative transcripts and chromosome location of the human LIM domain binding protein 1 gene LDB1. *Cytogenet Cell Genet* **87**, 119-24 (1999).
154. Mukhopadhyay, M. et al. Functional ablation of the mouse Ldb1 gene results in severe patterning defects during gastrulation. *Development* **130**, 495-505 (2003).
155. Cohen-Kaminsky, S. et al. Chromatin immunoselection defines a TAL-1 target gene. *Embo J* **17**, 5151-60 (1998).
156. Lecuyer, E. et al. The SCL complex regulates c-kit expression in hematopoietic cells through functional interaction with Sp1. *Blood* **100**, 2430-40 (2002).
157. Anderson, K. P., Crable, S. C. & Lingrel, J. B. Multiple proteins binding to a GATA-E box-GATA motif regulate the erythroid Kruppel-like factor (EKLF) gene. *J Biol Chem* **273**, 14347-54 (1998).
158. Lahlil, R., Lecuyer, E., Herblot, S. & Hoang, T. SCL assembles a multifactorial complex that determines glycophorin A expression. *Mol Cell Biol* **24**, 1439-52 (2004).
159. Huang, S., Qiu, Y., Stein, R. & Brandt, S. p300 functions as a transcriptional coactivator for the TAL1/SCL oncoprotein. *Oncogene* **18**, 4858-67 (1999).
160. Huang, S. & Brandt, S. mSin3A regulates murine erythroleukemia cell differentiation through association with the TAL1 (or SCL) transcription factor. *Mol Cell Biol*. **20**, 2248-59 (2000).
161. Miller, I. & Bieker, J. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. *Mol Cell Biol* **13**, 2776-86 (1993).
162. Raich, N. & Romeo, P. Erythroid regulatory elements. *Stem Cells* **11**, 95-104 (1993).
163. Chen, X. & Bieker, J. Erythroid Krüppel-like factor (EKLF) contains a multifunctional transcriptional activation domain important for inter- and intramolecular interactions. *Embo J* **15**, 5888-96 (1996).
164. Pandya, K., Donze, D. & townes, T. Novel transactivation domain in erythroid Kruppel-like factor (EKLF). *J Biol Chem* **276**, 8239-43 (2001).
165. Southwood, C., Downs, K. & Bieker, J. Erythroid Kruppel-like factor exhibits an early and sequentially localized pattern of expression during mammalian erythroid ontogeny. *Dev Dyn* **206**, 248-59 (1996).
166. Crossley, M., Tsang, A., Bieker, J. & Orkin, S. Regulation of the erythroid Kruppel-like factor (EKLF) gene promoter by the erythroid transcription factor GATA-1. *J Biol Chem* **269**, 15440-4 (1994).
167. Anderson, K. P., Crable, S. C. & Lingrel, J. B. The GATA-E box-GATA motif in the EKLF promoter is required for in vivo expression. *Blood* **95**, 1652-5 (2000).
168. Nuez, B., Michalovich, D., Bygrave, A., Ploemacher, R. & Grosveld, F. Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature* **375**, 316-8 (1995).
169. Perkins, A., Sharpe, A. & Orkin, S. Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *nature* **375**, 318-22 (1995).
170. Fautino, P., Lavinha, J., Marini, M. & Moi, P. beta-Thalassemia mutation at -90C-->T impairs the interaction of the proximal CACCC box with both erythroid and nonerythroid factors. *Blood* **88**, 3248-9 (1996).
171. Feng, W., Southwood, C. & Bieker, J. Analyses of beta-thalassemia mutant DNA interactions with erythroid Kruppel-like factor (EKLF), an erythroid cell-specific transcription factor. *J Biol Chem* **269**, 1493-500 (1994).
172. Kulozik, A., Bellan-Koch, A., Bail, S., Kohne, E. & Kleihauer, E. Thalassemia intermedia: moderate reduction of beta globin gene transcriptional activity by a novel mutation of the proximal CACCC promoter element. *Blood* **77**, 2054-8 (1991).
173. Wijgerde, M. et al. The role of EKLF in human beta-globin gene competition. *Genes Dev* **10**, 2894-902 (1996).
174. Drissen, R. et al. The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev* **15**, 2485-9 (2004).

175. Drissen, R. et al. The erythroid phenotype of EKLF-null mice: defects in hemoglobin metabolism and membrane stability. *Mol Cell Biol* **25**, 5205-14 (2005).
176. Welch, J. J. et al. Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood* **104**, 3136-47 (2004).
177. Chen, X. & Bieker, J. Unanticipated repression function linked to erythroid Kruppel-like factor. *Mol Cell Biol* **21**, 3118-25 (2001).
178. Zhang, W. & Bieker, J. Acetylation and modulation of erythroid Krüppel-like factor (EKLF) activity by interaction with histone acetyltransferases. *Proc Natl Acad Sci U S A* **95**, 9855-9860 (1998).
179. Amstrong, J., Bieker, J. & Emerson, B. A SWI/SNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLF in vitro. *Cell* **95**, 93-104 (1998).
180. Fisher, R. & Scott, E. Role of PU.1 in hematopoiesis. *Stem Cells* **16**, 25-37 (1998).
181. McKercher, S. et al. Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities. *Embo J* **15**, 5647-58 (1996).
182. Zhang, P. et al. PU.1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding. *Blood* **86**, 2641-8 (2000).
183. Zhang, P. et al. Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. *Proc Natl Acad Sci U S A* **96**, 8705-10 (1999).
184. Rekhtman, N. et al. PU.1 and pRB interact and cooperate to repress GATA-1 and block erythroid differentiation. *Mol Cell Biol* **23**, 7460-74 (2003).
185. Rhodes, J. et al. Interplay of pu.1 and gata1 determines myelo-erythroid progenitor cell fate in zebrafish. *Dev Cell* **8**, 97-108 (2005).
186. Goodman, R. & Smolik, S. CBP/p300 in cell growth, transformation, and development. *Genes Dev* **14**, 1553-77 (2000).
187. Lipinski, M. & Jacks, T. The retinoblastoma gene family in differentiation and development. *Oncogene* **18**, 7873-82 (1999).
188. Clarke, A. et al. Requirement for a functional Rb-1 gene in murine development. *Nature* **359**, 328-30 (1992).
189. Clark, A., Doyle, K. & Humbert., P. Cell-intrinsic requirement for pRb in erythropoiesis. *Blood* **104**, 1324-6 (2004).
190. Horak, C. et al. GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIp-chip analysis. *Proc Natl Acad Sci U S A* **99**, 2924-9 (2002).
191. Vakoc, C. et al. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol Cell* **17**, 453-62 (2005).
192. Splinter, E., Grosveld, F. & Laat, W. d. in *Methods Enzymol* (2004).
193. Rylski, M. et al. GATA-1-mediated proliferation arrest during erythroid maturation. *Mol Cell Biol* **23**, 5031-42 (2003).
194. Lacombe, C. & Mayeux, P. The molecular biology of erythropoietin. *Nephrol Dial Transplant* **14**, 22-8 (1999).
195. Grass, J. A. et al. GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. *Proc Natl Acad Sci U S A* **100**, 8811-6 (2003).
196. Ackerman, S., Kephart, G., Habermann, T., Greipp, P. & Gleich, G. Localization of eosinophil granule major basic protein in human basophils. *J Exp Med* **158**, 946-61 (1983).
197. Ackerman, S. et al. Distinctive cationic proteins of the human eosinophil granule: major basic protein, eosinophil cationic protein, and eosinophil-derived neurotoxin. *J Immunol* **131**, 2977-82 (1983).
198. Du, J. et al. Novel combinatorial interactions of GATA-1, PU.1, and C/EBPepsilon isoforms regulate transcription of the gene encoding eosinophil granule major basic protein. *J Biol Chem* **277**, 43481-94 (2002).
199. Calligaris, R., Bottardi, S., Cogoi, S., Apezteguia, I. & Santoro, C. Alternative translation initiation site usage results in two functionally distinct forms of the GATA-1 transcription factor**.** *Proc Natl Acad Sci U S A* **92**, 11598-602 (1995).
200. Wechsler, J. et al. Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. *Nat Genet* **21**, 148-52 (2002).

201. Crispino, J. GATA1 in normal and malignant hematopoiesis. *Semin Cell Dev Biol* **16**, 137-47 (2005).
202. Wechsler, J. et al. Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. *Nat Genet* **32**, 148-52 (2002).
203. Peterson, C. & Laniel, M. Histones and histone modifications. *Curr Biol.* **14**, 546-51 (2004).
204. Gilbert, N., Gilchrist, S. & Bickmore, W. A. Chromatin organization in the Mammalian nucleus. *Int Rev Cytol* **242**, 283-336 (2004).
205. Weintraub, H. & Groudine, M. Chromosomal subunits in active genes have an altered conformation. *Sience* **193**, 848-56 (1976).
206. Luger, K., Rechsteiner, T., Flaus, A., Waye, M. & hmond, T. Characterization of nucleosome core particles containing histone proteins made in bacteria. *J Mol Biol* **272**, 310-11 (1997).
207. Workman, J. & Kingston, R. Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu Rev Biochem*, 545-79 (1998).
208. Roth, S., Denu, J. & Allis, C. Histone acetyltransferases. *Annu Rev Biochem.*, 81-120 (2001).
209. Goldman, P., Tran, V. & Goodman, R. The multifunctional role of the co-activator CBP in transcriptional regulation. *Recent Prog Horm Res*, 103-19 (1997).
210. Mizzen, C. et al. The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. *Cell* **87**, 1261-70 (1996).
211. Dhalluin, C. et al. Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399**, 491-6 (1999).
212. Davie, J. & Chadee, D. Regulation and regulatory parameters of histone modifications. *J Cell Biochem Suppl* **30**, 203-13 (1998).
213. Brehm, A. et al. Retinoblastoma protein recruits histone deacetylase to repress transcription. *Nature* **391**, 597-601 (1998).
214. Drummond, D. et al. Clinical development of histone deacetylase inhibitors as anticancer agents. *Annu Rev Pharmacol Toxicol*, 495-528 (2005).
215. Murray, K. The Occurence of Epsilon-N-methyl lysine in histones. *Biochemistry*, 10-5 (1964).
216. Tachibana, M., Sugimoto, K., Fukushima, T. & Shinkai, Y. Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *J Biol Chem* **276**, 25309-17 (2001).
217. Zhang, Y. & Reinberg, D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev* **15**, 2343-60 (2001).
218. Nakayama, J., Rice, J., Strahl, B., Allis, C. & Grewal, S. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**, 110-3 (2001).
219. Bannister, A. et al. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120-4 (2001).
220. Santos-Rosa, H. et al. Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407-11 (2002).
221. Jenuwein, T. Re-SET-ting heterochromatin by histone methyltransferases. *Trends Cell Biol.* **11**, 266-73 (2001).
222. Shi, Y. et al. Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature* **422**, 735-8 (2003).
223. Humphrey, G. et al. Stable histone deacetylase complexes distinguished by the presence of SANT domain proteins CoREST/kiaa0071 and Mta-L1. *J Biol Chem* **276**, 6817-24 (2001).
224. Lo, W. et al. Snf1--a histone kinase that works in concert with the histone acetyltransferase Gcn5 to regulate transcription. *Science* **293**, 1142-6 (2001).
225. Talasz, H., Helliger, W., Puschendorf, B. & Lindner., H. In vivo phosphorylation of histone H1 variants during the cell cycle. *Biochemistry* **35**, 1761-7 (1996).
226. Strahl, B. & Allis, C. The language of covalent histone modifications. *Nature* **403**, 41-5 (2000).
227. Margueron, R., Trojer, P. & Reinberg., D. The key to development: interpreting the histone code? *Curr Opin Genet Dev* **15**, 163-76 (2005).
228. Ragvin, A. et al. Nucleosome binding by the bromodomain and PHD finger of the transcriptional cofactor p300. *J Mol Biol* **337**, 773-88 (2004).
229. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein., T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **4410**, 116-20 (2001).

230. Schneider, R., Bannister, A., Weise, C. & Kouzarides, T. Direct Binding of INHAT to H3 Tails Disrupted by Modifications. *J Biol Chem* **278**, 23859-62 (2004).
231. Zegerman, P., Canas, B., Pappin, D. & Kouzarides., T. Histone H3 lysine 4 methylation disrupts binding of nucleosome remodeling and deacetylase (NuRD) repressor complex. *J Biol Chem* **277**, 11621-4 (2002).
232. Danzer, J. & Wallrath, L. Mechanisms of HP1-mediated gene silencing in Drosophila. *Development* **131**, 3571-80 (2004).
233. Rea, S. et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593-9 (2000).
234. Cheung, P. et al. Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. *Mol Cell* **5**, 905-15 (2000).
235. Henikoff, S., Furuyama, T. & Ahmad, K. Histone variants, nucleosome assembly and epigenetic inheritance. *Trends in Genetics* **20**, 320-6 (2004).
236. Ahmad, K. & Henikoff, S. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol Cell* **9**, 1191-200 (2002).
237. Schwartz, B. & Ahmad, K. Transcriptional activation triggers deposition and removal of the histone variant H3.3. *Genes Dev* **19**, 804-14 (2005).
238. Narlikar, G. J., Fan, H. Y. & Kingston, R. E. Cooperation between complexes that regulate chromatin structure and transcription. *Cell* **108**, 475-87 (2002).
239. Eberharter, A. & Becker, P. ATP-dependent nucleosome remodelling: factors and functions. *J Cell Sci* **117**, 3707-11 (2004).
240. Winston, F. & Carlson, M. Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection. *Trends Genet* **8**, 387-91 (1992).
241. Khavari, P., Peterson, C., Tamkun, J., Mendel, D. & Crabtree, G. BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription. *Nature* **366**, 170-4 (1993).
242. Muchardt, C. & Yaniv, M. A human homologue of Saccharomyces cerevisiae SNF2/SWI2 and Drosophila brm genes potentiates transcriptional activation by the glucocorticoid receptor. *Embo J* **12**, 4279-90 (1993).
243. Neely, K. et al. Activation domain-mediated targeting of the SWI/SNF complex to promoters stimulates transcription from nucleosome arrays. *Mol Cell* **4**, 649-55 (1999).
244. Yudkovsky, N., Logie, C., Hahn, S. & Peterson, C. Recruitment of the SWI/SNF chromatin remodeling complex by transcriptional activators. *Genes Dev* **13**, 2369-74 (1999).
245. Cosma, M., Tanaka, T. & Nasmyth., K. Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter. *Cell* **97**, 299-311 (1999).
246. Lee, C., Murphy, M., Lee, J. & Chung., J. Targeting a SWI/SNF-related chromatin remodeling complex to the beta-globin promoter in erythroid cells. *Proc Natl Acad Sci U S A* **96**, 12311-5 (1999).
247. Kowenz-Leutz, E. & Leutz., A. A C/EBP beta isoform recruits the SWI/SNF complex to activate myeloid genes. *Mol Cell* **4**, 735-43 (1999).
248. Kadam, S. et al. Functional selectivity of recombinant mammalian SWI/SNF subunits. *Genes Dev* **14**, 2441-51 (2000).
249. Holstege, F. et al. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717-28 (1998).
250. Corona, D. F. & Tamkun, J. W. Multiple roles for ISWI in transcription, chromosome organization and DNA replication. *Biochim Biophys Acta* **1677**, 113-9 (2004).
251. Zhou, Y., Santoro, R. & Grummt, I. The chromatin remodeling complex NoRC targets HDAC1 to the ribosomal gene promoter and represses RNA polymerase I transcription. *Embo J* **21**, 4632-40 (2002).
252. Collins, N. et al. An ACF1-ISWI chromatin-remodeling complex is required for DNA replication through heterochromatin. *Nat Genet* **32**, 627-32 (2002).
253. Bultman, S. et al. A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. *Mol Cell* **6**, 1287-95 (2000).

254. Stopka, T. & Skoultchi, A. I. The ISWI ATPase Snf2h is required for early mouse development. *Proc Natl Acad Sci U S A* **100**, 14097-102 (2003).
255. Tsukiyama, T. & Wu, C. Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell* **83**, 1011-20 (1995).
256. Aalfs, J., Narlikar, G. & Kingston, R. Functional differences between the human ATP-dependent nucleosome remodeling proteins BRG1 and SNF2H. *J Biol Chem* **276**, 34270-8 (2001).
257. Zhang, Y., LeRoy, G., Seelig, H., Lane, W. & Reinberg., D. The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell* **95**, 279-89 (1998).
258. Wade, P., Jones, P., Vermaak, D. & Wolffe., A. A multiple subunit Mi-2 histone deacetylase from Xenopus laevis cofractionates with an associated Snf2 superfamily ATPase. *Curr Biol.* **8**, 843-6 (1998).
259. Feng, Q. & Zhang, Y. The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. *Genes Dev* **15**, 827-32 (2001).
260. Hendrich, B. & Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *18* **11** (1998).
261. Zhang, Y. et al. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev* **13**, 1924-35 (1999).
262. Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V. A. & Bird, A. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes Dev* **15**, 710-23 (2001).
263. Ng, H. H. et al. MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat Genet* **23**, 58-61 (1999).
264. Feng, Q. et al. Identification and functional characterization of the p66/p68 components of the MeCP1 complex. *Mol Cell Biol* **22**, 536-46 (2002).
265. Ahringer, J. NuRD and SIN3 histone deacetylase complexes in development. *Trends Genet* **16**, 351-6 (2000).
266. Hutchins, A. S. et al. Gene silencing quantitatively controls the function of a developmental trans-activator. *Mol Cell* **10**, 81-91 (2002).
267. Kehle, J. et al. dMi-2, a hunchback-interacting protein that functions in polycomb repression. *Science* **282**, 1897-900 (1998).
268. Wang, J. et al. Selective defects in the development of the fetal and adult lymphoid system in mice with an Ikaros null mutation. *Immunity* **5**, 537-49 (1996).
269. Kim, J. et al. Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes. *Immunity* **10**, 345-55 (1999).
270. Bird, A. & Wolffe, A. Methylation-induced repression--belts, braces, and chromatin. *Cell* **99**, 451-4 (1999).
271. Williams, C. J. et al. The chromatin remodeler Mi-2beta is required for CD4 expression and T cell development. *Immunity* **20**, 719-33 (2004).
272. Jiang, S. J., Rauch, T., Li, H. & Pfeifer., G. MBD3L2 interacts with MBD3 and components of the NuRD complex and can oppose MBD2-MeCP1-mediated methylation silencing. *J Biol Chem* **280**, 12700-9 (2005).
273. Hong W, N. M., Chen YY, Kori R, Vakoc CR, Rakowski C, Blobel GA. FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. *Embo J* 24, 2367-78 (2005).

Chapter 2:

Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice

Ernie de Boer, Patrick Rodriguez, Edgar Bonte, Jeroen Krijgsveld[1], Eleni Katsantoni, Albert Heck[1], Frank Grosveld and John Strouboulis

Proc Natl Acad Sci U S A **100**, 7480-5 (2003)

Department of Cell Biology, Erasmus Medical Center, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
[1]Department of Biomolecular Mass Spectrometry, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht , The Netherlands

Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice

Ernie de Boer*, Patrick Rodriguez*, Edgar Bonte*, Jeroen Krijgsveld†, Eleni Katsantoni*, Albert Heck†, Frank Grosveld*, and John Strouboulis*‡

*Department of Cell Biology, Erasmus Medical Center, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands; and †Department of Biomolecular Mass Spectrometry, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, The Netherlands

Communicated by Gary Felsenfeld, National Institutes of Health, Bethesda, MD, May 1, 2003 (received for review February 26, 2003)

Proteomic approaches require simple and efficient protein purification methodologies that are amenable to high throughput. Biotinylation is an attractive approach for protein complex purification due to the very high affinity of avidin/streptavidin for biotinylated templates. Here, we describe an approach for the single-step purification of transcription factor complex(es) based on specific *in vivo* biotinylation. We expressed the bacterial BirA biotin ligase in mammalian cells and demonstrated very efficient biotinylation of a hematopoietic transcription factor bearing a small (23-aa) artificial peptide tag. Biotinylation of the tagged transcription factor altered neither the factor's protein interactions or DNA binding properties *in vivo* nor its subnuclear distribution. Using this approach, we isolated the biotin-tagged transcription factor and at least one other known interacting protein from crude nuclear extracts by direct binding to streptavidin beads. Finally, this method works efficiently in transgenic mice, thus raising the prospect of using biotinylation tagging in protein complex purification directly from animal tissues. Therefore, BirA-mediated biotinylation of tagged proteins provides the basis for the single-step purification of proteins from mammalian cells.

In the postgenome-sequencing era, focus has shifted toward the identification and characterization of the protein complement of cells, the proteome. A crucial aspect of this effort is the utilization of simple and efficient methodologies that are amenable to high-throughput approaches for the purification of proteins and protein complexes (1, 2). As a result, a number of generic affinity-based methodologies have been developed for these purposes, based primarily on the use of specific antibodies or affinity tags that are fused to the protein of interest (3, 4).

Prominent among the affinity-based purification methodologies is the biotin/avidin system. Biotin is a naturally occurring cofactor for metabolic enzymes, which is active only when covalently attached to the enzymes through the action of specific protein–biotin ligases (5). Any biotinylated substrate can be bound very tightly by the proteins avidin and streptavidin. Biotin/avidin binding is the strongest noncovalent interaction known in nature ($K_d = 10^{-15}$ M), several orders higher than that of commonly used antibodies or other affinity tags. As a result, the biotin/avidin affinity system has numerous applications in modern biological techniques (6). For the purposes of protein purification, in particular, biotinylation offers a number of advantages. For example, the high affinity of biotin for avidin/streptavidin allows purification of the biotinylated protein under high stringency conditions, thus reducing background binding often observed with other affinity tags that elute more easily. In addition, there are very few naturally biotinylated proteins, thus reducing the chance for crossreaction when using biotinylation in protein purification, as opposed to antibodies that may crossreact with several species.

The potential advantages of biotinylation tagging in protein purification have not gone unnoticed (7). The characterization of the minimal amino acid sequence requirements of naturally biotinylated proteins has led to the development of sequence tags that can be biotinylated in bacterial, yeast, insect, and mammalian cells (7–11). Biotinylation can occur either by the cell's endogenous protein–biotin ligases or through the coexpression of an exogenous biotin ligase, in most cases that of the bacterial BirA enzyme. These tags, however, are large in size (at least 63 aa) and may thus affect the structure of the proteins they are fused to. In addition, that these tags can be recognized by endogenous enzymes excludes applications where biotinylation of the tagged protein may need to be regulated. Furthermore, biotinylation using these tags is not very efficient, particularly in mammalian cells (10, 11). Another approach, so far only demonstrated in bacteria (12–15), utilizes small (<23-aa) artificial tags that have been selected through multiple rounds of screening combinatorial peptide libraries for specific biotinylation by BirA biotin ligase (16). These tags have been shown to be biotinylated *in vitro* with kinetics comparable to those of natural biotin acceptor sequences (17) and may thus serve as excellent substrates for efficient biotinylation in cells by coexpressed biotin ligases.

Simple generic affinity purification methodologies have been increasingly applied for the purposes of large-scale proteomic studies, particularly in yeast (18, 19). However, these often use reagents of variable affinities (e.g., antibodies) that increase background and/or intermediate steps (e.g., prepurification or affinity tag removal) that restrict their simple application in more complex protein sources such as mammalian cells. Considering the advantages of biotinylation, we explored the feasibility of using it for the simple high-affinity one-step purification of tagged proteins from mammalian cells. To these ends, we coexpressed bacterial BirA biotin ligase and hematopoietic transcription factors tagged by an N-terminal fusion of a small artificial peptide previously shown to be biotinylated by BirA (15–17). We show that tagged proteins can be very efficiently and specifically biotinylated in mammalian cells and transgenic mice and can be efficiently purified in a single step by binding to streptavidin beads.

Experimental Procedures

Constructs. The coding region of the *Escherichia coli* birA biotin–protein ligase gene (20) was cloned from genomic DNA by PCR as an ≈1-kb fragment by using Deep-VENT DNA polymerase (New England Biolabs) and verified by sequencing. The birA gene was recloned into the *Bgl*II site of the erythroid expression vector pEV-puromycin, which consists of the human β-globin Locus Control Region (miniLCR), the human β-globin promoter, and the β-globin second intron (21). GATA-1 cDNA cloned in pEV-Neomycin was N-terminally tagged by introducing into the *Nco*I site at the start codon, an oligonucleotide linker with *Nco*I overhangs

Abbreviations: MEL, mouse erythroleukemic; HRP, horseradish peroxidase; EKLF, erythroid Krüppel-like factor.

‡To whom correspondence should be addressed. E-mail: i.strouboulis@erasmusmc.nl.

www.pnas.org/cgi/doi/10.1073/pnas.1332608100

coding for the 23-aa biotinylation tag (16). Tagged erythroid Krüppel-like factor (EKLF) cDNA was constructed in pBluescript by cloning sequentially into the *Nco*I site start codon firstly an oligonucleotide coding for three copies of the hemagglutinin tag, followed by an oligonucleotide coding for the 23-aa biotinylation tag. Tagged EKLF was then subcloned into pEV-Neomycin.

Mouse Erythroleukemic (MEL) Cell Transfections. MEL cells (22) were initially electroporated with linearized BirA/pEV, and stable clones were selected under puromycin. Clones were screened for BirA RNA expression by Northern blot analysis. A selected BirA/pEV MEL clone was transfected with tagged GATA-1/pEV, and stable clones were double selected for puromycin and neomycin (Invitrogen).

Nuclear Extract Preparation. MEL cells cultured in DMEM supplemented with 10% FCS at 37°C were induced to differentiate into mature erythroblasts with 2% DMSO for at least 3 days (22). Cells were harvested by centrifugation at 640 × *g* and washed once with cold PBS. The cell pellet was resuspended in 2.2 M sucrose in 10 mM Hepes, pH 7.9/25 mM KCl/0.15 mM Spermine/0.5 mM Spermidine/1 mM EDTA and incubated for 20 min. Cells were lysed with a blender, and lysis was checked under a light microscope by staining nuclei with Unna (Methylgreen-Pyronin). Nuclei were pelleted by ultracentrifugation at 141,000 × *g* for 2 h at 4°C, resuspended in lysis buffer (10 mM Hepes, pH 7.9/100 mM KCl/3 mM $MgCl_2$/0.1 mM EDTA/20% glycerol), and extracted by dropwise addition of 3.3 M KCl until the final concentration was ≈400 mM. Insoluble material was removed by ultracentrifugation at 300,000 × *g* for 1 h at 4°C. Nuclear extracts were aliquoted and stored at −70°C.

Immunoblot Analysis. For analysis of GATA-1 expression and *in vivo* biotinylation, nuclear extracts (1–2.5 μg/lane) were resolved by SDS/PAGE in an 8% gel and blotted onto ProTran nitrocellulose membrane (Schleicher & Schuell) by using standard procedures. Filters were blocked for 1 h in 5% BSA/1× TBS/0.2% Tween-20 and incubated for 1 h at room temperature with anti-GATA-1 N6 rat monoclonal antibody (Santa Cruz Biotechnology, dilution 1:5,000) or streptavidin–horseradish peroxidase conjugate (HRP) (NEN, dilution 1:10,000). Filters were washed in 1× TBS/0.5M NaCl/0.3% Triton X-100 and incubated in secondary rabbit anti-rat antibody (DAKO, dilution 1:3,000), as above. No secondary antibody step was required after incubating the filters with streptavidin–HRP. Filters were developed by using enhanced chemiluminescence (Amersham Pharmacia).

Binding to Streptavidin Beads. Paramagnetic streptavidin beads [Dynabeads M-280, Dynal (Great Neck, NY)] were blocked by washing three times in TBS with 200 ng/μl purified chicken serum albumin (Sigma-Aldrich). We used ≈20 μl of beads per 1 mg of nuclear extract. Binding was done in 1× TBS/0.3% Nonidet P-40 at 4°C for 1 h to overnight on a rocking platform, followed by six washes in binding solution at room temperature. Bound material was eluted by boiling for 5 min in Laemmli protein sample loading buffer and analyzed by immunoblotting as above.

MS. Proteins eluted from the beads after binding of the BirA nuclear extract were separated by SDS/PAGE electrophoresis on an 8% polyacrylamide gel and stained with Colloidal blue (Invitrogen). The entire lane was cut out and divided into at least 20 gel plugs, which were each further reduced to 1 mm^3 gel pieces and dried by using 100% acetonitrile (Fluka) in 60% acetonitrile pretreated tubes (Bioquote, York, U.K.). Proteins were in-gel-digested by using modified trypsin (Roche Diagnostics) in 50 mM ammonium bicarbonate. Digests were analyzed by nanoflow liquid chromatography–tandem MS by using an electrospray ionization quadrupole time-of-flight mass spectrometer (Q-Tof, Micromass, Manchester, U.K.) operating in positive ion mode. A nanoLC system was coupled to the Q-Tof essentially as described (23). Peptide mixtures were delivered to the system by using a Famos autosampler (LC Packings, Amsterdam) at 3 μl/min and trapped on an Aqua C18RP column [Phenomenex (Belmont, CA); column dimensions 1 cm × 100 μm ID, packed in-house]. After flow splitting down to 150–200 nl/min, peptides were transferred to the analytical column (PepMap, LC Packings; column dimensions 25 cm × 50 μm ID, packed in-house) in a gradient of acetonitrile (1% per min). Fragmentation of eluting peptides was performed in data-dependent mode, and mass spectra were acquired in full-scan mode. Database searches were performed by using MASCOT (www.matrixscience.com).

Chromatin Pull-Down Assays. MEL cells were treated with 1% formaldehyde in 40 mM Hepes, pH 7.9, at room temperature for 10 min, followed by quenching with 0.125 M glycine. Crosslinked chromatin was fragmented by sonication [10 × 30-s bursts at amplitude 5, followed by 10 × 30-s bursts at amplitude 8 by using a Sanyo Soniprep (Loughborough, U.K.) 150 sonicator] and aliquots corresponding to 4–10 OD^{260} units were snap-frozen. Pull-downs were carried out by incubating an aliquot of crosslinked chromatin for at least 1 h at 4°C with 20 μl of streptavidin-coated Dynabeads, preblocked with 1 mg/ml BSA and 0.4 mg/ml sonicated salmon sperm DNA. Washes were done according to standard protocols (www.upstate.com/misc/protocols.asp). Elution of bound material and reversal of crosslinks was done in 1% SDS in TE buffer (10 mM Tris/1 mM EDTA, pH 8.0) by overnight incubation at 60°C with shaking. DNA was recovered after deproteinization, and aliquots of pulled-down DNA were assayed by PCR using primers against the erythroid mouse βmaj globin promoter as described in *Supporting Materials and Methods*, which is published as supporting information on the PNAS web site, www.pnas.org.

Transgenic Mice. DNA fragments containing the tagged EKLF, and BirA erythroid expression cassettes were released from prokaryotic vector sequences by double digestion with Aat II/Asp 718. The DNA was gel-purified by using Gelase (Epicentre Technologies, Madison, WI) and prepared for microinjection by using EluTip (Schleicher & Schuell) according to the manufacturers' instructions. Microinjection into mouse fertilized eggs was carried out according to standard procedures (24). Screening of mice for transgenics was carried out by Southern blot analysis by using the BirA cDNA and the human β-globin intron II as probes.

Results

Efficient Biotinylation of Tagged GATA-1 in Transfected Cells. The scheme for the specific *in vivo* biotinylation of tagged proteins in mammalian cells is outlined in Fig. 1*A*. According to this procedure, a small (23 aa) peptide tag is fused to the protein of interest and coexpressed in cells together with BirA, a bacterial protein–biotin ligase (20). The peptide tag used has been previously isolated from a synthetic peptide library screened for BirA-mediated biotinylation, which occurs specifically at the lysine residue of the tag (16). Protein database searches have identified no naturally occurring proteins that possess a sequence motif similar to that of the peptide tag.

The protein we tagged in testing this system was the essential murine hematopoietic transcription factor GATA-1 (for review, see ref. 25). The tag was fused N-terminally to GATA-1 and expressed under the control of a human β-globin expression cassette in MEL cells, which can be induced to undergo terminal erythroid cell differentiation (22). BirA was also cloned and expressed in MEL cells by using the human globin expression cassette. Clones corresponding to single and double stable transfectants for tagged GATA-1 and BirA were isolated and initially screened for expression of both constructs. In the case of GATA-1, Western blot analysis using a GATA-1 antibody detects the slower-migrating tagged protein as well as the endogenous GATA-1 in nuclear

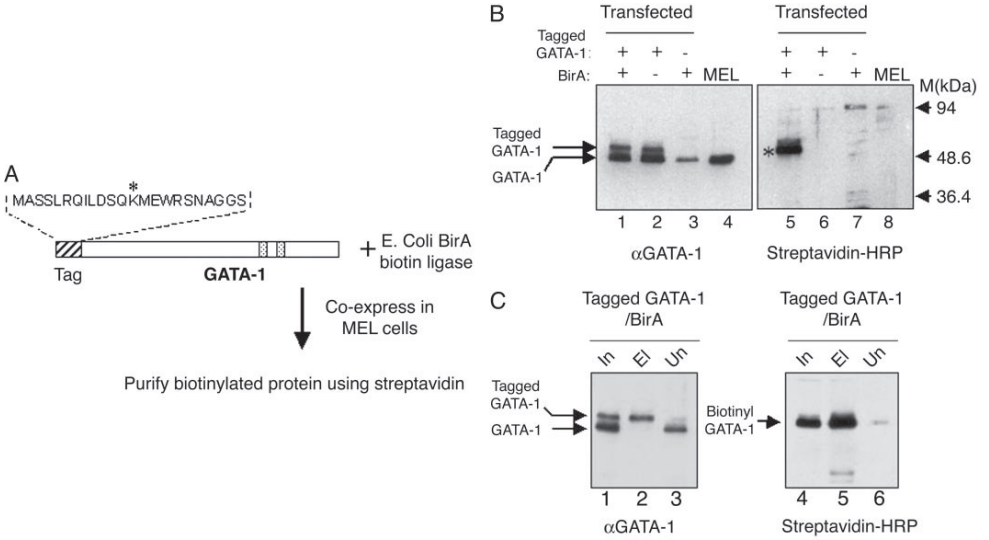


Fig. 1. (*A*) Scheme for the specific biotinylation of tagged GATA-1 by BirA biotin ligase in MEL cells. The sequence of the 23-aa peptide tag fused to the N terminus of GATA-1 is shown. The asterisk indicates the lysine residue that becomes specifically biotinylated by BirA. Speckled boxes indicate the positions of the two GATA-1 Zinc-fingers. Tagged GATA-1 and BirA were cloned separately in a mammalian erythroid expression cassette and coexpressed in MEL cells. (*B*) Biotinylation of tagged GATA-1 in MEL cells. (*Left*) Western blot with an anti-GATA-1 antibody to detect endogenous and tagged GATA-1 proteins. (*Right*) Western blot of the same extracts with streptavidin–HRP conjugate to detect biotinylated GATA-1. Nuclear extracts (5 μg per lane) from the double transfectants (lanes 1 and 5) and single transfectants (lanes 2, 3, 6, and 7) for tagged GATA-1 and Bir A were tested. Lanes 4 and 8, nuclear extract from nontransfected MEL cells. Biotinylated GATA-1 (asterisk) is clearly visible in only in the lane of the double transfected cells. Also indicated is the low background detected by streptavidin in MEL nuclear extracts from cells expressing only BirA (*Right*, lane 7). (*C*) Efficiency of GATA-1 biotinylation and binding to streptavidin beads. (*Left*) Western blot using anti-GATA-1 antibody to detect binding of tagged GATA-1 to streptavidin beads (lane 2; starting material for the binding was 2.5 times the amount of nuclear extract shown in the input lane). Input and unbound material are shown in lanes 1 and 3. (*Right*) The same filter stripped and reprobed with streptavidin–HRP to detect the binding of biotinylated GATA-1 to streptavidin beads (lane 5). Lane 6 shows that very little tagged GATA-1 remains unbound by streptavidin. In this binding experiment, the beads were washed under stringent conditions (0.5 M NaCl/0.3% Triton X-100 in PBS). In, input (nuclear extract); El, eluted material; Un, unbound material.

extracts from transfected cells (Fig. 1*B*, lanes 1 and 2). Expression of BirA was analyzed at the RNA level (data not shown).

We next tested, on selected stable transfectants, whether the tagged GATA-1 protein was biotinylated by BirA. Assaying nuclear extracts from MEL cell clones using a streptavidin–HRP conjugate showed a robust signal corresponding to the tagged GATA-1 protein, detectable only in the lane of the tagged GATA-1/BirA double transfectant (Fig. 1*B*, lane 5). No biotinylation of tagged GATA-1 is visible in the absence of BirA (Fig. 1*B*, lane 6). These findings confirm that BirA protein is synthesized in an active form in transfected MEL cells. In addition, very little nonspecific background biotinylation is observed in MEL cell nuclear extracts expressing only BirA (Fig. 1*B*, lane 7). We therefore conclude that expression of bacterial BirA protein–biotin ligase in MEL cells can specifically biotinylate a mammalian transcription factor bearing a unique peptide tag.

We next tested the efficiency of biotinylation by binding tagged GATA-1 in crude nuclear extracts to streptavidin paramagnetic Dynabeads (Fig. 1*C*). Analysis of the material eluted from the beads showed that almost all of the tagged GATA-1 protein was bound (compare lane 2 to lanes 1 and 3, Fig. 1*C*). Reprobing the same filter with streptavidin–HRP shows ≈100% efficiency in the biotinylation and capture of tagged GATA-1 by the beads (Fig. 1*C*, lanes 4 and 5). In addition, consistent with what was seen in Fig. 1*B*, there is little background binding of endogenously biotinylated proteins to the beads, as detected by streptavidin–HRP (Fig. 1*C*, lane 5). These data demonstrate that tagged GATA-1 is very efficiently biotinylated and recovered from extracts by streptavidin binding, with negligible background biotinylation.

Single-Step Purification of Biotinylated GATA-1 from Crude Nuclear Extracts by Streptavidin Binding. We also explored the feasibility of a single-step purification in isolating biotinylated GATA-1 from crude nuclear extracts by directly binding to streptavidin beads under moderate stringency (150 mM NaCl/0.3% Nonidet P-40/200 ng/μl chicken serum albumin). We first tried a control binding from 5 mg of crude nuclear extracts from MEL cells expressing only BirA (Fig. 2, lane 4). The eluted material consisted of approximately five strongly stained bands against a backdrop of much fainter bands (Fig. 2, lane 5). We identified the background binding proteins by excising the whole lane from the gel and analyzing it by liquid chromatography–tandem MS. The proteins thus identified are classified in Table 1 according to biological function and cellular compartment, as defined by the Gene Ontology Consortium (www.geneontology.org). The results showed that the most abundant background proteins identified were the naturally biotinylated carboxylases and associated enzymes, which largely coincided with the intensely staining bands. We also found background binding of abundant nuclear proteins involved in mRNA processing, such as splicing factors, as well as of ribosomal proteins. Together, these

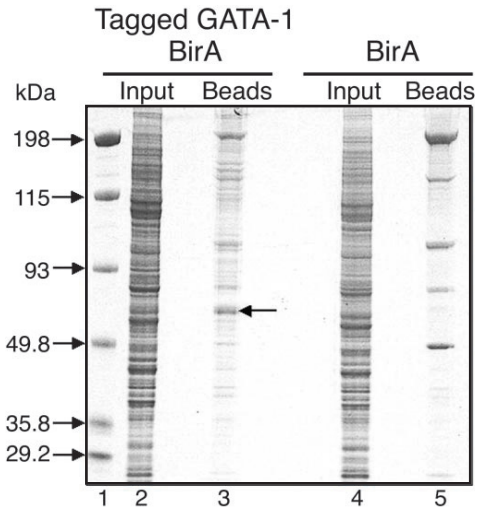


Fig. 2. Colloidal blue-stained gel of a binding experiment of crude nuclear extracts to streptavidin beads. Lane 1, marker (M). Lane 2, input nuclear extract from tagged GATA-1/BirA double transfected cells (≈12 μg). Lane 3, proteins eluted after direct binding to streptavidin beads of ≈5 mg of crude nuclear extracts from tagged GATA-1/BirA transfected cells. Lane 4, input nuclear extract from tagged GATA-1/BirA transfected cells. Lane 5, proteins eluted after binding to streptavidin beads ≈5 mg of nuclear extract from BirA transfected cells. Arrow in lane 3 indicates protein band containing purified biotinylated GATA-1, as determined by MS.

three classes of proteins accounted for >80% of background binding under the conditions used (Table 1). It is notable that very few peptides were identified as corresponding to factors involved in transcriptional regulation (Table 1). We conclude that background binding to streptavidin beads is mainly due to endogenous biotinylated proteins as well as abundant nuclear factors involved in mRNA processing and ribosome synthesis and assembly, with very little nonspecific binding of factors involved in transcriptional regulation/activation.

We next tried binding a similar amount of crude nuclear extract (5 mg) from tagged GATA-1/BirA double transfectants, under the same conditions (Fig. 2, lane 2). The staining pattern of the lane with the eluted material (Fig. 2, lane 3) was significantly different to that observed with the background binding in lane 5, indicating a significant enrichment in proteins coeluting with tagged GATA-1 (Fig. 2, lane 3 vs. lane 5). Biotinylated GATA-1 and coeluting proteins may dilute out or compete for binding with the nonspecific proteins observed in lane 5. The protein enrichment visible in lane 3 may thus correspond to copurified GATA-1-interacting proteins. In the eluted material, we also observed a strongly stained band migrating with a size similar to that expected for tagged GATA-1 (Fig. 2, arrow, lane 3). We confirmed the presence of GATA-1 in this band by gel excision and MS. The detailed analysis of all proteins copurifying with tagged GATA-1 will be published elsewhere (P.R., F.G., and J.S., unpublished results). Taken together, these data demonstrate the quantitative biotinylation of tagged GATA-1 in MEL cells and its efficient purification from crude protein extracts in a single-step procedure by direct binding to streptavidin beads, with few background binding bands corresponding primarily to endogenously biotinylated proteins and easily identifiable abundant nuclear/nucleolar proteins.

Biotinylation Does Not Affect the Protein-Interacting or DNA-Binding Properties of GATA-1. Because it is possible that addition of the peptide tag and/or biotinylation may affect the properties of the

Table 1. Summary of background binding proteins

Biological process	Cellular component	Total no. of peptides	Remarks
Metabolism	Mitochondrion	>180	Carboxylases, acyltransferases, etc.
mRNA processing	Nucleus	120	Splicing factors, hnRNPs, ATP-dependent RNA helicases, etc.
Protein biosynthesis	Cytosol	93	Ribosomal proteins, etc.
Receptor activity?	Unknown	17	Single protein: thyroid hormone receptor-associated protein
RNA processing?	Unknown	16	Four proteins with RNA-binding motifs
Cytoskeleton		13	Actin, lamin
Chromatin assembly	Nucleosome	13	Histones
Apoptotic program	Nucleolus	10	Single protein: apoptotic chromatin condensation inducer in the nucleus
No information	Unknown	10	Bcl-2-associated transcription factor
Ribosome biogenesis	Nucleolus	9	SnoRNA-binding proteins, fibrillarin
Protein targeting	Nucleus	5	Single protein: cdc5-like
Electron transport	Microsomes	5	Single protein: oxidoreductase
No information	Nucleolar	4	myb-binding protein 1a
Chromatin Modification	Nucleus	3	Single protein: SAP 18
DNA recombination	Nucleus	2	Two proteins: pontin and reptin (RuvB-like)
Transcription regulation	Nucleus	2	Two proteins: Y box transcription factor 1 (1 peptide); thymocyte selection associated HMG box (1 peptide)
Total no. of peptides	>500		
Gel slices	~21		
Protein molecular mass range, kDa	15–274		

Proteins identified by MS were classified according to biological function and cellular localization, with the criteria used by the Gene Ontology Consortium. Also shown is the number of peptides identified for each class of proteins.

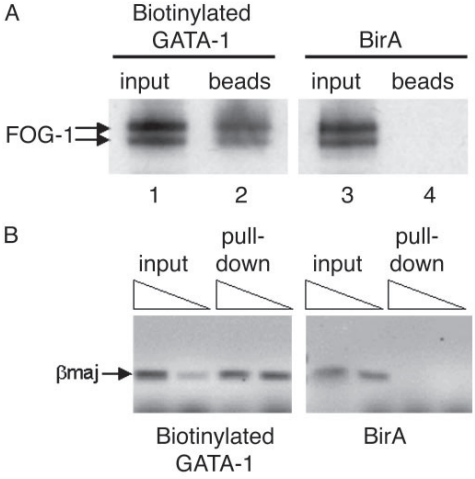


Fig. 3. (*A*) Binding biotinylated GATA-1 to streptavidin beads specifically pulls down FOG-1, as detected by Western blotting by using a FOG-1 antibody. By contrast, FOG-1 cannot be pulled down by streptavidin in nuclear extracts expressing biotinylated expressing BirA only (*Right*). FOG-1 is detected as a doublet (26). (*B*) Streptavidin pull-down of βmaj globin promoter sequences from crosslinked chromatin from MEL cells expressing biotinylated GATA-1 (*Left*) or BirA only (*Right*). Triangles indicate increasing amounts of pulled-down crosslinked chromatin used as template in PCR reactions in detecting amplification of the βmaj sequences. Specific enrichment for βmaj sequences is observed in pulled-down chromatin from cells expressing biotinylated GATA-1 but not from cells expressing BirA only.

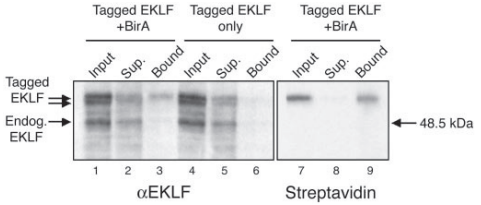


Fig. 4. Specific biotinylation of tagged EKLF in transgenic mouse embryos. Nuclear extracts from the fetal liver of 13.5-days postcoitum embryos from a tagged EKLF/BirA double transgenic line (lanes 1–3 and 7–9) and from a tagged EKLF transgenic line (lanes 4–6) were bound to streptavidin beads. Tagged and biotinylated EKLF in input nuclear extract, unbound material (sup., supernatant), and bound material was detected by an EKLF antibody (*Left*) or by streptavidin–HRP (*Right*). EKLF biotinylation and binding to the beads is detected only in extracts from double transgenic embryos.

tagged protein, we tested whether biotinylated GATA-1 could still undergo protein–protein interactions with a known GATA-1 partner, such as FOG-1 (26), and whether it could bind *in vivo* to known GATA-1 gene targets such as the mouse βmaj globin promoter. We carried out a pull-down experiment of biotinylated GATA-1 by binding nuclear extracts to streptavidin beads and tested whether FOG-1 was also copurified. We found FOG-1 to be pulled down from extracts expressing biotinylated GATA-1 but not from extracts expressing BirA only (Fig. 3*A*, lanes 2 and 4). We have also found FOG-1 to be copurifying with biotinylated GATA-1 by MS. We also carried out a chromatin pull-down (ChIP) experiment in which sonicated chromatin from formaldehyde-crosslinked MEL cells was incubated with streptavidin beads, followed by elution of the bound material and recovery of the pulled-down DNA. Using primers specific for the βmaj promoter, we found enrichment for these sequences in the DNA pulled down from the chromatin of cells expressing biotinylated GATA-1 but not from cells expressing BirA (Fig. 3*B*). We have also found that biotinylation does not affect the subnuclear distribution normally observed with GATA-1 (Fig. 5, which is published as supporting information on the PNAS web site; ref. 27) and that biotinylated GATA-1 displays an identical biochemical fractionation profile as endogenous GATA-1 in MEL cell nuclear extracts (data not shown). Taken together, these results provide strong evidence that the properties of GATA-1 are not affected by biotinylation tagging. In addition, these data also demonstrate the application of biotinylation tagging as an alternative to antibodies in methods involving an affinity purification step, such as protein pull-downs or a ChIP assay.

Biotinylation Tagging in Transgenic Mice. The ability to directly isolate the biotin-tagged protein from crude extracts in a single step raises the prospect of using this approach in the purification of tagged proteins from limiting sources such as mouse tissues. We therefore tested whether BirA-mediated biotinylation tagging would also work *in vivo* in transgenic mice. Because GATA-1 overexpression leads to embryonic lethality in mice (28), we tested this approach by tagging the essential erythropoietic transcription factor EKLF (29). Mouse EKLF cDNA was tagged with the biotinylation tag and a double hemagglutinin epitope and microinjected in mouse eggs to establish transgenic mouse lines. Similarly, transgenic mouse lines were also established by microinjecting the BirA/erythroid expression cassette construct. Transgenic mouse lines with detectable expression of tagged EKLF and BirA in erythroid cells were selected and crossbred. *In vivo* biotinylation of tagged EKLF was assessed in nuclear extracts prepared from the fetal livers of 13.5-day postcoitum embryos. Western blot analysis with an anti-EKLF antibody detected endogenous EKLF as well as tagged EKLF, which was visualized as a doublet (Fig. 4, lane 1). Binding of the fetal liver nuclear extracts to streptavidin beads shows that only the top band in the doublet is retained, suggesting that it is biotinylated (Fig. 4, lanes 2 and 3). This observation is confirmed by probing the same blot with streptavidin–HRP, which detects only a single band (Fig. 4, lanes 7 and 9). The doublet detected by the EKLF antibody is most likely due to the differential utilization of translation initiation codons at the N-terminally fused biotinylation tag (top band) and at the double hemagglutinin epitope present immediately downstream, which also contains an initiation codon (bottom band). As a result, only the top band bearing the biotinylation tag serves as a substrate for BirA, thus further demonstrating the *in vivo* specificity of this approach. The binding of nuclear extracts to streptavidin beads also showed that a significant proportion (≈50%) of tagged EKLF is biotinylated in the fetal livers of transgenic mice. We speculate that the difference in biotinylation efficiencies between transfected cells and fetal livers (apart from the use of different fusion proteins) may reflect an *in vivo* limitation in the availability of biotin (biotin is abundant in the FCS used to supplement cell culture media) or a difference in BirA expression levels. These results demonstrate that the specific biotinylation of tagged EKLF by bacterial BirA can also be achieved with high efficiency *in vivo* in transgenic mice.

Discussion

In this paper, we have demonstrated that expression of the bacterial BirA biotin ligase in mammalian cells and transgenic mice leads to the quantitative biotinylation of specific transcription factors bearing a small artificial peptide tag. We also showed that, at least for GATA-1, fusion of the peptide tag and specific biotinylation do not interfere with the protein's properties. We demonstrated that

biotinylation can be effectively used for the single-step affinity purification of the tagged protein by binding to streptavidin beads.

Small (<23-aa) biotinylation tags have been previously obtained through multiple rounds of screening combinatorial peptide libraries for specific biotinylation by the BirA biotin ligase (16) and have been shown to be biotinylated at rates similar to those of naturally occurring substrates (17). Such peptide tags have been subsequently used for the specific biotinylation of fusion proteins in *E. coli* (12–15). Our work shows the utility of such a small tag for the efficient BirA-mediated biotinylation of specific fusion proteins in mammalian cells. Larger (>63-aa) tags derived from biotin acceptor domains present in naturally biotinylated proteins have been previously used in biotinylating fusion proteins in mammalian cells (10, 11). However, there are obvious advantages in using smaller artificial tags. First, small tags are much less likely to affect the structure and thus the properties of the fusion protein *in vivo*. It should be noted that tag size can be reduced even further to ≈14 aa without compromising biotinylation efficiencies (17). Second, the use of small tags avoids the extra complication of tag removal by proteolytic cleavage that is often necessary with larger tags (e.g., TAP tag, ref. 4). Third, artificial tags are unlikely to be recognized and biotinylated by endogenous biotin ligases. Indeed, in our assays, we observed no biotinylation (in nuclear extracts) of the fusion proteins in the absence of BirA. It is also important to note that expression of BirA in mammalian cells did not lead to an increase in nonspecific background biotinylation in nuclear extracts. We conclude that, using this approach, biotinylation of tagged proteins in mammalian cells is a highly specific tightly regulated process that occurs only in the presence of BirA.

The demonstration that a small peptide tag can be efficiently biotinylated in mammalian cells provides a very useful tool with a number of advantages for protein purification. First, we showed that the biotinylated protein can be efficiently purified directly from a crude extract in a single-step procedure, whereas most commonly used affinity tags require a number of purification steps before the affinity-binding step. Second, there is low specific background binding primarily due to the small number of endogenous naturally biotinylated proteins. Under the mild conditions we used in our purification procedure, we observed five strongly staining background protein bands binding to streptavidin that corresponded to naturally biotinylated proteins, such as carboxylases, as well binding by abundant nuclear proteins such as splicing factors. Third, the very strong binding of biotin to avidin/streptavidin offers another advantage in that it allows increasingly higher stringencies to be used during purification without fear of early elution of the tagged protein. Not only may this higher stringency further reduce nonspecific background binding (e.g., by splicing factors), but it can also serve as a measure of the strength of interactions of proteins copurifying with the biotin-tagged protein. Indeed, a systematic comparison of different tags for purifying a rat neurotensin receptor in *E. coli* showed biotinylation to be the best approach in terms of efficiency and purity (12). Furthermore, the inclusion of specific protease cleavage sites downstream of the biotinylation tag may be used to specifically cleave and elute the tagged protein complex from streptavidin, while leaving the endogenous biotinylated background proteins still bound. Fourth, another advantage is the availability of modified forms of avidin with lower binding affinities for biotin (e.g., monomeric avidin with a K_d of $\approx 5 \times 10^{-8}$ M), which allow elution of bound biotinylated proteins under native conditions. This option offers the possibilities of purifying active tagged protein complexes as well as the better resolution and identification of purified proteins by 2D gel electrophoresis. Finally, it is very significant that the BirA-mediated specific biotinylation can also be efficiently achieved in transgenic animals, thus raising the prospect of using limiting animal tissues for protein purification by using high-affinity interaction with streptavidin. This is, in principle, a major advantage over other tagging approaches, such as the highly efficient TAP tag approach (4), which also does not require prepurification steps. The TAP tag includes two Ig-binding domains from the *Stahylococcus aureus* protein A, which may be a problem when expressed in transgenic animals that naturally express antibodies.

In conclusion, the efficient biotinylation of specific proteins in mammalian cells demonstrated here raises the prospect of applying the advantages and flexibility of the well-developed biotin/avidin technology in the single-step purification and characterization of mammalian protein complexes.

We are grateful to Dr. Stuart Orkin (Harvard Medical School) for the FOG-1 antibody, Dr. Mariette van de Corput for help with immunofluorescence, and John Kong-a-San for mouse transgenesis. This work was supported by the European Union (to J.S., P.R., and E.K.) and the Netherlands Organization for Scientific Research (NWO).

1. Lamond, A. I. & Mann, M. (1997) *Trends Cell Biol.* **7,** 139–142.
2. Kumar, A. & Snyder, M. (2002) *Nature* **415,** 123–124.
3. Jarvik, J. W. & Telmer, C. A. (1998) *Annu. Rev. Genet.* **32,** 601–618.
4. Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M. & Seraphin, B. (1999) *Nat. Biotechnol.* **17,** 1030–1032.
5. Chapman-Smith, A. & Cronan, J. E., Jr. (1999) *Trends Biochem. Sci.* **24,** 359–363.
6. Savage, M. D., Mattson, G., Desai, S., Nielander, G. W., Morgensen, S. & Conklin, E. J. (1992) *Avidin-Biotin Chemistry: A Handbook* (Pierce Chemical Company, Rockford, IL).
7. Cronan, J. E., Jr. (1990) *J. Biol. Chem.* **265,** 10327–10333.
8. Duffy, S., Tsao, K. L. & Waugh, D. S. (1998) *Anal. Biochem.* **262,** 122–1228.
9. Cronan, J. E., Jr., & Reed, K. E. (2000) *Methods Enzymol.* **326,** 440–458.
10. Parrott, M. B. & Barry, M. A. (2000) *Mol. Ther.* **1,** 96–104.
11. Parrott, M. B. & Barry, M. A. (2001) *Biochem. Biophys. Res. Commun.* **281,** 993–1000.
12. Tucker, J. & Grisshammer, R. (1996) *Biochem. J.* **317,** 891–899.
13. Tatsumi, H., Fukuda, S., Kikuchi, M. & Koyama, Y. (1996) *Anal. Biochem.* **243,** 176–180.
14. Tsao, K. L., DeBarbieri, B., Michel, H. & Waugh, D. S. (1996) *Gene* **169,** 59–64.
15. Smith, P. A., Tripp, B. C., DiBlasio-Smith, E. A., Lu, Z., LaVallie, E. R. & McCoy, J. M. (1998) *Nucleic Acids Res.* **26,** 1414–1420.
16. Schatz, P. J. (1993) *Biotechnology* **11,** 1138–1143.
17. Beckett, D., Kovaleva, E. & Schatz, P. J. (1999) *Protein Sci.* **8,** 921–929.
18. Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, *et al.* (2002) *Nature* **415,** 141–147.
19. Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., *et al.* (2002) *Nature*, **415,** 180–183.
20. Howard, P. K., Shaw, J. & Otsuka, A. J. (1985) *Gene* **35,** 321–331.
21. Needham, M., Gooding, C., Hudson, K., Antoniou, M., Grosveld, F. & Hollis, M. (1992) *Nucleic Acids Res.*, **20,** 997–1003.
22. Antoniou, M. (1991) in *Methods in Molecular Biology: Gene Transfer and Expression Protocols*, ed. Murray E. J. (Humana, Clifton, NJ), Vol. 7, pp. 421–434.
23. Meiring, H. D., van der Heeft, E., ten Hove, G. J. & de Jong, A. P. J. M. (2002) *J. Sep. Sci.* **25,** 557–568.
24. Hogan, B., Beddington, R., Costantini, F. & Lacey, E. (1994) *Manipulating the Mouse Embryo: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
25. Cantor, A. B. & Orkin, S. H. (2002) *Oncogene* **21,** 3368–3376.
26. Tsang, A. P., Visvader, J. E., Turner, C. A., Fujiwara, Y., Yu, C., Weiss, M. J., Crossley, M. & Orkin, S. H. (1997) *Cell* **90,** 109–119.
27. Elefanty, A. G., Antoniou, M., Custodio, N., Carmo-Fonseca, M. & Grosveld, F. G. (1996) *EMBO J.* **15,** 319–333.
28. Whyatt, D., Lindeboom, F., Karis, A., Ferreira, R., Milot, E., Hendriks, R., de Bruijn, M., Langeveld, A., Gribnau, J., Grosveld, F. & Philipsen, S. (2000) *Nature* **406,** 519–524.
29. Bieker, J. J. (2001) *J. Biol. Chem.* **276,** 34355–34358.

Chapter 3:

GATA-1 forms distinct activating and repressive complexes in erythroid cells

Patrick Rodriguez[1], Edgar Bonte[1], Jeroen Krijgsveld[2], Katarzyna Kolodziej[1], Boris Guyot[3], Albert J.R. Heck [2], Paresh Vyas[3], Ernie de Boer[1], Frank Grosveld[1] and John Strouboulis[1]

Part of this Chapter was published in EMBO, 2005, Jul 6; (24): 2354-66.

[1] Department of Cell Biology, Erasmus University Medical Center, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
[2] Department of Biomolecular Mass Spectrometry, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, The Netherlands
[3] Department of Haematology, The Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DU, United Kingdom

GATA-1 forms distinct activating and repressive complexes in erythroid cells

Abstract

GATA-1 is a key transcription factor essential for the differentiation of the erythroid, megakaryocytic and eosinophilic lineages. In erythropoiesis, GATA-1 functions involve lineage-specific gene activation, as well as repression, suppression of cell proliferation and anti-apoptotic functions. GATA-1 was reported to interact with other transcription factors, such as FOG-1, TAL-1 (and its co-factors Ldb1, E2A and Lmo2) and Sp1, and also with CBP/p300 and the SWI/SNF chromatin-remodeling complex in vitro. Nevertheless, the basis of all of GATA-1's multiple functions in erythropoiesis remains unclear. Using an in vivo biotinylation tagging approach we isolated and characterized GATA-1 complex(es) from red cells. Biotinylated GATA-1 from nuclear extracts was bound directly to streptavidin beads and co-purifying proteins were identified by mass spectrometry. In addition to the known GATA-1-interacting transcription factors FOG-1, TAL-1 and Ldb1, we describe novel interactions with the essential hematopoietic transcription factor Gfi-1b and the chromatin remodeling complexes MeCP1 and ACF/WCRF. The most abundant interactions are with FOG-1 and MeCP1 where FOG-1 is essential for the interaction. We show these complexes to be distinct with interactions differentially mediated via the zinc-fingers of GATA-1. Our findings on GATA-1 complexes provide important insight by revealing novel GATA-1 partners.

Introduction

Hematopoiesis is the process responsible for the generation of the different blood cells and has often served as a model for understanding the basis of cellular commitment and differentiation. Hematopoiesis proceeds from multipotential, self-renewing hematopoietic stem cells (HSCs) to an increasing number of progressively more lineage restricted cells, each lineage giving rise to one type of mature blood cell. The differentiation of distinct hematopoietic lineages is the result of specific transcription programs regulated by a number of essential transcription factors[1,2].

An example of a key hematopoietic transcription factor is GATA-1, the prototypical member of the GATA family of zinc-finger transcription factors[3,4]. GATA-1 is essential for the differentiation of the erythroid, megakaryocytic and eosinophilic lineages[5-7] and is also expressed in mast cells and in some multipotential hematopoietic precursor cells. GATA-1 null embryonic stem cells fail to contribute to the generation of mature red blood cells in chimeric mice[5]. GATA-1 null mice die from severe anemia at day 11.5 dpc because erythroid cells fail to mature beyond the proerythroblast stage and die of apoptosis[8]. Importantly, the forced ectopic expression of GATA-1 in an early myeloid cell line promotes megakaryocyte differentiation[9]. Similarly, enforced expression of GATA-1 in myelomonocytic and lymphoid precursor cells reprograms them to erythroid, megakaryocytic and eosinophilic fates, with the concomitant repression and upregulation of respective lineage-specific marker genes[10-12]. These observations suggest that GATA-1 is capable of dictating expression of a specific genetic program in a cell where it is not normally expressed. Thus, GATA-1 acts as a key regulator in the specification of the aforementioned lineages, a role that must involve the activation as well as repression of lineage-restricted transcriptional programs.

Previous studies have identified several protein partners of GATA-1. The most prominent amongst the GATA-1 interacting partners is the zinc-finger protein FOG-1, which was originally identified in a yeast two-hybrid screen using GATA-1 as bait[13]. A direct interaction between the two factors is required for erythroid differentiation since a mutant of GATA-1 unable to bind to FOG-1 fails to support terminal erythroid differentiation[14]. FOG-1 is co-expressed with GATA-1 in fetal liver, mast cells and megakaryocytes. The overall GATA-1 and FOG-1 knockout phenotypes are very similar in that they both result in a block in erythropoiesis and early embryonic lethality due to anemia[15]. Ablation of GATA-1 leads to a block of erythroid differentiation at the proerythroblast stage and apoptosis, whereas ablation of FOG-1 leads to a marked but partial block at the same stage. In megakaryopoiesis, FOG-1$^{-/-}$ embryos exhibit a complete failure of development whereas megakaryocytes lacking GATA-1 are increased in number but have a block in their development at mid-maturation.

In addition, in erythroid cells GATA-1 is part of a pentameric protein complex that includes the essential hematopoietic transcription factor TAL-1 and its associated co-factors Ldb1, LMO2 and E2A[16]. This complex was first shown to bind to GATA-E-box motifs *in vitro* and has since been implicated in the positive regulation of the expression of a number of erythroid genes, such as protein 4.2, c-kit and glycophorin A and GATA-1 itself[17-19]. In addition, GATA-1 interacts with the erythroid specific transcription factor EKLF, which is essential for the expression of β-globin genes, with PU.1, which is essential for normal myelopoiesis and lymphopoiesis, and also with the ubiquitously expressed transcription factor Sp1 (reviewed by Cantor [2]).

A role for GATA-1 in regulation of chromatin structure was suggested by its interactions with chromatin remodeling/modification proteins, including the CBP/p300 histone

acetyltransferases and the SWI/SNF chromatin remodeling complex *in vitro*[20,21]. Acetylation of histones is associated with an open chromatin configuration and the interaction of GATA-1 with CBP was shown to stimulate transcriptional activation by GATA-1. In addition, the SWI/SNF complex was shown to interact with the erythroid specific transcription factor EKLF in activating the β-globin locus[22]. In the same biochemical studies it was shown that the GATA-1 Zn-fingers also have the capacity to interact with the SWI/SNF complex[21], but so far no functional evidence for these interactions has been demonstrated *in vivo*.

Despite all these data, important questions remain as to how can GATA-1 accommodate all these interactions at the same time in erythroid cells and how these interactions relate to the multiple GATA-1 functions? In addressing these questions, we have undertaken a biotinylation tagging-proteomics approach to characterize GATA-1 complexes from erythroid cells[23]. First, we show that GATA-1 interacts with previously reported interacting proteins thus validating our approach. In addition, we identify novel partners of GATA-1, such as the essential hematopoietic transcription factor Gfi-1b and the chromatin remodeling and modification complexes MeCP1 and ACF/WCRF. Moreover, we show a number of these complexes to represent distinct as well as overlapping GATA-1 interactions in erythroid cells. Using GATA-1 mutants, we show that the GATA-1 zinc fingers differentially mediate these interactions and, lastly, we show that the interaction of GATA-1 with the MeCP1 complex is mediated via FOG-1. Our findings provide an explanation for a number of previous observations regarding GATA-1 functions and protein interactions.

Results

Identification of GATA-1 complexes from erythroid cells by biotinylation tagging and mass spectrometry

GATA-1 was tagged by fusing of a small (23aa) peptide sequence to its N-terminus. This tag is efficiently biotinylated by the bacterial BirA biotin ligase which is co-expressed in stably transfected mouse erythroleukemic (MEL) cells, allowing the single step purification of biotinylated GATA-1 using streptavidin beads under mild conditions[23]. Proteins co-purified with GATA-1 from MEL cells chemically induced to undergo terminal differentiation, were identified by mass spectrometry, classified according to Gene Ontology terms or by BLAST searches and compared to the background (Table 1). Additional experiments employed more stringent conditions and different nuclear extract preparations from induced MEL cells. We rejected proteins that appeared in the background binding experiments[23], or proteins that belonged to a subnuclear compartment from which GATA-1 is excluded, e.g. the nucleolus[24]. Streptavidin pull-downs of nuclear extracts under more stringent conditions (Fig. 1A-C) and immunoprecipitations of induced non-transfected MEL nuclear extracts provided further validation (Fig. 1D-G). The identities of proteins confirmed in this way as co-purifying with GATA-1 are shown in Table 2.

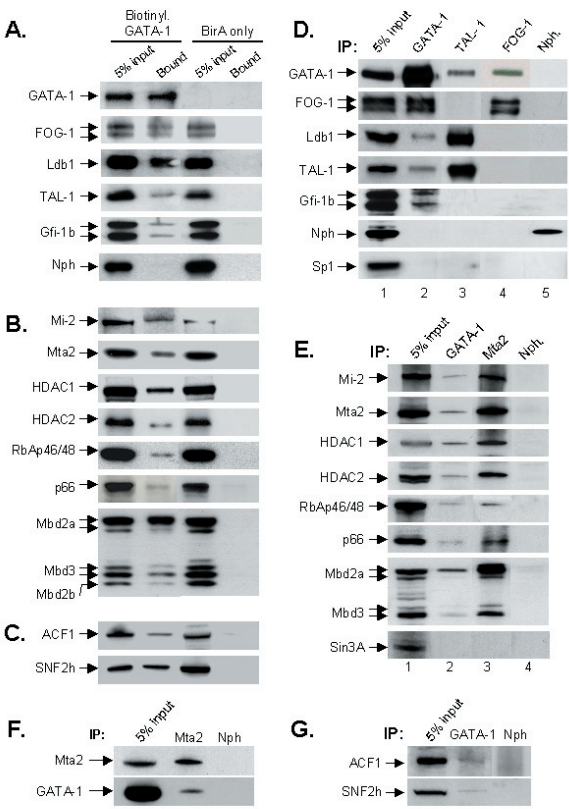


Figure 1: Confirmation by streptavidin pull-downs (A-C) and immunoprecipitations (D-G) of proteins identified co-purifying with GATA-1.

(A) Streptavidin pull-downs of transcription factors. Biotinylated GATA-1 (top panel) is detected by streptavidin-HRP and is absent from the BirA only transfected cells. (B) Pull-downs of the MeCP1 complex. (C) Pull-downs of ISWI-containing complexes. SB: streptavidin-bound. (D) Immunoprecipitations (IP) using antibodies against GATA-1, TAL-1, FOG-1 (lanes 2, 3 and 4, respectively) and nucleophosmin as negative control (lane 5). (E) IP of the MeCP1 complex by antibodies against GATA-1 and MTA2 (lanes 2 and 3) and nucleophosmin (lane 4). (F) GATA-1, can be specifically IP'd by an antibody against MTA2 (G) IP of the ACF/WCRF complex by GATA-1 antibodies. Nuclear extract equivalent to 5% used in each pull-down or IP was loaded as control for input material. IP: immunoprecipitating antibody. Arrows show the detecting.

We found FOG-1, TAL-1 and Ldb1 co-purifying with GATA-1[13,16] thus validating our approach. The Gfi-1b hematopoietic transcription factor was also identified under moderate

GO Biological process	GO Cell. component	Total no. of peptides	Remarks
chromatin modeling/modification transcriptional regulation	nucleus	289	ATPase-dependent chromatin remodeling, HDACs
Unknown		158	putative transcription factors, RNA binding proteins, nucleolar proteins, receptors etc.
mRNA processing	nucleus	139	splicing factors
DNA topological change	nucleus	130	topoisomerases
transcription regulation, DNA dependent	nucleus	105	Transcription factors
Metabolism	mitochondrion	54	carboxylases
protein biosynthesis	cytosol	53	ribosomal proteins
ribosome biogenesis	nucleolus	37	apoptotic chromatin condensation inducer in the nucleus
DNA repair	nucleus	28	XRCC1, Ku autoantigen
nucleolus biogenesis	nucleus	24	
cell proliferation	nucleus	23	Ki-67
DNA replication	nucleus	23	ORC
chromatin assembly	nucleus	21	histones
Cytoskeleton		18	tubulin
structural molecule	nucleus	11	NuMA1
protein targeting	nucleus	7	
Apoptosis	nucleus	4	CDC5
protein chaperone	nucleus	4	
cell cycle	nucleus	3	RCC1
signal transduction		2	GTP binding protein 3
Total # of peptides:		**>1100**	
Gel slices:		**~45**	
MW range:		**15-324kDa**	

Table 1: **Classification by Gene Ontology (GO) criteria of proteins identified by mass spectrometry as co-purifying with GATA-1**

PROTEIN IDENTITY	NUMBER OF PEPTIDES	ADDITIONAL PURIFICATIONS
MeCP1 complex		
Mi-2	82	+
HDAC 1	10	+
HDAC 2	17	+
MTA1	47	+
MTA2	10	+
MTA3	4	-
Mbd2	14	+
Mbd3	9	+
p66	11	+
RbAp46	10	+
RbAp48	8	+
ACF/WCRF complex		
SNF2h	21	+
ACF1	4	-
Transcription factors		
FOG-1 (Hem.)	47	+
TAL-1 (Hem.)	2	+
Gfi-1b (Hem.)	1	-
Ldb1 (Ubiq.)	1	+
DNA repair		
Rfc5	10	+
XRCC1	3	-
Ku70	9	+
PARP	10	+
DNA ligase III-β	1	-
DNA Topological change		
DNA Topo I	34	+
DNA Topo II α	64	+
DNA Topo II β	32	-

Table 2: Proteins specifically co-purifying with biotin tagged GATA-1 as compared to the control purification[23].
A number of these proteins have been validated by immunoprecipitations and other assays (see text). Hem: hematopoietic transcription factors. Ubiq: ubiquitous transcription factors.

stringency conditions and verified by immunoprecipitation (Fig. 1D, lane 2) demonstrating an interaction between the two factors. This is in line with the similarities observed in the Gfi-1b and GATA-1 knockout phenotypes which result in differentiation arrest of the erythroid and megakaryocytic lineages[8,25]. Chromatin remodeling and modification proteins also co-eluted with GATA-1 (Table 1) including the entire MeCP1 complex (Fig. 1). MeCP1 consists of the methyl-DNA binding protein MBD2[26], p66/p68[27] and the multi-subunit Mi-2/NuRD complex containing the nucleosome stimulated Mi-2β ATPase, the histone deacetylases HDAC1 and HDAC2 and other subunits of unknown function. The Mi2/NuRD and MeCP1 complexes are associated with epigenetic mechanisms of repression during development[28], potentially linking the GATA-1 repressive functions to the MeCP1 complex.

The SNF2h and ACF1 members of mammalian ISWI chromatin remodeling complexes also co-purified with GATA-1 (Table 1, Fig. 1C, G). SNF2h, a homologue of the Drosophila

protein ISWI, is the “signature” ATPase of this class of complexes and participates in three distinct complexes in human cells: RSF, hACF/WCRF, hCHRAC[29]. We did not detect by mass spectrometry or immunoprecipitation (not shown) the additional p15 and p17 protein partners present in the hCHRAC complex hence GATA-1 appears to interact with SNF2h/ACF1 in the context of the ACF/WCRF complex[30]. ISWI/SNF2h-containing chromatin remodeling complexes have been associated with gene activation and repression (reviewed by[29]). The interaction between SNF2h and GATA-1 may help to explain the observation that knocking down SNF2h expression in primary hematopoietic progenitor cells blocked erythroid differentiation[31].

Further validation for the GATA-1 interactions was provided by reverse immunoprecipitations using antibodies against TAL-1, FOG-1 (Fig. 1D) or MTA2 (Fig. 1E). TAL-1 antibodies specifically immunoprecipitated GATA-1 and Ldb1 (Fig. 1D), as previously observed[32,33]. LMO2 or E2A were not detected co-purifying with GATA-1 from induced MEL cells, but it cannot be excluded that their absence is due to the very low abundance of the GATA-1/TAL-1/Ldb1/E2A/LMO2 complex (Table 1), in agreement with previous reports of a very small fraction of LMO2 being immunoprecipitated by GATA-1 antibodies[32]. Interestingly, FOG-1 antibodies immunoprecipitated GATA-1 but not TAL-1, Ldb1, or Gfi-1b (Fig. 1D, lane 4). The converse was also true using TAL-1 antibodies (Fig. 1D, lane 3). Thus, GATA-1 interactions with TAL-1, FOG-1 and Gfi-1b are non-overlapping and must thus occur in distinct complexes. GATA-1 was also immunoprecipitated by MTA2 antibodies (Fig. 1F). By contrast, the Sin3A co-repressor, which interacts with HDACs but not in the MeCP1 complex, was not immunoprecipitated by GATA-1 or MTA2 antibodies (Fig. 1E) further supporting the specificity of the GATA-1 interactions with MeCP1.

Other abundant chromatin-associated proteins also co-purified with GATA-1 (Table 1), including topoisomerases and Ku autoantigen or ADP ribosyltransferase (PARP). We tested the association of these proteins with DNA and GATA-1 by treating nuclear extracts with DNase I. In contrast to MTA2, there were no topoisomerase I or PARP co-purifying with GATA-1 following DNase I treatment (Fig. 2). Though it remains formally possible that interactions of GATA-1 with topoisomerase I or PARP are relevant and require DNA, on the basis of our DNase I results and on previous evidence by other groups describing topoisomerases as a common contaminant[34], we did not pursue these further.

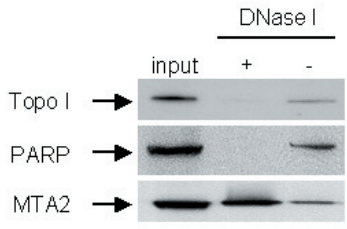


Figure 2: DNase I treatment of nuclear extracts shows the indirect co-purification of abundant chromatin associated proteins with GATA-1. *Nuclear extracts from MEL cells treated or not treated with DNAse I were immunoprecipitated with GATA-1 antibodies. Co-immunoprecipitation of abundant chromatin associated proteins such as topoisomerase I orPARP, identified by mass spectroscopy as co-purifying with GATA-1, was lost upon DNAse I treatment. By contrast, co-immunoprecipitation of MTA2, a member of the MeCP1 complex co-purified with GATA-1, was unaffected by DNase I treatment thus showing a direct interaction with GATA-1.*

GATA-1 and co-purifying proteins interact as large complexes

We also tested whether GATA-1 and interacting proteins share overlapping size-fractionation profiles, as would be predicted for proteins participating in the same multi-protein complexes. Nuclear extracts from induced MEL cells expressing biotinylated GATA-1 were size-

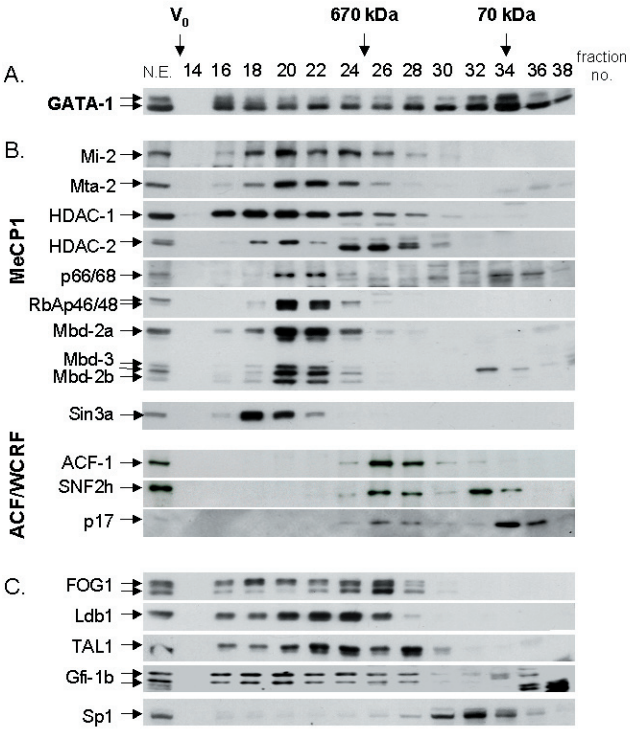


Figure 3: Size-fractionation profiles by Superose 6 column of GATA-1 (panel A), members of the MeCP1 and ACF/WCRF complexes (panel B) and transcription factors (panel C).
Molecular mass markers are indicated on the top. V_0: void volume. N.E.: input nuclear extract.
GATA-1 displays a broad fractionation profile with several peaks. The profile of tagged GATA-1 closely follows that of endogenous GATA-1, is stable in salt concentrations up to 1M and is not dependent on the presence of DNA (data not shown). Members of the MeCP1 complex showed overlapping peaks around fractions 20-22 (e.g. MTA2, RbAp46/48, Mbd2/3) in contrast to Sin3A which peaked around fractions 18-20. SNF2h and ACF1 peaked around fractions 26-28 and are distinctly different from those of the MeCP1 complex. p17 elutes in fractions 34-36, further suggesting that GATA-1 interactions with SNF2h and ACF1 occur in the context of the ACF/WCRF complex. These observations also suggest that GATA-1 interactions with MeCP1 and ACF/WCRF occur in distinct complexes. Fractionation profiles of transcription factors FOG-1, Ldb1, TAL-1 and Gfi-1b are largely coincident within the higher molecular weight fractions 16-28, though peaks vary between them. For most of these factors little or no protein is detected in the free protein fractions (i.e. fractions 32-38), in contrast to Sp1, which elutes as free protein.

fractionated by Superose-6 gel filtration and the fractionation profiles of GATA-1 and co-purifying proteins were determined (Fig. 3). GATA-1 displays a broad fractionation profile with several peaks at the high (fractions 16-25) and low (e.g. fraction 34 and below) molecular weight ends (Fig. 3A). The profile of tagged GATA-1 closely follows that of endogenous GATA-1 (Fig. 3A) is stable in salt concentrations of up to 1M and is not dependent on the presence of DNA (data not shown). The fractionation profiles of members of the MeCP1 complex showed overlapping peaks around fractions 20-22 (e.g. MTA2, RbAp46/48, Mbd2/3; Fig. 3B). By contrast, Sin3A fractionated with a peak around fractions 18-20 (Fig. 3B). The fractionation peaks of SNF2h,

ACF1 overlap between fractions 26-28 and are distinct from those of the MeCP1 complex (Fig. 3B). The peak of p17 in fractions 34-36 provides further support for GATA-1 interactions with SNF2h and ACF1 occurring in the context of the ACF/WCRF complex. Furthermore, indirect purification in our experiments of SNF2h by association with the MeCP1 complex, as has been previously described[35], is unlikely. These observations also suggest that GATA-1 interactions with MeCP1 and ACF/WCRF occur in distinct complexes.

The fractionation profiles of transcription factors FOG-1, Ldb1, TAL-1 and Gfi-1b are largely coincident within the higher molecular weight fractions 16-28 (Fig. 3C), though peaks vary between them. For most of these factors little or no protein is detected fractionating with a molecular weight corresponding to that of the free protein (i.e. around fractions 32-38). By contrast, the fractionation profile of Sp1 corresponds to that of the free protein (Fig. 3C), in agreement with the lack of a detectable complex between GATA-1 and Sp1 in MEL cells. In conclusion, the broad fractionation profile of GATA-1 most likely reflects the participation of GATA-1 in several distinct high molecular weight complexes with transcription factors and/or chromatin remodeling/modification complexes.

GATA-1 forms several distinct complexes

To directly confirm the distinct GATA-1 interactions and to assess how the GATA-1 partners may be partitioned in the GATA-1 complexes, we carried out sequential immunodepletion experiments. First, we used an antibody against one of the GATA-1 partners, i.e. FOG-1, TAL-1 or MTA2 in order to immunodeplete from a nuclear extract the fraction of GATA-1 that is in complex with this factor (Fig. 4A). The remaining GATA-1 in the supernatant was subsequently immunoprecipitated with a GATA-1 antibody and both immunoprecipitates were tested for the presence or absence of GATA-1 and interacting proteins (Fig. 4A).

We first established that the antibodies against GATA-1, TAL-1, FOG-1 and MTA2 were efficient in immunodepleting most of these proteins from nuclear extracts (Fig. 4B). As expected, FOG-1 antibodies immunoprecipitated a fraction of GATA-1 (Fig. 4C, lane 2). Surprisingly, MTA2 was also specifically immunoprecipitated by FOG-1 antibodies (Fig. 4C, lane 2), suggesting an interaction between FOG-1 and the MeCP1 complex. This was confirmed by the reverse immunoprecipitation of FOG-1 by an MTA2 antibody (Fig. 4E). There was no immunoprecipitation of TAL-1, Gfi-1b or ACF1 by FOG-1 antibodies (Fig. 4C, lane 2). Importantly, MTA2 could no longer be detected in the subsequent immunoprecipitation of the supernatant with GATA-1 antibodies (Fig. 4C, lane 3). Thus the fraction of GATA-1 that is in complex with MTA2 (and MeCP1) was depleted in the first step by the FOG-1 antibodies, leading to the conclusion that GATA-1 and FOG-1 interact together in the same complex with MeCP1. Following the FOG-1 immunodepletion, GATA-1 antibodies could still immunoprecipitate TAL-1, Gfi-1b and ACF1 (Fig. 4C, lane 3). This confirms that GATA-1 participates in a complex with FOG-1 and MeCP1 that is distinct from those with TAL-1, Gfi-1b or ACF/WCRF. Using TAL-1 antibodies in the first immunodepletion step, a small fraction of GATA-1, but no MTA2 or ACF1 was immunoprecipitated (Fig. 4D, lane 2).

We also carried out an MTA2 immunodepletion to determine whether the entire fraction of GATA-1 interacting with FOG-1 does so in the context of the MeCP1 complex. Following the immunodepletion of MTA2 (Fig. 4B), an appreciable amount of FOG-1 was subsequently immunoprecipitated by GATA-1 antibodies (Fig. 4E, lane 3). Thus, GATA-1 also interacts with FOG-1 independently of the MeCP1 complex. Interestingly, the MTA2 antibody specifically immunoprecipitated the slower migrating of the two bands detected by the FOG-1 antibody,

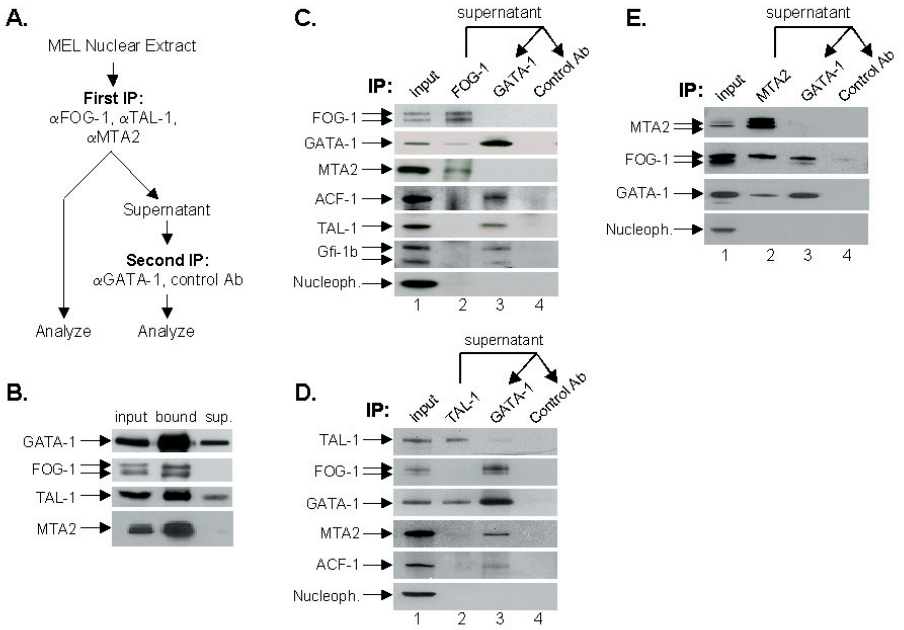


Figure 4: Distinct GATA-1 complexes by sequential immunoprecipitations (IP). (A) Experimental procedure.
(B) Efficiency of immunoprecipitating antibodies (also used to detect the immunoprecipitated protein). Sup: supernatant after IP. (C) FOG-1 immunodepletion, FOG-1 IP (lane 2) followed by IP of supernatant with GATA-1 or control antibodies (lanes 3 and 4). (D) TAL-1 immunodepletion, same as (C) using TAL-1 antibodies in first IP. (E) MTA2 immunodepletion, same as (C) and (D) using MTA2 antibodies in first IP. The MTA2 antibody used in panels B and E is different to that used in panels C and D (see Suppl. Methods). IP: immunoprecipitating antibody. Arrows show the detecting antibodies.

suggesting differential interaction with one of the two FOG-1 isoforms, while GATA-1 can interact with both FOG-1 isoforms (Fig. 4E, lane 3). Taken together, these experiments show that GATA-1 forms at least five complexes: first with FOG-1 and MeCP1, second with FOG-1 alone, third with TAL-1 (and Ldb1 since it can be almost completely immunodepleted by TAL-1 antibodies [not shown]), fourth with Gfi-1b and fifth with ACF/WCRF.

The GATA-1 zinc fingers mediate differential protein interactions

GATA-1 contains two evolutionarily conserved, closely spaced zinc finger domains. The C-terminal zinc finger (C-ZnF) is essential for DNA binding whereas the N-terminal zinc finger (N-ZnF) is primarily involved in protein-protein interactions, for example with FOG-1, which contribute to the specificity and stability of DNA binding by the C-ZnF (review[36]). Significantly, the C-ZnF is essential for all *in vivo* GATA-1 functions, whereas the N-ZnF is required for definitive, but not primitive erythropoiesis[37]. We addressed how the GATA-1 zinc fingers mediated its multiple protein interactions by expressing in MEL cells biotin-tagged mutants lacking the N-ZnF or the C-ZnF followed by streptavidin pull-downs (Fig. 5A-C). As described[15], GATA-1 interaction with FOG-1 requires the N-ZnF. Interactions of the MeCP1 members MTA2, MBD2

and HDAC 1 also occur through the N-ZnF of GATA-1 (Fig. 5A). Interestingly, interactions of GATA-1 with TAL-1 require both zinc fingers (Fig. 5B), whereas interactions with SNF2h or Gfi-1b require only the C-ZnF (Fig. 5C). We tested by immunoprecipitation using Gfi-1b antibodies whether Gfi-1b and SNF2h were in complex but found no evidence of such an interaction (not shown). Thus the multiple, distinct interactions of GATA-1 are differentially mediated through its zinc finger domains.

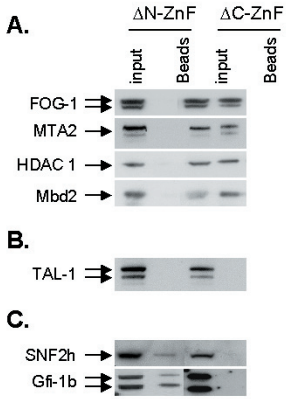


Figure 5: (A-C): Differential interactions mediated by the GATA-1 zinc fingers.
GATA-1 zinc finger deletions were expressed as biotin tagged proteins in MEL cells and interactions were assessed by streptavidin pull-downs and Western blots. (A) FOG-1 and the MeCP1 complex require the N-ZnF for interactions. (B) TAL-1 requires both zinc fingers. (C) Gfi-1b and SNF2h require the C-ZnF for interactions with GATA-1. The TAL-1 antibody used is different to that used in Fig. 1.

GATA-1 and FOG-1 associate with histone deacetylase activity

We further focused on the GATA-1 association with MeCP1 for the reasons that: (i) this is a novel finding, (ii) it represents the most abundant of the GATA-1 complexes linked to chromatin structure and (iii) it involves FOG-1, an essential hematopoietic transcription factor. We first tested whether GATA-1 and FOG-1 are associated with histone deacetylase (HDAC) activity, as would be predicted by their association with the MeCP1 complex. We first assayed for HDAC activity a number of immunoprecipitates using antibodies against GATA-1 as well as antibodies against Rpd3 (class I HDAC enzyme), FOG-1, TAL-1 and members of the MeCP1 complex (Fig. 6A). As expected, the highest HDAC activity was immunoprecipitated by the anti-Rpd3 antibody. HDAC activity was also clearly detectable in the GATA-1 immunoprecipitation. FOG-1 antibodies immunoprecipitated significant HDAC activity, further establishing its association with the MeCP1 complex. By contrast, little HDAC activity was immunoprecipitated by TAL-1 antibodies. All immunoprecipitates with antibodies against members of the MeCP1 complex contained significant HDAC activity (Fig. 6A). The HDAC activity immunoprecipitated by HDAC2, FOG-1 and GATA-1 antibodies was sensitive to the Class I HDAC inhibitors trichostatin A (TSA) and OSI-2040 (Fig. 6B)[38,39]. In addition, we immunodepleted the MeCP1 complex from nuclear extracts using MTA2 antibodies and found that most of the HDAC activity associated with the FOG-1 and GATA-1 immunoprecipitates was also depleted (Fig. 6C). Under the same conditions, there is considerable total HDAC activity remaining in the supernatant following MTA2 immunodepletion (not shown).

FOG-1 mediates interactions of GATA-1 with the MeCP1 complex in repressing transcription

We next tested whether FOG-1 mediates interactions between GATA-1 and the MeCP1 complex. GATA-1 was transiently expressed in HeLa cells (which express endogenous

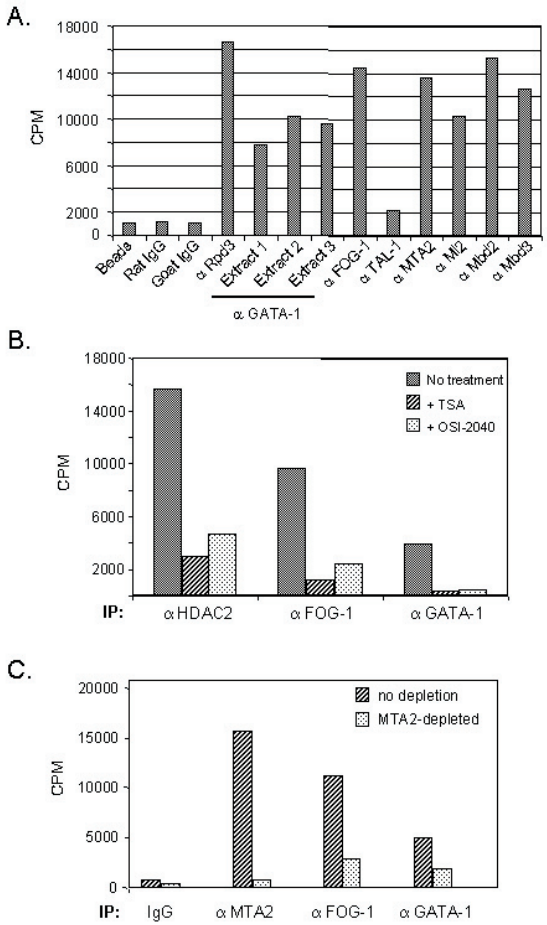


Figure 6: Histone deacetylase (HDAC) assays.
(A) HDAC activity associated with proteins immunoprecipitated by the indicated antibodies. Protein G beads and rat and goat immunoglobulins (IgG) were used as background controls. Three different nuclear extracts were used in the GATA-1 immunoprecipitations (extract 1: non-transfected MEL cells; extracts 2 and 3: biotinylated GATA-1). (B) HDAC activity immunoprecipitated by HDAC2, FOG-1 and GATA-1antibodies is sensitive to the Class I HDAC inhibitors TSA and OSI-2040. (C) Antibodies against the Mi2/NuRD-associated protein MTA2 deplete a considerable part of the HDAC activity associated with FOG-1 and GATA-1 immunoprecipitates. Immunoglobulins (IgG) were used as background control.

MeCP1, but not GATA-1 or FOG-1) with or without FOG-1, followed by immunoprecipitation using FOG-1 or MTA2 antibodies (Fig. 7A-B). We find that the interaction of GATA-1 with MTA2 occurs only in the presence of FOG-1 (Fig. 7B), whereas FOG-1 interacts with MTA2 regardless of the presence or absence of GATA-1 (Fig. 7B, upper panel). Expression of the GATA-1 zinc-finger deletion mutants (Fig. 7B, lower panel) confirmed these observations. We conclude that interaction of GATA-1 with the MeCP1 complex requires interaction with FOG-1, which thus serves as the bridging factor.

We next tested whether the well known GATA-1 and FOG-1 mediated repression is

due to the recruitment of the MeCP1 complex to a GATA-dependent promoter. To this end, we used a reporter plasmid containing the rabbit β-globin minimal promoter (pOVEC-1,[40], carrying four copies of an optimal GATA-1 binding sequence, or four copies of a mutated sequence that abolish GATA-1 binding[41]). The GATA-binding promoter was activated more than six-fold by co-transfection of GATA-1 alone (Fig. 7C). As expected, co-transfection of FOG-1 and GATA-1 repressed activation of the GATA-dependent promoter (Fig. 7C). Chromatin immunoprecipitations (ChIP) showed that repression by GATA-1 and FOG-1 was due to the specific recruitment of the MeCP1 complex. Binding of Mi-2β to the repressed gene was specifically enriched in GATA-1 and FOG-1 transfected cells (Fig. 7D), but not in cells transfected with GATA-1 only. The promoter bearing the mutated GATA binding sites does not bind MeCP1, even in the presence of FOG-1 (Fig. 7D). Thus, FOG-1/MeCP1 repression is mediated through GATA-1 binding at its cognate binding sites.

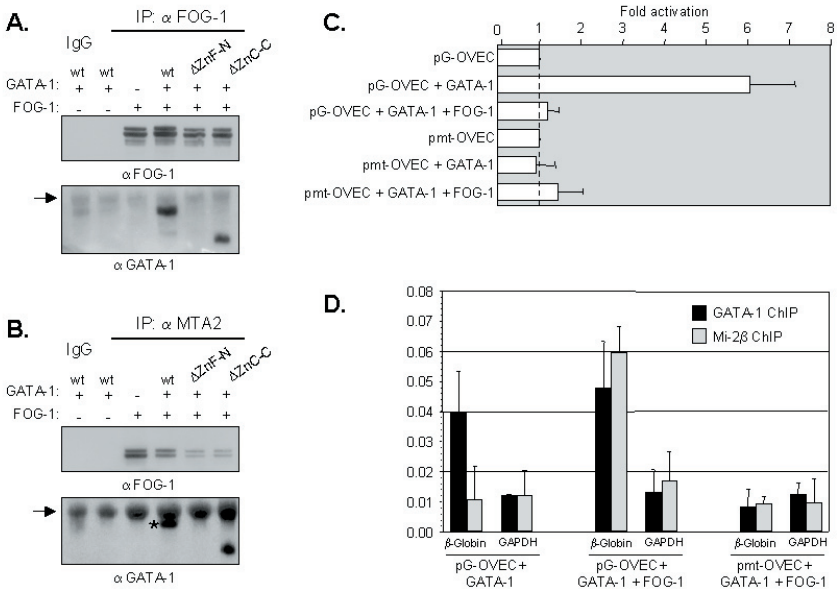


Figure 7: (A-B): FOG-1 bridges GATA-1 and MeCP1.
Nuclear extracts from HeLa cells transfected with the FOG-1 and GATA-1 combinations indicated, were immunoprecipitated with FOG-1 (panel A) or MTA2 antibodies (panel B) and detected with FOG-1 and GATA-1 antibodies. ΔZn-N and ΔZn-C: GATA-1 N- and C-terminal zinc-finger deletion mutants. Arrows: cross-reacting IgG. Asterisk (panel B): GATA-1 signal.

(C) Real-Time PCR transcription assays in transfected HeLa cells. *GATA-1 activates transcription of pG-OVEC, whereas co-transfection of FOG-1 represses to basal levels.*

(D) Specific recruitment of Mi-2β by co-transfected GATA-1 and FOG-1 by ChIP assays in HeLa cells. *Mi-2β recruitment to the repressed gene requires GATA-1 binding to the promoter.*

Discussion

We describe here the characterization of GATA-1 complexes from erythroid cells by *in vivo* biotinylation tagging and purification by streptavidin beads. This work has led to a number of important findings. First, we identified novel GATA-1 partners, including the essential hematopoietic factor Gfi-1b and the chromatin remodeling and modification complexes MeCP1 and ACF/WCRF, in addition to the known GATA-1 interacting factors FOG-1, TAL-1 and Ldb1. Second, we showed that GATA-1 forms several distinct complexes with FOG-1, with FOG-1 and MeCP1, with TAL-1/Ldb1, with Gfi-1b and with the ACF/WCRF complex. Third, we found that the most abundant of the GATA-1 complexes are those with FOG-1 and with FOG-1 and MeCP1, with FOG-1 serving as the bridging factor between GATA-1 and the MeCP1 complex. Fourth, we showed that the distinct interactions of GATA-1 with its protein partners are differentially mediated through the two GATA-1 zinc finger domains. Finally, our work demonstrates the utility of biotinylation tagging as an efficient approach for the rapid isolation and identification by mass spectrometry of multiple protein complexes.

Biotinylation tagging and protein complex purification

From our previous work[23] and the work described here, we show that background using biotinylation tagging consists of naturally biotinylated proteins, of abundant nuclear proteins such as splicing factors binding nonspecifically to the beads[23] and, potentially, of abundant chromatin-associated proteins, such as topoisomerase I, which are indirectly pulled-down with the tagged transcription factor (Table 1). We have validated a number of the remaining proteins as being true GATA-1 partners, some of which represent low abundance or weaker GATA-1 interactions, e.g. with TAL-1/Ldb1, Gfi-1b and ACF/WCRF. Importantly, purification required a single capture step.

We cannot be certain that we identified all GATA-1 complexes in differentiated MEL cells. Indeed, some of the size-fractionation profiles (Fig. 3) suggest that there may be additional protein partners that were not identified perhaps due to their very low abundance or instability. This may be the case for the multimeric GATA-1/TAL-1/Ldb1/E2A/LMO2 complex. Several lines of evidence have suggested the presence of this complex in erythroid cells binding to distinct E-box and GATA motifs spatially arranged 9 to 12 nucleotides apart (review by[42]). Many erythroid genes identified to-date contain such motifs, including GATA-1 itself, EKLF, glycophorin A and 4.2 protein[42]. Evidence for the multimeric GATA-1/TAL-1 complex binding to erythroid genes *in vivo*, such as α globin and glycophorin A, has been provided recently by ChIP assays[19,43]. Nevertheless, we did not find any co-purification of E2A or LMO2 with GATA-1 from induced MEL cells. The complementary isolation by biotinylation tagging of protein partners, such as TAL-1, will be informative in that respect and may also reveal additional protein partners. Indeed, work presented in chapter 4 shows the co-purification of GATA-1 with complexes purified by biotin tagged Ldb1.

Novel GATA-1 protein partners

We describe here, for the first time, an interaction of GATA-1 with the essential hematopoietic transcription factor Gfi-1b. This factor contains six C-terminal C_2H_2 zinc fingers, which bind a defined DNA consensus sequence, and an N-terminal SNAG domain associated with repression[44,45]. The Gfi-1b knockout is remarkably similar to that of GATA-1, i.e. it shows embryonic lethality E15 due to the developmental arrest of erythroid and megakaryocytic

differentiation in the fetal liver[25]. It is important to note that although there are also similarities between the FOG-1 and Gfi-1b knockout phenotypes, we did not find FOG-1 and Gfi-1b to directly interact in induced MEL cells (Fig. 1 and 4). Possibly, the two factors regulate common gene targets through distinct complexes and binding sites. Alternatively, the functions of GATA-1 with FOG-1 or Gfi-1b could be separate, e.g. differentiation (FOG-1) versus proliferation arrest (Gfi-1b), with each function being essential for erythropoiesis. We also provide preliminary evidence for distinct gene targets for FOG-1 and Gfi-1b in Chapter 4.

We also describe, for the first time, interactions of GATA-1 with the MeCP1 and ACF/WCRF complexes, linking GATA-1 to repressive functions (with MeCP1) and chromatin structure. Previous evidence linking GATA-1 to chromatin structure involved interactions with the histone acetyltransferases (HATs) CBP and p300[20] and *in vitro* experiments where GATA-1 co-operated with the SWI/SNF remodeling complex in transcriptional activation[21]. However, we did not observe these interactions in our GATA-1 purification from induced MEL cells or in immunoprecipitations (data not shown).

Our observations on the interactions of GATA-1 (and FOG-1) with the MeCP1 complex add to previous reports linking MeCP1 (and the closely related NuRD complex) to transcription factors in hematopoiesis[46-48]. Significantly, the transcription factor Ikaros was shown to interact with the NuRD complex in erythroid cells (O'Neill et al. 2000). In addition, the third member of the GATA family, GATA-3 was shown to functionally interact with the NuRD complex in differentiating T cells, where GATA-3 was shown to displace Mbd2 from the repressed IL-4 gene thus allowing its activation. These observations suggest that GATA-3 is antagonistic to MeCP1-mediated gene repression, in contrast to GATA-1 which tethers MeCP1 for gene repression[46]. Furthermore, the conditional knockout of Mi-2β in thymocytes revealed a requirement in different stages of T cell maturation[49]. Lastly, the characterization of the MTA3 member of NuRD in B lymphocytes showed an interaction with BCL-6, a key repressor of the mature plasma cell transcription program[50]. In these experiments, it was suggested that MTA3 and the NuRD complex play a role in the maintenance of a population of less differentiated, "poised" B lymphocytes[50]. By contrast, our data in erythroid cells suggest that the MeCP1 complex works with tissue-specific transcription factors to effect terminal differentiation by shutting down transcription programs associated with early multipotential ("poised") states (see next chapter).

A specific interaction between GATA-1 and ACF/WCRF was also found. It is of note that the knockdown of SNF2h in primary human proerythroblasts blocks erythroid differentiation, a phenotype similar to that of the GATA-1 knockout[31]. Thus, the GATA-1/ACF/WCRF complex could provide an essential aspect in GATA-1 functions in erythroid differentiation. However, it remains unclear as to whether the GATA-1/ACF/WCRF interactions lead to repression or activation. Interestingly, in Drosophila, mutations in the NURF complex, an ISWI family complex, were shown to lead to blood defects. Mutants of the Nurf 301 subunit, a homolog of ACF, are characterized by a high number of circulating blood cells, suggesting a potential role in malignancy[51]. Whereas ISWI complexes were originally associated with activation[29] recent evidence in Drosophila showed by immunofluorescence that localization of ISWI does not overlap with transcriptionally active sites, as visualized by staining for RNA polymerase II[52]. By contrast, the SWI/SNF complex is almost exclusively associated with active genes[53]. These data suggest that ISWI is linked to both gene activation and gene repression.

GATA-1 had previously been reported to bind to additional transcription factors (Sp1) and co-factors (SWI/SNF, CBP/P300) that we did not find by mass spectrometry to co-purify with biotin-tagged GATA-1 or by immunoprecipitations[20,54-56]. There are a number of reasons that

could account for these discrepancies. Firstly, the nature of GATA-1 complexes may vary during developmental stages of erythroid cells. Therefore, we can imagine that the interaction between CBP/P300 and GATA-1 in non-differentiated MEL cells is no longer detectable in the final stages of terminal differentiation. However, preliminary data from the characterization of GATA-1 complexes from non-induced MEL cells showed that CBP/P300 are again absent. Nevertheless, we cannot exclude that the binding of CBP/P300 to GATA-1 is highly dynamic or labile or it only takes place on DNA, as evidenced by ChIP assays (REF). Secondly, the discrepancies may be due to the nature of the assay used previously to identify these interactions. The transcription factor Sp1 was shown to bind to GATA-1 either indirectly by using EMSA assays or by GST pull-downs[54] . Though we found no evidence of a pre-formed, soluble GATA-1/Sp1 complex, we can think of a model in which Sp1 is interacting with GATA-1 via DNA to activate transcription. GATA-1 was also shown to interact with the chromatin remodeling complex SWI/SNF by GST-GATA-1 pull-downs[56]. Interestingly, we have not identified any SWI/SNF complex components in the GATA-1 purification, but instead all components of the MeCP1 and ACF/WCRF complexes to link GATA-1 to chromatin structure regulation. Again the absence of SWI/SNF in the GATA-1 purification may be due to limitations in the original assay used in the original report, or due to a highly dynamic nature of the interactions.

GATA-1 and FOG-1 interactions

Considerable evidence has linked GATA-1 functions to FOG-1 (reviewed by[2,36]). A single amino acid change in the N-terminal zinc finger of GATA-1 which abolishes interaction with FOG-1[14], resulted in lethality in mice due to severe anemia[57] and is associated with dyserythropoietic anemia in patients[58]. Our work suggests that the overlapping functions of GATA-1 and FOG-1 in erythropoiesis occur in the context of two distinct complexes, a GATA-1/FOG-1/MeCP1 complex and a GATA-1/FOG-1 complex. Clearly, the association of GATA-1 and FOG-1 with the MeCP1 complex provides the molecular basis for the well-documented repressive properties of GATA-1 and FOG-1 interactions[14,59-61]. This is supported by our experiments using transcription and ChIP assays in transfected HeLa cells. It is of note that only the slower migrating isoform of FOG-1 (Fig. 4) interacts with the GATA-1/MeCP1 complex, providing a potential mechanism for the selective formation of the GATA-1/FOG-1/MeCP1 complex.

Hong and colleagues have purified a similar FOG-1 complex using an alternative approach[62]. These authors did not detect the presence of MBD2 during purification of the FOG-1 complex, which led them to suggest that FOG-1 interacts with the NuRD complex instead of the MeCP1 complex. Interestingly, GATA-1 was also not identified as co-purifying with the FOG-1/NuRD complex. Since Hong and colleagues used only the N-terminal 45 amino acids of FOG-1 in the purification, it is possible that Mbd2 might bind to another domain of FOG-1 outside the 45aa used in the purification. This would certainly be the case for GATA-1 since it has been shown to interact with FOG-1 through the latter's zinc fingers and not the N-terminal 45aa domain used in the purification. Another explanation could be that MBD2 would interact with the FOG-1/NuRD complex only in the presence of GATA-1.

We suggest that the separate GATA-1/FOG-1 complex without MeCP1 is responsible for the transcriptional activation functions that have been previously described for GATA-1 with FOG-1. For example, disruption of GATA-1 and FOG-1 interactions down-regulates erythroid genes such as α and β globin, Band 3, DC11 and HD2 genes[14,61]. ChIP assays have also shown GATA-1 and FOG-1 to be bound *in vivo* to active genes such as the α globin locus and the GATA-1 gene itself[43,60]. Significantly, in the α globin locus the GATA-1/FOG-1 complex occupies

sites distinct from those occupied by the GATA-1/TAL-1/Ldb1 complex[43], in agreement with our findings of distinct GATA-1 complexes.

Our finding that FOG-1 bridges GATA-1 to the repressive MeCP1 complex partly explains the common features of the GATA-1 and FOG-1 knockouts and the phenotypes caused by the single amino acid change in the N-terminal zinc finger of GATA-1 in mice and patients. In the GATA-1 knockout FOG-1/MeCP1 cannot be tethered to target genes, whereas in the FOG-1 knockout, the interaction between GATA-1 and the MeCP1 complex cannot take place. In patients, the lack of interaction between GATA-1 and FOG-1 would also fail to tether the MeCP1 complex to some of their target genes.

In conclusion, we have identified at least five distinct complexes involving essential transcription factors and chromatin remodeling and/or modifying complexes (Figure 8). These observations on the GATA-1 interactions raise important questions as to what the functions of these complexes are *in vivo* and which genes are the *in vivo* targets of these complexes in erythroid cells.

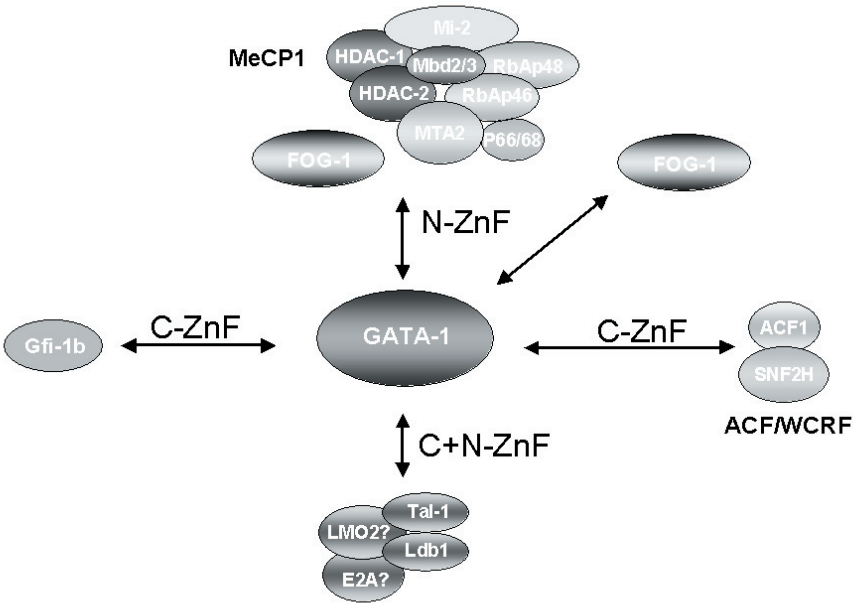


Figure 8: Model for the distinct GATA-1 complexes in erythroid cells.

Materials and methods

Constructs, nuclear extract preparation, streptavidin binding, mass spectrometry and immunoblot analysis. Tagged constructs and procedures involving MEL cells, biotinylated proteins and mass spectrometry were previously described[23]. The GATA-1 zinc finger deletions have been described [41]. G1E cells and induction were described[13,63].

Superose 6 gel filtration. Size fractionation of protein complexes was done on an AKTA FPLC apparatus with a Superose 6 10/30 column (Amersham Biosciences, Piscataway NJ). Fractions were precipitated with 100% trichloroacetic acid and analyzed by Western immunoblotting, as described[23]. Molecular size standards were thyroglobulin (670kDA) and albumin (66kDa) (Amersham Biosciences, Piscataway NJ).

Immunoprecipitations. Nuclear extracts were ple-cleared at 4°C using Protein G sepharose beads and affinity-purified IgG (rat [Santa Cruz, CA, sc-2026], rabbit [Santa Cruz, sc-2027], goat [Santa Cruz, sc-2028]) in HENG150 buffer. GATA-1 and TAL-1 antibodies were crosslinked to beads using dimethyl pimelimidate. Immunoprecipitations were performed in HENG150 / 0.3% NP-40 buffer overnight at 4°C using protein-G Sepharose beads. Washes were done at room temperature in HENG250 / 0.3% NP-40 buffer. Bound material was eluted by boiling in 1x Laemmli buffer.

HDAC assays. Immunoprecipitations for HDAC assays were carried out using 0.3-0.5mg of nuclear extracts in HENG150 / 0.1% NP-40, as described above. Immunoprecipitates were washed once for 10 minutes in HENG150 / 0.3% NP-40 and three times for 10 minutes each in HENG300 / 0.3%NP-40. Beads were resuspended in HENG50 and HDAC assays were done using approximately 32,000cpm of ^{3}H-labelled core histones per reaction, as previously described[64].

HeLa transient transfection and transcription assays. GATA-1 and FOG-1 cDNAs cloned in pCDNA 3.1 (Invitrogen, Carlsbad CA) were transiently transfected using 2μg DNA and Lipofectamine 2000 (Invitrogen, Carlsbad CA). Cells were harvested after 24 hours and nuclear extracts were used for immunoprecipitations as above. pEGFP-N1 (Invitrogen, Carlsbad CA) was included as transfection efficiency control. Transcription was assayed by Real-Time PCR with primers for exon 2 of the pOVEC reporter plasmid. ChIP assays were done as below using GATA-1 and Mi-2β antibodies. The endogenous human GAPDH gene was used as control.

Acknowledgements

We are grateful to our many colleagues cited for generously sharing reagents. We are also grateful to Karim Bouazoune and Alexander Brehm (Adolf Butenandt Institut, Munich) for help with the HDAC assays, Sjaak Philipsen and Marieke von Lindern (Erasmus MC) for comments on the manuscript and Raymond Poot (CSC, London) for helpful discussions. We are grateful to Gerd Blobel (Children's Hospital of Philadelphia) for exchanging pre-publication data. This work has been supported by the Netherlands Research Organization (NWO), the European Union (grant HPRN-CT-2000-00078), the NIH (grant RO1 HL 073455-01) and the Netherlands Proteomics Center.

References

1. Sieweke, M. H. & Graf, T. A transcription factor party during blood cell differentiation. *Curr Opin Genet Dev* **8**, 545-51 (1998).
2. Cantor, A. B. & Orkin, S. H. Transcriptional regulation of erythropoiesis: an affair involving multiple partners. *Oncogene* **21**, 3368-76 (2002).
3. Weiss, M. J. & Orkin, S. H. GATA transcription factors: key regulators of hematopoiesis. *Exp Hematol* **23**, 99-107 (1995).
4. Weiss, M. J. & Orkin, S. H. Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis. *Proc Natl Acad Sci U S A* **92**, 9623-7 (1995).
5. Pevny, L. et al. Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* **349**, 257-60 (1991).
6. Shivdasani, R. A., Fujiwara, Y., McDevitt, M. A. & Orkin, S. H. A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. *Embo J* **16**, 3965-73 (1997).
7. Yu, C. et al. Targeted deletion of a high-affinity GATA-binding site in the GATA-1 promoter leads to selective loss of the eosinophil lineage in vivo. *J Exp Med* **195**, 1387-95 (2002).
8. Pevny, L. et al. Development of hematopoietic cells lacking transcription factor GATA-1. *Development* **121**, 163-72 (1995).
9. Visvader, J., Elefanty, A., Strasser, A. & Adams, J. GATA-1 but not SCL induces megakaryocytic differentiation in an early myeloid line. *Embo J* **11**, 4557-64 (1992).
10. Kulessa, H., Frampton, J. & Graf, T. GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts. *Genes Dev* **9**, 1250-62 (1995).
11. Heyworth, C., Pearson, S., May, G. & Enver, T. Transcription factor-mediated lineage switching reveals plasticity in primary committed progenitor cells. *Embo J* **21**, 3770-81 (2002).
12. Iwasaki, H. et al. GATA-1 converts lymphoid and myelomonocytic progenitors into the megakaryocyte/erythrocyte lineages. *Immunity* **19**, 451-62 (2003).
13. Tsang, A. P. et al. FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell* **90**, 109-19 (1997).
14. Crispino, J. D., Lodish, M. B., MacKay, J. P. & Orkin, S. H. Use of altered specificity mutants to probe a specific protein-protein interaction in differentiation: the GATA-1:FOG complex. *Mol Cell* **3**, 219-28 (1999).
15. Tsang, A. P., Fujiwara, Y., Hom, D. B. & Orkin, S. H. Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG. *Genes Dev* **12**, 1176-88 (1998).
16. Wadman, I. A. et al. The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *Embo J* **16**, 3145-57 (1997).
17. Xu, Z., Huang, S., Chang, L. S., Agulnick, A. D. & Brandt, S. J. Identification of a TAL1 target gene reveals a positive role for the LIM domain-binding protein Ldb1 in erythroid gene expression and differentiation. *Mol Cell Biol* **23**, 7585-99 (2003).
18. Lecuyer, E. et al. The SCL complex regulates c-kit expression in hematopoietic cells through functional interaction with Sp1. *Blood* **100**, 2430-40 (2002).
19. Lahlil, R., Lecuyer, E., Herblot, S. & Hoang, T. SCL assembles a multifactorial complex that determines glycophorin A expression. *Mol Cell Biol* **24**, 1439-52 (2004).
20. Blobel, G. A., Nakajima, T., Eckner, R., Montminy, M. & Orkin, S. H. CREB-binding protein cooperates with transcription factor GATA-1 and is required for erythroid differentiation. *Proc Natl Acad Sci U S A* **95**, 2061-6 (1998).
21. Kadam, S. & Emerson, B. M. Transcriptional specificity of human SWI/SNF BRG1 and BRM chromatin remodeling complexes. *Mol Cell* **11**, 377-89 (2003).
22. Amstrong, J., Bieker, J. & Emerson, B. A SWI/SNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLF in vitro. *Cell* **95**, 93-104 (1998).

23. de Boer, E. et al. Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc Natl Acad Sci U S A* **100**, 7480-5 (2003).
24. Elefanty, A. G., Antoniou, M., Custodio, N., Carmo-Fonseca, M. & Grosveld, F. G. GATA transcription factors associate with a novel class of nuclear bodies in erythroblasts and megakaryocytes. *Embo J* **15**, 319-33 (1996).
25. Saleque, S., Cameron, S. & Orkin, S. H. The zinc-finger proto-oncogene Gfi-1b is essential for development of the erythroid and megakaryocytic lineages. *Genes Dev* **16**, 301-6 (2002).
26. Feng, Q. & Zhang, Y. The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. *Genes Dev* **15**, 827-32 (2001).
27. Feng, Q. et al. Identification and functional characterization of the p66/p68 components of the MeCP1 complex. *Mol Cell Biol* **22**, 536-46 (2002).
28. Ahringer, J. NuRD and SIN3 histone deacetylase complexes in development. *Trends Genet* **16**, 351-6 (2000).
29. Corona, D. F. & Tamkun, J. W. Multiple roles for ISWI in transcription, chromosome organization and DNA replication. *Biochim Biophys Acta* **1677**, 113-9 (2004).
30. Bochar, D. A. et al. A family of chromatin remodeling factors related to Williams syndrome transcription factor. *Proc Natl Acad Sci U S A* **97**, 1038-43 (2000).
31. Stopka, T. & Skoultchi, A. I. The ISWI ATPase Snf2h is required for early mouse development. *Proc Natl Acad Sci U S A* **100**, 14097-102 (2003).
32. Osada, H., Grutz, G., Axelson, H., Forster, A. & Rabbitts, T. H. Association of erythroid transcription factors: complexes involving the LIM protein RBTN2 and the zinc-finger protein GATA1. *Proc Natl Acad Sci U S A* **92**, 9585-9 (1995).
33. Visvader, J. E., Mao, X., Fujiwara, Y., Hahm, K. & Orkin, S. H. The LIM-domain binding protein Ldb1 and its partner LMO2 act as negative regulators of erythroid differentiation. *Proc Natl Acad Sci U S A* **94**, 13707-12 (1997).
34. Eberharter, A. et al. Acf1, the largest subunit of CHRAC, regulates ISWI-induced nucleosome remodelling. *Embo J* **20**, 3781-8 (2001).
35. Hakimi, M. A. et al. A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. *Proc Natl Acad Sci U S A* **99**, 7420-5 (2002).
36. Blobel, G. A. & Weiss, M. J. in *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management.* (eds. Steinberg, M. H., Forget, B. G., Higgs, D. R. & Nagel, R. L.) 72-94 (Cambridge University Press, Cambridge, 2001).
37. Shimizu, R., Takahashi, S., Ohneda, K., Engel, J. D. & Yamamoto, M. In vivo requirements for GATA-1 functional domains during primitive and definitive erythropoiesis. *Embo J* **20**, 5250-60 (2001).
38. Yoshida, M., Kijima, M., Akita, M. & Beppu, T. Potent and specific inhibition of mammalian histone deacetylase both in vivo and in vitro by trichostatin A. *J Biol Chem* **265**, 17174-9 (1990).
39. Haley, J. D. et al. Identification and characterization of mechanistically distinct inducers of gamma-globin transcription. *Biochem Pharmacol* **66**, 1755-68 (2003).
40. Westin, G., Gerster, T., Muller, M. M., Schaffner, G. & Schaffner, W. OVEC, a versatile system to study transcription in mammalian cells and cell-free extracts. *Nucleic Acids Res* **15**, 6787-98 (1987).
41. Whyatt, D. J., deBoer, E. & Grosveld, F. The two zinc finger-like domains of GATA-1 have different DNA binding specificities. *Embo J* **12**, 4993-5005 (1993).
42. Lecuyer, E. & Hoang, T. SCL: from the origin of hematopoiesis to stem cells and leukemia. *Exp Hematol* **32**, 11-24 (2004).
43. Anguita, E. et al. Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. *Embo J* **23**, 2841-52 (2004).
44. Duan, Z. & Horwitz, M. Gfi-1 oncoproteins in hematopoiesis. *Hematology* **8**, 339-44 (2003).
45. Doan, L. L. et al. Targeted transcriptional repression of Gfi1 by GFI1 and GFI1B in lymphoid cells. *Nucleic Acids Res* **32**, 2508-19 (2004).
46. Hutchins, A. S. et al. Gene silencing quantitatively controls the function of a developmental trans-activator. *Mol Cell* **10**, 81-91 (2002).
47. O'Neill, D. W. et al. An ikaros-containing chromatin-remodeling complex in adult-type erythroid cells. *Mol Cell Biol* **20**, 7572-82 (2000).

48. Kim, J. et al. Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes. *Immunity* **10**, 345-55 (1999).
49. Williams, C. J. et al. The chromatin remodeler Mi-2beta is required for CD4 expression and T cell development. *Immunity* **20**, 719-33 (2004).
50. Fujita, N. et al. MTA3 and the Mi-2/NuRD complex regulate cell fate during B lymphocyte differentiation. *Cell* **119**, 75-86 (2004).
51. Badenhorst, P., Voas, M., Rebay, I. & Wu, C. Biological functions of the ISWI chromatin remodeling complex NURF. *Genes Dev* **16**, 3186-98 (2002).
52. Deuring, R. et al. The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo. *Mol Cell* **5**, 355-65 (2000).
53. Armstrong J. A., P. O., Daubresse G, Sperling A. S.,Lis J. T., Scott M. P. and Tamkun J. W. The Drosophila BRM complex facilitates global transcription by RNA polymerase II. *Embo J* **21**, 5245–5254 (2002).
54. Merika, M. & Orkin, S. Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Kruppel family proteins Sp1 and EKLF. *Mol Cell Biol* **15**, 2437-47 (1995).
55. Hung, H. L., Lau, J., Kim, A. Y., Weiss, M. J. & Blobel, G. A. CREB-Binding protein acetylates hematopoietic transcription factor GATA-1 at functionally important sites. *Mol Cell Biol* **19**, 3496-505 (1999).
56. Kadam, S. et al. Functional selectivity of recombinant mammalian SWI/SNF subunits. *Genes Dev* **14**, 2441-51 (2000).
57. Chang, A. N. et al. GATA-factor dependence of the multitype zinc-finger protein FOG-1 for its essential role in megakaryopoiesis. *Proc Natl Acad Sci U S A* **99**, 9237-42 (2002).
58. Nichols, K. E. et al. Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. *Nat Genet* **24**, 266-70 (2000).
59. Fox, A. H. et al. Transcriptional cofactors of the FOG family interact with GATA proteins by means of multiple zinc fingers. *Embo J* **18**, 2812-22 (1999).
60. Pal, S. et al. Coregulator-dependent facilitation of chromatin occupancy by GATA-1. *Proc Natl Acad Sci U S A* **101**, 980-5 (2004).
61. Letting, D. L., Chen, Y. Y., Rakowski, C., Reedy, S. & Blobel, G. A. Context-dependent regulation of GATA-1 by friend of GATA-1. *Proc Natl Acad Sci U S A* **101**, 476-81 (2004).
62. Hong W, N. M., Chen YY, Kori R, Vakoc CR, Rakowski C, Blobel GA. FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. *Embo J* **24**, 2367-78 (2005).
63. Weiss, M. J., Yu, C. & Orkin, S. H. Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. *Mol Cell Biol* **17**, 1642-51 (1997).
64. Taunton, J., Hassig, C. A. & Schreiber, S. L. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* **272**, 408-11 (1996).

Chapter 4:

Distinct GATA-1 complexes bind to different target gene subsets in erythroid cells

Patrick Rodriguez[1], Edgar Bonte[1], Jeroen Krijgsveld[2], Katarzyna Kolodziej[1], Boris Guyot[3], Albert J.R. Heck [2], Paresh Vyas[3], Ernie de Boer[1], Frank Grosveld[1] and John Strouboulis[1]

Part of this Chapter was published in EMBO, 2005, Jul 6; (24): 2354-66.

[1] Department of Cell Biology, Erasmus University Medical Center, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
[2] Department of Biomolecular Mass Spectrometry, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, The Netherlands
[3] Department of Haematology, The Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DU, United Kingdom

Distinct GATA-1 complexes bind to different target gene subsets in erythroid cells

Abstract

The hematopoietic transcription factor GATA-1 belongs to a family of factors that binds DNA through a GATA consensus-binding site. GATA-1 is essential for the generation of the erythroid, megakaryocytic, eosinophilic and mast lineages. Its functions involve lineage-specific gene activation and repression of early hematopoietic or alternative transcriptional programs. Roles in suppressing cell proliferation and in protecting differentiating erythroid cells from apoptosis have also been suggested. However, little is known about the molecular basis of GATA-1 function. In erythroid cells, we have shown that GATA-1 forms several distinct complexes with transcription factors and chromatin remodeling and/or modifying complexes. Here, we show that distinct GATA-1 complexes bind to specific subsets of target genes. Firstly, the repressive GATA-1/FOG-1/MeCP1 complex binds to the early hematopoietic GATA-2 gene locus. The interaction of GATA-1 with the FOG-1/MeCP1 complex is required for GATA-2 silencing during erythroid differentiation. The same GATA-1 repressive complex binds to the promoter of the eosinophilic Major Basic Protein (MBP) gene, which is upregulated by GATA-1 in eosinophils but repressed in erythroid cells. Similar observations were made for another eosinophilic gene (IL-5RA) and for the myeloid PU.1 transcription factor. Secondly, we show evidence of GATA-1 interacting with Gfi-1b in binding to the cell proliferation related c-myb and c-myc genes, which are repressed during erythroid differentiation. In contrast, the GATA-1/TAL-1 complex involved in erythroid gene activation, is stably bound to regulatory elements of the active EKLF gene. Our results suggest that the distinct GATA-1 complexes regulate subsets of genes involved in different aspects of GATA-1 function.

Introduction

Hematopoiesis is characterized by the generation of all blood cells from a unique hematopoietic stem cell (HSC) that has the capacity to self-renew or to differentiate into distinct lineages. Progressive differentiation gives rise to an increasing number of divergent and tightly defined lineages expressing a more restricted genetic program at each branch point. This stepwise process is regulated by signaling molecules and transcription factors that modulate the response of the downstream target genes[1,2].

Recent evidence has indicated that considerable plasticity exists in the transcriptional programs governing hematopoietic lineage specification. It has been shown that HSCs and multipotential progenitor cells are transcriptionally "promiscuous" in that they express, at low levels, genes associated with committed hematopoietic lineages[3-6]. These findings led to a model whereby multipotential progenitors are transcriptionally "primed" for differentiation along several different lineages. Commitment to a specific lineage occurs through the (stochastic) selection of a particular transcription program and the concomitant suppression of programs specifying alternative lineages[7-9]. Enforced ectopic expression of a key lineage-specific transcription factor in a lineage where it is not normally expressed, or expressed at low levels, can result in the reprogramming of this lineage towards other fates normally regulated by this factor (reviewed by Graf T.[10]). This reprogramming is accompanied by the repression of the original lineage-specific transcriptional program. Thus, hematopoietic lineage specification involves two key aspects, firstly, the appearance or increase in levels of key transcription factors which up-regulate lineage-specific transcription and, secondly, the suppression of alternative, "poised" lineage transcription programs, often by the same transcription factors that up-regulate lineage specific programs[7-9].

An example of such a key hematopoietic transcription factor is GATA-1. It is essential for the differentiation of the erythroid, megakaryocytic (Meg/E) and eosinophilic lineages[11-13]. GATA-1 is also expressed in mast cells and in some multipotential hematopoietic precursor cells. GATA-1 null mice are characterized by an absence of these lineages. Specifically in the erythroid lineage, the absence of mature red blood cells is due to a block of the progenitors at the proerythroblast stage, which leads to embryonic lethal anemia. The instructive role of GATA-1 to generate Meg/E and eosinophils was demonstrated in an avian *in vitro* system using a Myb-Ets-transformed multipotential chicken cell line[14]. In the murine model, the ectopic expression of GATA-1 in an early myeloid cell line, which can differentiate towards megakaryocytic or granulocytes lineages, led to the generation of megakaryocytic only cells[15]. In addition, the expression of megakaryocytic and erythroid-associated genes in the same system was shown to be up regulated[16,17]. Significantly, enforced expression of GATA-1 in common lymphoid progenitors and common myeloid progenitors reprograms them to erythroid, megakaryocytic and eosinophilic fates, with the concomitant repression and up regulation of respective lineage-specific marker genes[14,18,19]. Thus, GATA-1 acts as a key regulator in the specification of the aforementioned lineages, a role that apparently involves the activation as well as repression of lineage-restricted transcriptional programs. This hypothesis was confirmed by microarray analysis of GATA-1 dependent erythroid differentiation showing the up and down regulation of several classes of genes[20].

GATA-1 carries out activating and repressive functions in erythroid cells. GATA-1 was shown to activate several target genes, such as β-globin genes or EKLF[21,22]. In association with the pentameric complex which includes TAL-1/Ldb1/E2A/Lmo2, GATA-1 has been reported

to activate glycophorin A and the α-globin genes[21,23]. GATA-1 was shown to repress early hematopoietic programs genes, such as GATA-2[24], and was also linked to the down-regulation of cell proliferation related genes, such as c-myc and c-myb[25]. Analysis of GATA-1 dependent changes in gene expression obtained by microarray data clearly shows gene activation and repression providing further evidence for the dual functions of GATA-1[20].

Despite this evidence, it is still unclear how GATA-1 functions as an activator and a repressor in the same cell. Using the *in vivo* biotinylation tagging method (Chapter 2), we purified five distinct GATA-1 complexes containing other hematopoietic transcription factors and chromatin remodeling complexes. We show that distinct GATA-1 complexes are bound *in vivo* to genes that belong to different classes. We also provide direct evidence for GATA-2 gene repression by GATA-1 through the recruitment of the FOG-1/MeCP1 complex. On the basis of the evidence presented here and in Chapter 3, we propose a model in which GATA-1 has specific early versus late differentiation functions in the context of distinct complexes and subsets of target genes. This model is also supported by the observations of Welch et colleagues[20].

Results

GATA-1/ FOG-1/ MeCP1 and GATA-1/Gfi-1b are bound to repressed genes *in vivo*

The results presented in Chapter 2 suggest that GATA-1/FOG-1 interactions can tether MeCP1 to repressed GATA-1 target sequences *in vivo*. We therefore employed ChIP assays using the GATA-2 locus, the best characterized example of a target gene being repressed by GATA-1 in a FOG-1 dependent manner[22,24,26]. We had previously found that the -2.8kb region upstream of the GATA-2 locus was enriched for GATA-1 and FOG-1 binding (Fig. 1A). Significantly, the same sequence was also enriched for MBD2 binding (Fig. 1A). Similar results were obtained with an antibody against Mi-2β, another component of the MeCP1 complex (Fig. 1B). No binding of TAL-1 or Gfi-1b was observed in any of the GATA-2 sequences. Interestingly, FOG-1 and MBD2 were shown also to bind to the -3.4kb region, but not to the -4.2kb and -2.2kb flanking sequences (Fig. 1A and 1B). This suggests that the FOG-1/MeCP1 binding at -3.4kb may reflect a localized spreading of these proteins over a few nucleosomes to sequences upstream of the -2.8kb element, or that they were accidentally crosslinked to neighbouring DNA. The latter possibility would suggest that the FOG-1/MeCP1 complex is closer to the upstream sequences around the GATA binding sites (see also below).

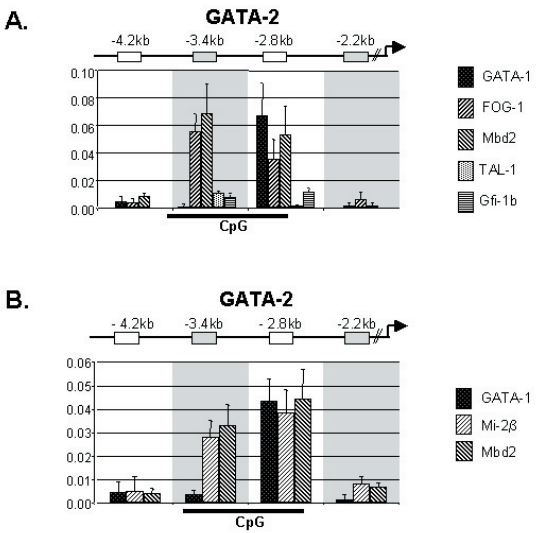


Figure 1: Binding of GATA-1 repressive complexes to target genes by ChIP assays in induced MEL cells.
Binding patterns of GATA-1, FOG-1, MBD2, TAL-1 and Gfi-1b (A) and of MBD2 and Mi2 (B) to the -2.8kb element of the GATA-2 locus. GATA-1 enriches the –2.8 kb region of the GATA-2 locus, with FOG-1 and MeCP1 components. MeCP1 presents a narrow spreading around the -2.8 kb region. None of the other antibodies were shown to enrich sequences tested. Relative enrichment has been normalized to input and corrected for background binding of species- and isotype-matched immunoglobulins. Antibodies: GATA-1, N6 (Santa Cruz); FOG-1 as in Tsang et al. (1997); MBD2 S923 [63]; TAL-1 as in Porcher et al. (1996); Gfi-1b, D19 (Santa Cruz), Mi2 (Paul Wade).

As MeCP1 contains methyl-binding protein, we investigated the methylation status of the GATA-2 locus around the MeCP1 binding region. A 3.9kb EcoRI DNA fragment was digested

either with XmnI as a control for complete DNA digestion or by EagI that is methylation sensitive (Fig 2). EagI maps to the GATA binding site in the -2,8kb element of the GATA-2 locus. A 1.2kb Apa I probe (solid line) detects 1kb and 2.1kb fragments after Eag I digestion and 1.7kb and 2.2kb fragments after Xmn I digestion. Eag I digests completely DNA of both induced and non-induced MEL cells, suggesting that its consensus sequence is not affected by DNA methylation despite the fact that MBD2 is a methyl DNA-binding protein (Fig. 3)[27]. However this does not exclude the possibility of highly localized methylation to specific CpG residues elsewhere in the GATA-2 locus. Thus, considering that GATA-1 binding is essential for GATA-2 repression[24], our findings strongly suggest that GATA-1, FOG-1 and MeCP1 form the repressive complex responsible for GATA-2 silencing.

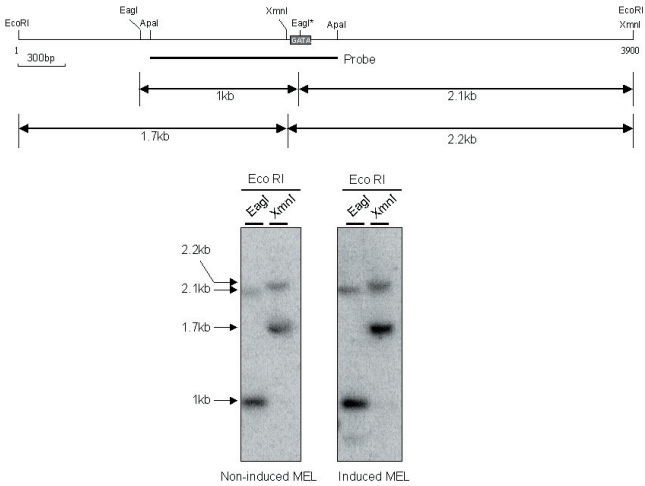


Figure 2: DNA methylation assay by restriction enzyme digestion at the -2.8kb element of the GATA-2 locus.

Genomic DNA (5-10μg) from non-induced (left panel) and induced (right panel) MEL cells was first digested with Eco RI, which releases a 3.9kb fragment, followed by digestion with Eag I (methylation sensitive, indicated by asterisk), or Xmn I as control for complete digestion of the genomic DNA samples. Eag I maps close to the GATA-1 binding sites in the -2.8kb element of the GATA-2 locus and within the PCR fragment (grey box, not to scale) amplified in the ChIP assays shown in Figures 5 and 6. Digested DNA was blotted and probed with a 1.2kb Apa I fragment (solid line) which detects 1kb and 2.1kb fragments on the Eag I digests and 1.7kb and 2.2kb fragments on the Xmn I digests. For both digests, the larger fragments are weaker due to their limited overlap with the probe. It can be seen that Eag I digests completely in DNA of both induced and non-induced MEL cells, thus suggesting that its recognition site is not affected by DNA methylation.

We extended our ChIP assays to investigate the binding of GATA-1 repressive complexes to other genes that are GATA-1 targets. In eosinophils, ectopic expression of FOG-1 results in the downregulation of eosinophilic GATA-1 target genes and the reprogramming of these cells towards an earlier, less differentiated cell type which may represent a common progenitor for the erythroid/megakaryocytic and eosinophilic lineages[28]. We thus reasoned that eosinophilic GATA-1 target genes, like the major basic protein (MBP)[29] which is inactive in erythroid cells[20], may be suppressed by the GATA-1/FOG-1/MeCP1 complex. We tested this

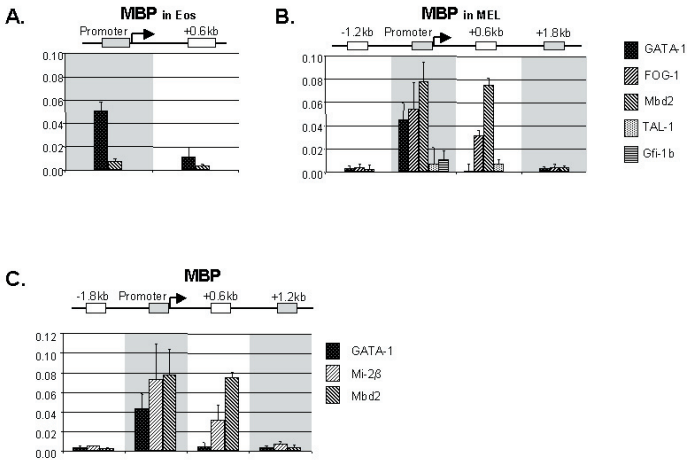


Figure 3: Binding of GATA-1 repressive complexes to target genes by ChIP assays in induced MEL cells.

Binding patterns of GATA-1, FOG-1, MBD2, TAL-1 and Gfi-1b to the MBP promoter in eosinophils (A) and in MEL cells (B). Binding of MBD2 and Mi2 to the MBP (C) promoter. GATA-1 was shown to bind to the MBP promoter in eosinophils as well as in erythroid cells. In erythroid cells, GATA-1 binds with MeCP1, whereas TAL-1 and Gfi-1b did not show any enrichment. Relative enrichment has been normalized to input and corrected for background binding of species- and isotype-matched immunoglobulins. Antibodies are similar to the ones used for Fig 1.

hypothesis by ChIP in induced MEL cells using as control chromatin from mouse eosinophils where MBP is expressed[30]. As expected, the promoter of the MBP gene was bound by GATA-1 in eosinophils (Fig. 3A). Importantly, GATA-1 was also bound to the inactive MBP promoter in induced MEL cells (Fig. 3B). FOG-1 and Mbd2 were also bound to the MBP promoter in MEL cells but not in eosinophils (Fig. 3B), consistent with the prediction above. Similar results were also obtained with an antibody against Mi-2β (Fig. 3C). Again, no TAL-1 of Gfi-1b binding was detected in the MBP promoter (Fig. 3B). Strikingly, in MEL cells, we found binding of the FOG-1 and MeCP1 complex, but not of GATA-1, to the +0.6kb sequence located close to the MBP promoter but not to other sequences located further upstream (-1.8kb) or downstream (+1.2kb) of the promoter (Fig. 3B and C). This observation is similar to that seen at the GATA-2 -2.8kb element.

We also investigated whether the interleukin-5 receptor α (IL-5Rα) gene is a repressed GATA-1 target gene in erythroid cells. IL-5Rα is a key regulator in the specification of the eosinophilic lineage and is involved in allergic responses[31]. The analysis of the upstream region of the IL-5Rα gene revealed the presence of several transcription factors binding sites, including GATA-1[32]. We thus reasoned that the eosinophilic IL-5Rα gene might also be down regulated by the GATA-1/FOG-1/MeCP1 complex in erythroid cells. Indeed, we detected binding of GATA-1 at the promoter of the IL-5Rα gene in erythroid cells (Fig. 4A). Significantly, FOG-1, Mi-2 and MBD2 were also detected binding to the IL5Rα promoter (Fig. 4B). These observations are consistent with the notion that the GATA-1/FOG-1/MecP1 complex is involved in the repression of alternative hematopoietic programs, such as the eosinophil program in erythroid cells. Thus,

GATA-1 may regulate diffrerent programs in the context of different complexes.

The Ets family hematopoietic transcription factor PU.1 is expressed in both myeloid and lymphoid progenitors[33] and the analysis of PU.1 knockout mice has shown that it is essential for the differentiation of both lineages[34]. Several lines of evidence have shown GATA-1 and PU.1 to be functionally antagonistic. GATA-1 was shown to bind to PU.1 in myeloid cells[35]. This interaction prevents recruitment of c-jun by PU.1 and therefore inhibits the ability of PU.1 to activate myeloid target genes. Similarly, in erythroid cells, PU.1 was shown to inhibit GATA-1-mediated transcriptional activation by interacting with GATA-1[36]. Ectopic expression of PU.1 in erythroid cells blocks terminal erythroid differentiation[37]. Thus we reasoned that the transcription factor PU.1, which would need to be repressed during erythroid differentiation, could also be a candidate for GATA-1 mediated repression. Indeed, we find that GATA-1 shows a clear enrichment for binding to the PU.1 promoter (Fig 4C). We also find enrichment for FOG-1, MBD2 and Mi-2 binding (Fig. 4D). We observe binding of the MeCP1 and FOG-1 also at the negative control sequence, suggesting a narrow spreading of the complex as observed for MBP and GATA-2. Although additional sequences in the PU.1 locus need to be investigated by ChIP assays, these preliminary data provide another example of a gene of an alternative lineage (the myeloid) that is repressed by the GATA-1/FOG-1/MeCP1 complex in erythroid cells.

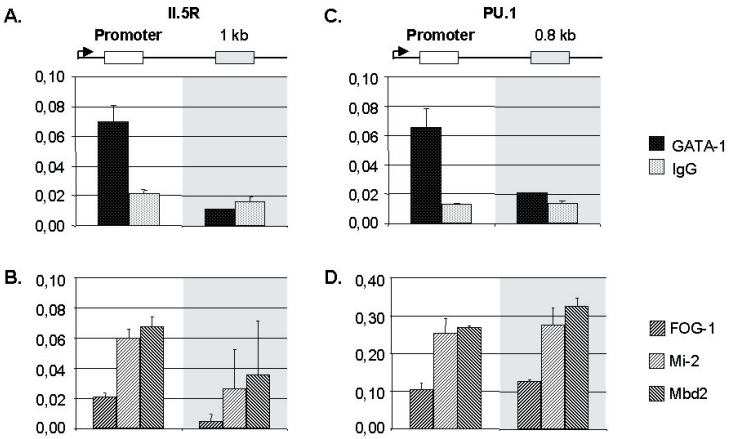


Figure 4: Binding of GATA-1 complexes to target genes by ChIP assays in erythroid cells.

The repressive GATA-1/FOG-1/MeCP1 complex binds to the promoters of the eosinophilic IL-5Rα (A and B) and of the myeloid PU.1 (C and D). Panels A and C represent relative enrichment for GATA-1 and IgG isotypes to input. Relative enrichment has been normalized to input and corrected for background binding of species- and isotype-matched immunoglobulins for panels B and D. Antibodies are similar to the ones used for Fig 1.

We next tested the c-myc and c-myb genes which are down-regulated with MEL cell differentiation[38,39]. Repression of the c-myc and c-myb genes has been linked to the proliferation arrest that accompanies terminal erythroid differentiation. The c-myc gene has also been shown to be a GATA-1 target gene in G1E cells[20,25]. We found GATA-1 binding to both promoters in induced MEL cells but we could not detect binding of FOG-1 or MBD2 to the same sequences (Fig. 5A and B). By contrast, Gfi-1b (absent from all other genes tested) was found binding to

both promoters (Fig. 5A and B), thus suggesting a role for the GATA-1/Gfi-1b complex in the repression of genes associated with cell proliferation. This may explain the observations of rapidly proliferating Gfi-1b$^{-/-}$ immature erythroid precursors in colony assays[40] and of Gfi-1b overexpression inducing proliferation arrest and differentiation in erythroid progenitors[41].

We also tested other genes to investigate functions of GATA-1 in apoptosis and cell survival. Bcl-X_L is a member of the Bcl2 family and has been associated with anti-apoptotic functions, also in erythropoiesis[42]. Bcl-X_L has been shown to be up-regulated during erythroid differentiation, this being dependent on Epo signaling[43]. Expression analysis of Bcl-X_L in G1E cells showed that GATA-1 is also required for Bcl-X_L up-regulation[42]. In addition, differentiation of Bcl-$X_L^{-/-}$ ES cells *in vitro* is characterized by a block of maturation of definitive erythroid precursors that die from apoptosis, an observation similar to that of the GATA-1 null ES cells[42]. Taken together, these observations suggest a regulatory role for GATA-1 in Bcl-X_L expression. Furthermore, the $p16^{INK4a}$ tumor suppressor that acts as an inhibitor of cyclin kinases such as CDK4 and CDK6[44] which phosphorylate retinoblastoma, was shown to be up-regulated in GATA-1 infected erythroid cells[45]. We tested GATA-1 binding at both Bcl-X_L and $p16^{INK4a}$ promoters but failed to detect any GATA-1 binding (data not shown), suggesting a possible indirect regulating effect of GATA-1 on those genes. For example, up-regulation of Bcl-X_L during erythroid differentiation is a late event compared to expression of GATA-1[42].

GATA-1 and TAL-1/Ldb1 are bound to an activated erythroid gene *in vivo*

Finally, we also tested the EKLF gene as an example of a gene that is activated during erythropoiesis. The EKLF enhancer sequence contains a GATA-E-box motif[46] which is bound *in vivo* by GATA-1 independently of FOG-1[22]. Strong GATA-1 binding and a clear enrichment for TAL-1 binding was indeed detected at the EKLF enhancer (Fig. 5C), thus providing a clear demonstration for the alternative (activating) GATA-1 complex with TAL-1 binding to a target gene *in vivo*. This may be related to the low level of HDAC activity associated with the TAL-1 immunoprecipitate (Chapter 2, Fig. 6A). No significant binding of FOG-1, MBD2 or Gfi-1b could be detected in the EKLF enhancer sequences (Fig. 5C).

Our analysis of the GATA-1/ACF/WCRF complex by ChIP, or any other, assays has been hindered by the quality of ACF/WCRF reagents available to us, hence it is presently not known whether the GATA-1 and ACF/WCRF complex binds to active or repressed genes.

GATA-1/FOG-1/MeCP1 binding in the GATA-2 gene locus

The binding of FOG-1 and MeCP1 to sequences adjacent to a GATA-1 binding site in the GATA-2 gene locus suggested that these proteins might spread along the DNA after GATA-1 mediated binding. We tested this possibility using ChIP analysis across the GATA-2 locus (Fig 1D and E, and Fig 6). The results showed that there is no binding of MeCP1 across the GATA-2 locus, suggesting that the spreading of the complex in differentiated MEL cells is restricted to sequences adjacent to the GATA binding sites of the -2.8kb region. This might also reflect the highly localized 3D structure of the locus where the two sequences that are showing enrichment are closer in a spatial organization than on a linear configuration. Within the GATA-2 gene locus, MeCP1 binding was detected only at a region neighbouring the -2.8 kb element. In contrast, GATA-1 was shown to bind, in addition to the -2.8kb region[26], to a sequence containing a GATA binding site located -3.9 kb upstream of the promoter. Both -2.8kb and -3.9kb elements coincide with the DNase I Hypersensitive sites in the active GATA-2 locus[47] (Fig. 6).

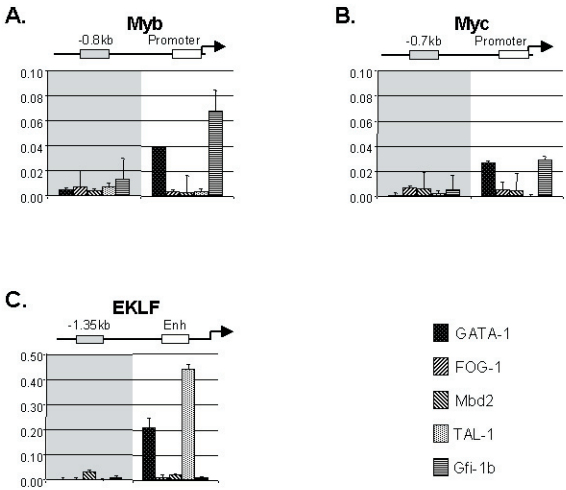


Figure 5: Binding of GATA-1 repressive and activating complexes to target genes by ChIP in induced MEL cells.

Binding patterns of GATA-1, FOG-1, MBD2, TAL-1 and Gfi-1b the myb (A) and myc (B) promoters and at the EKLF upstream enhancer (C). GATA-1 binds to the repressed c-myc and c-myb promoter genes with Gfi-1b, whereas GATA-1 and TAL-1 bind to the GATA-E-box motif of the EKLF enhancer. Relative enrichment has been normalized to input and corrected for background binding of species- and isotype-matched immunoglobulins. GATA containing sequences are indicated in bold.

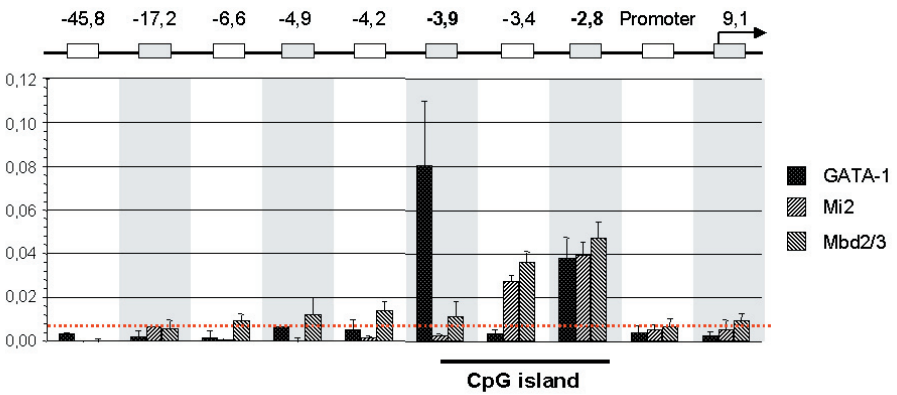


Figure 6: ChIP analysis of GATA-1 and MeCP1 (Mi2 and MBD2/3) across the GATA-2 locus.

GATA-1 is binding specifically at two regions defined as DNase Hypersensitive sites in the active locus, at -2.8 kb and -3.9 kb. The -2.8 kb region shows an enrichment for binding of FOG-1 and of the MeCP1 complex with a narrow spreading. In contrast, the -3.9 kb is not enriched by MeCP1 complex, suggesting that GATA-1 is bound to the GATA-2 locus as two independent protein complexes. Relative enrichment has been normalized to input and corrected for background binding of species- and isotype-matched

GATA-1 represses GATA-2 expression through the recruitment of FOG-1 and MeCP1

In order to confirm that GATA-2 repression during erythroid differentiation is specifically due to GATA-1 recruiting FOG-1 and the MeCP1 complex, we took advantage of the GATA-1 null G1E proerythroblastic cell line. These cells are derived from *in vitro* differentiated GATA-1 null ES cells and can undergo terminal differentiation only upon restoration of GATA-1 expression[48]. We used two G1E cell lines. The first one expresses wild type GATA-1 fused to an estrogen receptor (ER) ligand binding domain (GATA-1-ER) which can mediate terminal erythroid differentiation upon induction by estradiol[49]. The second cell line expresses ER fused to a mutant GATA-1 form bearing a single V205M amino acid substitution in the GATA-1 N-terminal zinc finger. Whilst not affecting GATA-1 DNA binding, this mutant abrogates interaction with FOG-1 and fails to rescue differentiation of G1E cells[50,51]. We first determined that repression of the GATA-2 gene in G1E cells was absolutely dependent on GATA-1 being capable of interacting with FOG-1 (Fig. 7A). We next tested by ChIP whether interaction of GATA-1 with FOG-1 binding at -2.8kb was responsible for the recruitment of MeCP1 to this sequence and to the neighboring -3.4kb sequence (Fig. 1A). As control, we also tested the more distal -4.2kb sequence, which did not show binding for any of these factors (Fig. 1A). In agreement with the MEL data, GATA-1, FOG-1 and Mi-2β were bound to the -2.8kb and to the -3.4kb sequence (for FOG-1 and Mi-2β) in differentiated GATA-1-ER cells (24 hours after induction with estradiol), albeit at lower levels compared to MEL cells (Fig. 7B, left panels). By contrast, in the GATA-1(V205M)-ER expressing cells GATA-1 was bound to the -2.8kb sequence, but no binding of FOG-1 and Mi-2β to the -2.8kb or -3.4kb sequences was detected (Fig. 7B right panels). We conclude that FOG-1 and the MeCP1 complex are specifically recruited by GATA-1 to the GATA-2 locus and are responsible for GATA-2 repression in terminal erythroid differentiation.

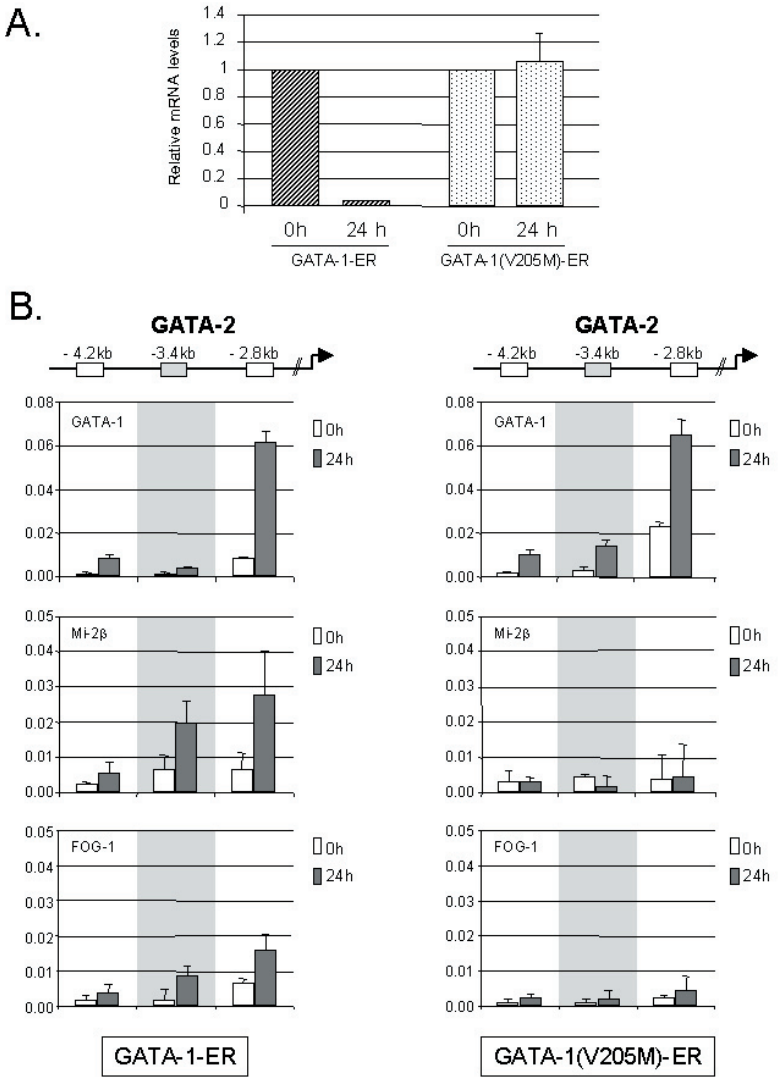


Figure 7: Silencing of GATA-2 requires recruitment of FOG-1 and MeCP1 by GATA-1.
(A) The V205M GATA-1 mutation fails to repress GATA-2. GATA-2 mRNA was measured by Real-Time PCR in G1E GATA-1-ER and G1E GATA-1(V205M)-ER cells before (0 hours) and after 24 hours of estradiol induction. Expression at 0 hour was normalized against GAPDH expression and set as 1. (B) ChIP to show binding of GATA-1, Mi-2β and FOG-1 in G1E GATA-1-ER (left panels) and in G1E GATA-1(V205M)-ER (right panels) at time 0 and 24 hours of estradiol induction. Relative enrichment has been normalized to input and corrected for background binding of species- and isotype-matched immunoglobulins. Data represent an of two independent IPs and three PCRs with duplicate samples. Antibodies used were as in the legend of Figure 5, except Mi-2β antibody[66].

Discussion

We provide here evidence for the *in vivo* binding of the five distinct GATA-1 complexes in erythroid cells. We show that the repression of the early hematopoietic gene GATA-2, which is mediated by GATA-1 and FOG-1, is due to the recruitment of the MeCP1 complex. In addition, we provide evidence for the *in vivo* binding of this complex to the repressed eosinophilic MBP and IL-5Rα genes and the myeloid PU.1 gene. Significantly, we also show the binding of GATA-1 in complex with TAL-1 to the active erythroid specific EKLF gene. Finally, we also show binding of the GATA-1/Gfi-1b complex to genes associated with cell proliferation functions, which become repressed with late stages of erythroid differentiation. These data strongly suggest that GATA-1 carries out different functions by targeting different genes through different complexes in erythroid cells.

GATA-1 complexes and target genes

We have identified the mechanism by which GATA-2 gene locus is repressed by GATA-1 during erythroid differentiation. GATA-2 gene expression was shown to be repressed by GATA-1 in a FOG-1 dependent manner, and GATA-1 was identified to bind at the -2.8kb region of the GATA-2 locus[24,26]. Here, we show in erythroid cells that GATA-1 recruits MeCP1 to the -2.8kb region through its interaction with FOG-1. Identification of MeCP1 as a co-factor of the GATA-2 gene repression is in agreement with the wave of deacetylation observed after GATA-1 binding, since MeCP1 carries histones deacetylase activities[26]. In addition, we have shown in chapter 2 that GATA-1 was associated with histone deacetylase activity. We show that GATA-1/FOG-1/MeCP1 is used in erythroid cells to repress early hematopoietic programs for the proper erythroid differentiation.

Determining cell fate is a complex process in which transcription factors play an essential role on the concomitant gene activation and repression. We found three non-erythroid lineage genes to be bound specifically by the repressive GATA-1/FOG-1/MeCP1 complex in erythroid cells. The MBP and IL-5Rα genes are expressed in eosinophils. MBP is activated by GATA-1[52], and the promoter of IL-5Rα contains GATA binding sites[53]. The fact that the GATA-1/MeCP1 complex is bound to these repressed genes in erythroid cells suggests that this complex acts early in erythroid cell differentiation to suppress alternative hematopoietic programs. The GATA-1 binding to the PU.1 promoter is also of interest since GATA-1 and PU.1 seem to have antagonistic functions in differentiation. For example, overexpression of PU.1 in MEL cells leads to a block of differentiation at the proerythroblast stage, similar to the GATA-1 knockout phenotype[54]. In addition, these two factors were shown to physically interact, perhpas affecting the DNA binding properties of the individual factors[36]. This direct interaction was thought to be sufficient to explain the down regulation of GATA-1 expression in myeloid cells, and conversely the down regulation of PU.1 in erythroid cells[55,56]. Here we show that in addition to the previously described GATA-1/PU.1 protein interactions, GATA-1 also regulates PU.1 gene expression directly as it binds to its promoter. However, we also detected MeCP1 binding to a region adjacent to this GATA site. Additional DNA sequences of the PU.1 gene locus must be examined to demonstrate specificity of the MeCP1 binding. These results suggest that the balance between myeloid and erythroid lineage is very tightly regulated by the fine-tuning of key transcription factor expression, such as GATA-1 and PU-1. We can speculate that the GATA-1/FOG-1/MeCP1 complex functions to suppress genes that need to be switched off early in hematopoiesis for erythroid differentiation, for example eosinophilic and myeloid lineages

associated genes.

DNA methylation is associated with gene repression. MeCP1 contains proteins that bind methylated DNA. Interestingly, FOG-1/MeCP1 binding at the GATA-2 and MBP loci show a narrow spreading into sequences adjacent to the GATA-1 binding sites, indicating that the chromatin complex can possibly spread along the chromatin. Alternatively, this enrichment might also reflect a cross-linking artefact due to the size of the complex that could be cross-linked to very close sequences. Other MeCP1 components such as Mi2-β and HDAC1 exhibit similar profiles at the same site (this thesis and [57]). Finding MeCP1 bound to GATA-2, MBP, IL-5Rα and PU.1 genes suggest therefore that these loci maybe methylated. However, we did not detect any methylation in the GATA-2 -2.8kb element in non-differentiated or differentiated MEL cells by southern blot. An analysis of CpGs in the GATA-2 locus has revealed the presence of a typical CpG island (CpG islands are usually located around the promoters of housekeeping genes) in the region between -2.2kb and -4.2kb of the GATA-2 locus (Fig. 6). It is very intriguing to detect MeCP1 binding in the -2.8kb region, since this is the first time to our knowledge that MeCP1 (and MBD2) is bound to non-methylated DNA *in vivo*. In contrast to our results, Hong and colleagues have not detected the presence of the MBD2 protein during purification of the FOG-1 complex, which led them to suggest that FOG-1 interacts with the NuRD complex instead of the MeCP1 complex. However, we have unequivocally identified MBD2 by mass spectrometry and also using a MBD2 specific antibody which detected interaction with GATA-1 by immunoprecipitations (data not shown) and binding to target genes. Despite these differences we would propose that GATA-1 uses the NuRD/MeCP1 complex to repress genetic programs that are “primed” in early stages of hematopoiesis. As this complex carries histone deacetylase activity, we can imagine modification of the chromatin structure that leads to repression throughout cell differentiation.

Loss of function of Gfi-1b in hematopoiesis has highlighted an essential role in megakaryocytic and erythroid linages[40]. The similar phenotypes of the Gfi-1b and GATA-1 knockouts suggest that they may act in a common pathway. Several lines of evidence have suggested that GATA-1 and Gfi-1b might be involved in cell cycle regulation. GATA-1 interacts with the retinoblastoma protein and GATA-1 overexpressing cells fail to arrest at G1[58]. Interestingly, Gfi-1b was shown to inhibit myeloid cell differentiation by repressing the cyclin kinase inhibitor $p21^{WAF1}$, [59]. The c-myc and c-myb genes have been reported to be down-regulated in GATA-1-mediated differentiation of erythroid cells[20]. The binding of GATA-1 and Gfi-1b to the repressed c-myc and c-myb genes provide important clues as to how these genes may regulate cell cycle and cell proliferation in erythroid differentiation. Gfi-1b contains a SNAG domain that has been linked to gene repression[59], for example by recruiting HDAC containing complexes[60]. It is possible that GATA-1 represses target genes *via* Gfi-1b recruiting co-repressors. Interestingly, we did not find Gfi-1b to bind to the GATA-2 locus, suggesting a tight regulation in the specific recruitment of complexes to different subsets of gene targets. However, previous studies have suggested a repression of c-myc by GATA-1 and FOG-1[61]. We therefore cannot exclude an indirect effect of GATA-1/FOG-1 on c-myc expression. A second discrepancy comes from ChIP analysis showing FOG-1 binding to the c-myc promoter in erythroid cells and recruitment of the NuRD complex[57]. This is in contrast to our data where we have failed to detect binding of FOG-1 despite using two different FOG-1 antibodies, or of MeCP1 members to the c-myc promoter.

GATA-1 complexes and erythropoiesis

An important aspect in hematopoietic differentiation, is the suppression of alternative “primed” lineage transcription programs and genes that maintain multipotentiality, while

upregulating genes associated with the differentiated cell type[7,9]. In addition, erythroid terminal differentiation is accompanied by cell cycle arrest. GATA-1 has been implicated in the regulation of most of these aspects[62]. In fact, a recent microarray analysis of GATA-1-dependent erythroid terminal maturation revealed an early wave of repression of genes like GATA-2, c-myc and c-myb, followed by the upregulation of erythroid specific genes[20]. Here we identified two GATA-1 repressive complexes acting on distinct sets of genes. We suggest that the GATA-1/Gfi-1b complex acts early and suppresses genes involved in cell proliferation, e.g. c-myc and c-myb. The GATA-1/FOG-1/MeCP1 complex also acts early to suppress genes required to maintain the "primed" multipotential state, e.g GATA-2 and alternative hematopoietic lineage genes, MBP, PU.1 and IL-5Rα (Fig. 1 and 2). In contrast, the GATA-1/FOG-1 and the GATA-1/TAL-1/Ldb1 complexes would play a major role in the later upregulation of erythroid genes (Fig. 4).

In conclusion, we suggest that GATA-1 provides specific early versus late differentiation functions in the context of distinct complexes (Fig. 8). The model of different GATA-1 complexes executing specific tasks in different stages of erythroid differentiation suggests a dynamic aspect in the GATA-1 complex interactions during differentiation and also raises the prospect of dissecting the contribution of distinct GATA-1 interactions in erythropoiesis (i.e. essential versus dispensable) by selectively manipulating a specific GATA-1 complex at a time.

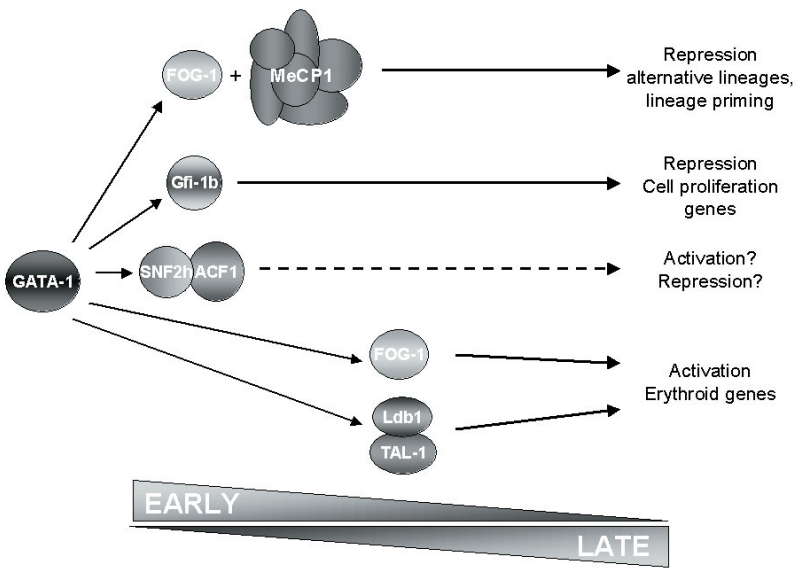


Figure 8: Model for the distinct GATA-1 complexes and their role in erythropoiesis. *Broken arrow indicates unknown function and timing. See text for explanation.*

Material and methods

ChIP assays. Preparation of crosslinked chromatin ($2x10^7$ induced MEL cells treated with 0.4% formaldehyde for 10 minutes at room temperature), sonication to 300-800 base pair fragments and immunoprecipitations were as described in the Upstate protocol (www.upstate.com). Anti-GATA-1 protein-DNA immunocomplexes were immunoprecipitated in an additional step with an AffiniPure rabbit anti-rat antibody (Jackson ImmunoResearch Laboratories, West Grove, PA). Eosinophilic chromatin was prepared as previously described[30]. At least two independent chromatin immunoprecipitations were carried out per experiment. Antibodies used : GATA-1, N6 (Santa Cruz); MBD2, S923 sheep polyclonal[63] and rabbit polyclonal anti MBD2/3 antibody (Upstate 07-199); FOG-1 rabbit polyclonal[49]; TAL-1 rabbit polyclonal[64]; Gfi-1b D19 goat polyclonal (Santa Cruz sc-8559).

Real time PCR. Quantitative RealTime PCR (Opticon I, MJ Research) was done using SYBR Green I. PCR primers were designed by Primer Express 2.0 (PE Applied Biosystems). The qPCR Core Kit (Eurogentec, Belgium) was used with 400 nM of each primer under the following cycling conditions: 95^0C for 10 minutes, 40 cycles of 30 seconds at 95^0C, 60 seconds at 60^0C, 15 seconds at 75^0C. Enrichment for a specific DNA sequence was calculated using the comparative C_T method [65]. PCR primer sequences are provided in Supplementary Materials and Methods.

ChIP PCR primer sequences:

EKLF upstream enhancer forward PCR primer: 5'-CTGGCCCCCCTACCTGAT-3'
EKLF upstream enhancer reverse PCR primer: 5'-GGCTCCCTTTCAGGCATTATC-3'

EKLF -1.35kb forward PCR primer: 5'-TGCTCCCCACTATGATAATGGA-3'
EKLF -1.35kb reverse PCR primer: 5'-GCCACAACCAAAGAAGACATTTT-3'

MBP -1.2kb forward PCR primer: 5'-GGGTCTAATTCCGAGGGTGAGT-3'
MBP -1.2kb reverse PCR primer: 5'-GGCCTGGAAATCACTGAGCTA-3'

MBP promoter forward PCR primer: 5'-CCGCCAAGGTGTCTATAAATGC-3'
MBP promoter reverse PCR primer: 5'-TGGGTCTTGTCAAGTTTGCAAA-3'

MBP +0.6kb forward PCR primer: 5'-GAAGTAGAGGCAGGATAATCAGGAA-3'
MBP +0.6kb reverse PCR primer: 5'-AGGATGAACCAGGGCTAATGC-3'

MBP +1.8kb forward PCR primer: 5'-TGTGACAGACGTGGACCTTCA-3'
MBP +1.8kb reverse PCR primer: 5'-TGCATCCAGAGTCACCCATAAG-3'

GATA2 -4.2kb region forward PCR primer: 5'-GAATTTCCTGCCGGTCCAT-3'
GATA2 -4.2kb region reverse PCR primer: 5'-GACGCGTTGGCTTTGTGTG-3'

GATA-2 -3.4kb forward PCR primer: 5'-TCCATCCAGCAGCTTTAGGAA-3'
GATA-2 -3.4kb region reverse PCR primer: 5'-GGGTTCGAAGCCACTCCAA-3'

GATA-2 -2.8kb region forward PCR primer: 5'-CCGGGCAGATAACGATTGG-3'
GATA-2 -2.8kb region reverse PCR primer: 5'-TTCATCTCGGCCGGCTAAT-3'

GATA-2 -2.2kb region forward PCR primer: 5'-AGGACCCCCTGCTTCTTGTTTC-3'
GATA-2 -2.2kb region reverse PCR primer: 5'-GGCAGTATGAGGCCCAGAATCTT-3'
Myb promoter forward PCR primer: 5'-GGGCGCCAGATTTGG-3'
Myb promoter reverse PCR primer: 5'-GGAGGAAACAGGTTGATATTAAAGT-3'

Myb -0.8kb forward PCR primer: 5'-GTAGGTTTGTCCAGCAAGTGTTTG-3'
Myb -0.8kb reverse PCR primer: 5'-AGGTGCCTACCACGCACTTCT-3'

C-myc promoter forward PCR primer: 5'-CCAGACATCGTTTTTCCTGCATA-3'
C-myc promoter reverse PCR primer: 5-'CCGCTCAGTGTGTGGAGTGATA-3'

C-myc -0.7kb forward PCR primer: 5'-ACACACACATACGAAGGCA-3'
C-myc -0.7kb reverse PCR primer: 5'-ACCGTTAACCCCTTCCTCCC-3'

PU.1 promoter forward PCR primer: 5'-CCAGGGCTGCCCTTTGA-3'
PU.1 promoter reverse PCR primer:5'-TTGCATAAATCTCTTGCGCTACA-3'

PU.1 -0.8kb forward PCR primer: 5'-GCATCTGGTGGGTGGACAAG-3'
PU.1 -0.8kb reverse PCR primer:5'- GCGCGCCATCTTCTGGTA-3'

Il-5R promoter forward PCR primer: 5'-GCAGGTGTCTTCTTAACCATGACA-3'
Il-5R promoter reverse PCR primer: 5'-GGGTGCACTGGTGTTTGACTT-3'

Il-5R -1 kb forward PCR primer: 5'-CTGAACCAAACCTGAACCCAT-3'
Il-5R -1kb reverse PCR primer: 5'-CACCTGCATGTCTGATTTGTC-3'

PCR primers for RNA analysis in G1E cells

GATA-2 mRNA primers:
Exon III forward PCR primer: 5'-ACTATGGCAGCAGTCTCTTCCATC-3'
Exon V reverse PCR primer: 5'-AAGGTGGTGGTTGTCGTCTGAC-3'

Antibodies

The following antibodies were purchased from Santa Cruz Biotechnology (Santa Cruz, CA): N6 GATA-1 rat monoclonal (sc-265); Gfi-1b D19 (sc-8559)
Rabbit polyclonal antibodies against TAL-1 were a generous gift by Richard Baer (Columbia University, NY) and by Catherine Porcher (WIMM, Oxford). Rabbit polyclonal antibodies against Ldb1 were kindly donated by Gordon N. Gill (Stanford University, CA). Rabbit polyclonal antibodies against FOG-1 were a generous gift by Stuart H. Orkin (Harvard Medical School, MA). Rabbit polyclonal antibodies against MTA2, Mi2 and MBD2/3 were kindly donated by Paul A. Wade (NIH/NIEHS, NC). Sheep anti-serum S923 against MBD2 was generously provided by Adrian Bird (University of Edinburgh, UK).

Acknowledgements

We are grateful to our many colleagues cited for generously sharing reagents. We are also grateful to Mitch Weiss (Children's Hospital of Philadelphia) for G1E cells, Sjaak Philipsen and Marieke von Lindern (Erasmus MC) for comments on the manuscript and Raymond Poot (CSC, London) for helpful discussions. We are grateful to Gerd Blobel (Children's Hospital of Philadelphia) for exchanging pre-publication data. This work has been supported by the Netherlands Research Organization (NWO), the European Union (grant HPRN-CT-2000-00078), the NIH (grant RO1 HL 073455-01) and the Netherlands Proteomics Center.

References

1. Sieweke, M. H. & Graf, T. A transcription factor party during blood cell differentiation. *Curr Opin Genet Dev* **8**, 545-51 (1998).
2. Cantor, A. B. & Orkin, S. H. Transcriptional regulation of erythropoiesis: an affair involving multiple partners. *Oncogene* **21**, 3368-76 (2002).
3. Hu, M. et al. Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev* **11**, 774-85 (1997).
4. Delassus, S., Titley, I. & Enver, T. Functional and molecular analysis of hematopoietic progenitors derived from the aorta-gonad-mesonephros region of the mouse embryo. *Blood* **94**, 1495-503 (1999).
5. Akashi, K. et al. Transcriptional accessibility for genes of multiple tissues and hematopoietic lineages is hierarchically controlled during early hematopoiesis. *Blood* **101**, 383-9 (2003).
6. Miyamoto, T. et al. Myeloid or lymphoid promiscuity as a critical step in hematopoietic lineage commitment. *Dev Cell* **3**, 137-47 (2002).
7. Enver, T., Heyworth, C. M. & Dexter, T. M. Do stem cells play dice? *Blood* **92**, 348-51; discussion 352 (1998).
8. Fisher, A. G. Cellular identity and lineage choice. *Nat Rev Immunol* **2**, 977-82 (2002).
9. Orkin, S. H. Diversification of haematopoietic stem cells to specific lineages. *Nat Rev Genet* **1**, 57-64 (2000).
10. Graf, T. Differentiation plasticity of hematopoietic cells. *Blood* **99**, 3089-101 (2002).
11. Pevny, L. et al. Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* **349**, 257-60 (1991).
12. Shivdasani, R. A., Fujiwara, Y., McDevitt, M. A. & Orkin, S. H. A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. *Embo J* **16**, 3965-73 (1997).
13. Yu, C. et al. Targeted deletion of a high-affinity GATA-binding site in the GATA-1 promoter leads to selective loss of the eosinophil lineage in vivo. *J Exp Med* **195**, 1387-95 (2002).
14. Kulessa, H., Frampton, J. & Graf, T. GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts. *Genes Dev* **9**, 1250-62 (1995).
15. Visvader, J., Elefanty, A., Strasser, A. & Adams, J. GATA-1 but not SCL induces megakaryocytic differentiation in an early myeloid line. *Embo J* **11**, 4557-64 (1992).
16. Yamaguchi, Y. et al. Mechanisms of transcription in eosinophils: GATA-1, but not GATA-2, transactivates the promoter of the eosinophil granule major basic protein gene. *Blood* **91**, 3447-58 (1998).
17. Seshasayee, D., Gaines, P. & Wojchowski., D. GATA-1 dominantly activates a program of erythroid gene expression in factor-dependent myeloid FDCW2 cells. *Mol Cell Biol.* **18**, 3278-88 (1998).
18. Heyworth, C., Pearson, S., May, G. & Enver, T. Transcription factor-mediated lineage switching reveals plasticity in primary committed progenitor cells. *Embo J* **21**, 3770-81 (2002).
19. Iwasaki, H. et al. GATA-1 converts lymphoid and myelomonocytic progenitors into the megakaryocyte/erythrocyte lineages. *Immunity* **19**, 451-62 (2003).
20. Welch, J. J. et al. Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood* **104**, 3136-47 (2004).
21. Anguita, E. et al. Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. *Embo J* **23**, 2841-52 (2004).
22. Letting, D. L., Chen, Y. Y., Rakowski, C., Reedy, S. & Blobel, G. A. Context-dependent regulation of GATA-1 by friend of GATA-1. *Proc Natl Acad Sci U S A* **101**, 476-81 (2004).
23. Lahlil, R., Lecuyer, E., Herblot, S. & Hoang, T. SCL assembles a multifactorial complex that determines glycophorin A expression. *Mol Cell Biol* **24**, 1439-52 (2004).
24. Pal, S. et al. Coregulator-dependent facilitation of chromatin occupancy by GATA-1. *Proc Natl Acad Sci U S A* **101**, 980-5 (2004).
25. Rylski, M. et al. GATA-1-mediated proliferation arrest during erythroid maturation. *Mol Cell Biol* **23**, 5031-42 (2003).

26. Grass, J. A. et al. GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. *Proc Natl Acad Sci U S A* **100**, 8811-6 (2003).
27. Zhang, Y. et al. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev* **13**, 1924-35 (1999).
28. Querfurth, E. et al. Antagonism between C/EBPbeta and FOG in eosinophil lineage commitment of multipotent hematopoietic progenitors. *Genes Dev* **14**, 2515-25 (2000).
29. Du, J. et al. Novel combinatorial interactions of GATA-1, PU.1, and C/EBPepsilon isoforms regulate transcription of the gene encoding eosinophil granule major basic protein. *J Biol Chem* **277**, 43481-94 (2002).
30. Guyot, B., Valverde-Garduno, V., Porcher, C. & Vyas, P. Deletion of the major GATA1 enhancer HS 1 does not affect eosinophil GATA1 expression and eosinophil differentiation. *Blood* **104**, 89-91 (2004).
31. Hirai, K. et al. Enhancement of human basophil histamine release by interleukin 5. *J Exp Med* **172**, 1525-8 (1990).
32. Imamura, F. et al. The murine interleukin-5 receptor alpha-subunit gene: characterization of the gene structure and chromosome mapping. *DNA Cell Biol.* **13**, 283-92 (1994).
33. Shivdasani, R. A. & Orkin, S. The transcriptional control of hematopoiesis. *Blood* **87**, 4025-39 (1996).
34. Scott, E., Simon, M., Anastasi, J. & Singh, H. Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science* **265**, 1573-7 (1994).
35. Zhang, P. et al. Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. *Proc Natl Acad Sci U S A* **96**, 8705-10 (1999).
36. Rekhtman, N., Radparvar, F., Evans, T. & Skoultchi, A. I. Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells. *Genes Dev* **13**, 1398-411 (1999).
37. Zhang, P. et al. PU.1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding. *Blood* **86**, 2641-8 (2000).
38. Lachman, H. M. & Skoultchi, A. I. Expression of c-myc changes during differentiation of mouse erythroleukaemia cells. *Nature* **310**, 592-4 (1984).
39. Chen, J. & Bender, T. P. A novel system to identify Myb target promoters in friend murine erythroleukemia cells. *Blood Cells Mol Dis* **27**, 429-36 (2001).
40. Saleque, S., Cameron, S. & Orkin, S. H. The zinc-finger proto-oncogene Gfi-1b is essential for development of the erythroid and megakaryocytic lineages. *Genes Dev* **16**, 301-6 (2002).
41. Garcon, L. et al. Gfi-1B plays a critical role in terminal differentiation of normal and transformed erythroid progenitor cells. *Blood* **105**, 1448-55 (2005).
42. Gregory T., Y. C., Ma A., Orkin S. H., Blobel G. A., and Weiss M. J. GATA-1 and Erythropoietin Cooperate to Promote Erythroid Cell Survival by Regulating bcl-xL Expression. *Blood* **94**, 87-96 (1999).
43. Gregoly, P. & Bondurant, M. The roles of Bcl-X(L) and apopain in the control of erythropoiesis by erythropoietin. *Blood* **90**, 630-40 (1997).
44. Sherr, C. & Roberts, J. CDK inhibitors: positive and negative regulators of G1-phase progression. *Genes Dev* **13**, 1501-12 (1999).
45. Pan X, O. O., Ohneda K, Lindeboom F, Iwata F, Shimizu R, Nagano M, Suwabe N, Philipsen S, Lim KC, Engel JD, Yamamoto M. Graded levels of GATA-1 expression modulate survival, proliferation, and differentiation of erythroid progenitors. *J Biol Chem* **280**, 22385-94 (2005).
46. Anderson, K. P., Crable, S. C. & Lingrel, J. B. Multiple proteins binding to a GATA-E box-GATA motif regulate the erythroid Kruppel-like factor (EKLF) gene. *J Biol Chem* **273**, 14347-54 (1998).
47. Martowicz, M., Grass, J., Boyer, M., Guend, H. & Bresnick, E. Dynamic GATA factor interplay at a multicomponent regulatory region of the GATA-2 locus. *J Biol Chem* **280**, 1724-32 (2005).
48. Weiss, M. J., Yu, C. & Orkin, S. H. Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. *Mol Cell Biol* **17**, 1642-51 (1997).
49. Tsang, A. P. et al. FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell* **90**, 109-19 (1997).

50. Crispino, J. D., Lodish, M. B., MacKay, J. P. & Orkin, S. H. Use of altered specificity mutants to probe a specific protein-protein interaction in differentiation: the GATA-1:FOG complex. *Mol Cell* **3**, 219-28 (1999).
51. Nichols, K. E. et al. Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. *Nat Genet* **24**, 266-70 (2000).
52. Yamaguchi Y, A. S., Minegishi N, Takiguchi M, Yamamoto M, Suda T. Mechanisms of transcription in eosinophils: GATA-1, but not GATA-2, transactivates the promoter of the eosinophil granule major basic protein gene. *Blood* **91**, 3447-58 (1998).
53. Imamura F, T. S., Akagi K, Ando M, Yamamura K, Takatsu K, Tominaga A. The murine interleukin-5 receptor alpha-subunit gene: characterization of the gene structure and chromosome mapping. *DNA Cell Biol.* **13**, 283-92 (1994).
54. Moreau-Gachelin, F., Ray, D., Tambourin, M. M. & Tavitian, A. The putative oncogene Spi-1: murine chromosomal localization and transcriptional activation in murine acute erythroleukemias. *Oncogene* **4**, 1449-56 (1989).
55. Nerlov, C. & Graf, T. PU.1 induces myeloid lineage commitment in multipotent hematopoietic progenitors. *Genes Dev* **12**, 2403-12 (1998).
56. McNagny, K., Sieweke, M., Doderlein, G., Graf, T. & Nerlov., C. Regulation of eosinophil-specific gene expression by a C/EBP-Ets complex and GATA-1. *Embo J* **17**, 3669-80 (1998).
57. Hong W, N. M., Chen YY, Kori R, Vakoc CR, Rakowski C, Blobel GA. FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. *Embo J* **24**, 2367-78 (2005).
58. Whyatt, D. et al. The level of the tissue-specific factor GATA-1 affects the cell-cycle machinery. *Genes Funct* **1**, 11-24 (1997).
59. Tong, B. et al. The Gfi-1B proto-oncoprotein represses p21WAF1 and inhibits myeloid cell differentiation. *Mol Cell Biol* **18**, 2462-73 (1998).
60. Peinado, H., Ballestar, E., Esteller, M. & Cano., A. Snail mediates E-cadherin repression by the recruitment of the Sin3A/histone deacetylase 1 (HDAC1)/HDAC2 complex. *Mol Cell Biol.* **24**, 306-19 (2004).
61. Crispino, J., Lodish, M., MacKay, J. & Orkin, S. Use of altered specificity mutants to probe a specific protein-protein interaction in differentiation: the GATA-1:FOG complex. *Mol Cell* **3**, 219-28 (1999).
62. Blobel, G. A. & Weiss, M. J. in *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management.* (eds. Steinberg, M. H., Forget, B. G., Higgs, D. R. & Nagel, R. L.) 72-94 (Cambridge University Press, Cambridge, 2001).
63. Ng, H. H. et al. MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat Genet* **23**, 58-61 (1999).
64. Porcher, C. et al. The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. *Cell* **86**, 47-57 (1996).
65. Litt, M. D., Simpson, M., Recillas-Targa, F., Prioleau, M. N. & Felsenfeld, G. Transitions in histone acetylation reveal boundaries of three separately regulated neighboring loci. *Embo J* **20**, 2224-35 (2001).
66. Fujita, N. et al. MTA3, a Mi-2/NuRD complex subunit, regulates an invasive growth pathway in breast cancer. *Cell* **113**, 207-19 (2003).

Chapter 5:

Novel binding partners of Ldb1 are required for hematopoietic development

Natalia Meier[1], Sanja Krpic[1], Patrick Rodriguez[1], John Strouboulis[1], Maria Monti[3], Jeroen Krijgsveld[3], Martin Gering[2], Roger Patient[2] , Frank Grosveld[1,4] and Arnd Hostert[1]

Submitted

[1]Department of Cell Biology and Genetics, ErasmusMC, PO Box 1738, 3000DR, Rotterdam, The Netherlands.
[2]Department of Haematology, The Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DU, United Kingdom
[3]Department of Biomolecular Mass Spectrometry, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Sorbonnelaan 16, 3584CA, Utrecht, The Netherlands

Novel binding partners of Ldb1 are required for hematopoietic development

Abstract

Ldb1, a ubiquitously expressed LIM domain binding protein, is essential in a number of tissues during development. It interacts with Gata1, Tal1, E2A and Lmo2 to form a transcription factor complex regulating late erythroid genes. We identify a number of novel Ldb1 interacting proteins in erythroleukemic cells, in particular the repressor protein Eto-2 (and its family member Mtgr1), the cyclin dependent kinase Cdk9 and the bridging factor Lmo4. Morpholino mediated knockdowns in zebrafish show these factors to be essential for definitive hematopoiesis. In accordance with the zebrafish results these factors are co-expressed in pre-hematopoietic cells of the early mouse embryo although we originally identified the complex in late erythroid cells. Based on the change in subcellullar localisation of Eto-2 we postulate that it plays a central role in the transition from the migration and expansion phase of the pre-hematopoietic cells to the establishment of definitive hematopoietic stem cells.

Introduction

Erythrocytes are derived from an infrequently dividing cell type, the hematopoietic stem cell (HSC)[1]. The first wave of hematopoiesis generates primitive erythrocytes From approximately E8.5 to E10.5 in the mouse definitive HSCs are derived from the aorta-gonado-mesonephros region (AGM)[2-6]. Recently the placenta has been identified as a further source of adult HSCs[7,8]. At E11 the mouse fetal liver becomes the main organ of hematopoiesis, later replaced by the bone marrow.

HSC differentiation involves co-ordinated and changing transcription, often by functionally conserved genes. For example, in mammals such a set of transcription factors (including Gata2, Tal1, Lmo2, Gata1 and Runx1/Aml-1) is required for the differentiation of HSCs[9].

Often the binding sites for transcription factor complexes are located at great distance from the genes that they control. In the human β-globin locus the interacting binding sites and genes are spread over a distance of 100kb. A three-dimensional structure resulting from long range interactions, the Active Chromatin Hub, has recently been demonstrated[10-13]. We anticipate that (novel) classes of proteins will establish such 3D structures by mediating interactions of protein complexes bound to distal regulatory elements. An example of a long range interaction protein is Chip/Ldb1[14], a protein that can interact with the insulator protein Su(Hw)[15]. Initially isolated in a screen for proteins that bind LIM domains[16], orthologs of Ldb1 have now been identified in a range of other species. Ldb1, a ubiquitously expressed nuclear protein, does not bind DNA but appears to participate in transcriptional control by acting as a co-factor for other proteins. It binds LIM-homeodomain (LIM-HD) and LIM-only (LMO) proteins via the C-terminal LIM Interaction Domain (LID)[15,17] and homodimerises via a Self Interaction Domain (SID) in the N-terminus and the Other Interaction Domain (OID)[15]. It is part of a protein complex in murine erythroid cells composed of the hematopoietic transcription factors Lmo2, Tal1, Gata1 and E2A[18] binding to a GATA-E box motif. This complex binds to the LCR and β globin promoter of murine erythroleukemic (MEL) cells[19], to the erythoid specific glycophorin A promoter[20] and to multiple sites in the α-globin locus during erythroid differentiation[21].

Correct transcriptional regulation by Ldb1 has been shown to be the result of an equilibrium of interactions between Ldb1 and its different binding partners. For example, in *D. melanogaster* Chip/Ldb1 and Apterous interact forming a functional complex where the relative levels of the two proteins are critical . This interaction can be modulated by dLMO, another interacting factor[24]. Consistent with its interaction with a broad range of transcription factors involved in development, the *Ldb1* knockout mouse dies between E9.5 and E10.5 of a series of developmental defects, including no hematopoieisis[25] and A. Hostert unpublished. The latter partly resembles the knockout phenotypes of the hematopoietic transcription factors Lmo2 and Tal1 [26-29].

In order to understand the role of Ldb1 in erythroid transcriptional interactions, we performed a biochemical screen to identify its binding partners. Using our in vivo biotinylation approach[30,31] we describe a number of novel partners. We show that Ldb1 forms complexes that change composition during MEL cell differentiation and that these complexes are bound to the target genes *in vivo*. We show that Ldb1 and its binding partners are co-expressed in situ at early stages of development of the hematopoietic system in the mouse embryo and that the novel erythroid binding partners are required for development of the definitive but not the primitive hematopoetic system of zebrafish embryos.

Materials and Methods

Ldb1 cDNA and bio-Ldb1 construct cloning

Ldb1 cDNA was cloned from D14.5 FvB fetal liver RNA (Trizol, Life Technologies) by RT-PCR (SuperScriptII RT, Invitrogen; Pfu Pol, Promega) into the EcoRI site of pBluescript (pBS) (construct AH-3).

ATG less Ldb-1 cDNA was amplified and cloned between the BamHI and NotI sites of pBS. The bio-tag was cloned directly in front of the ATG-less cDNAtagged cDNA was isolated as a XhoI–NotI fragment and cloned into the SalI and NotI sites of pEV-Neo (de Boer et al., 2003) to give construct pEV-Neo-bio-Ldb1.

MEL cell transfection and culture

C88 cells were cultured and induced for differentiation as described. $C88^{BirA}$ cells were transfected by electroporation with ScaI-linearised pEV-Neo-bioLdb1, cultured in 96 well plates containing medium with 1μg/ml puromycin and 0.8mg/ml neomycin to select single clones.

Nuclear extract preparation

Small scale nuclear extract preparation of MEL cell cultures (30-50ml) was prepared according to Andrews and Faller[78]. Nuclear extracts of larger cultures (5,5 to 9l) were prepared as described.

Strepatavidin Pulldown and Mass spectrometry

Strepatavidin coated Dynabeads M-280 (Dynal) were blocked for one hour with chicken serum albumin/PBS (200ng/μl). The salt and detergent concentrations of nuclear extract samples of 5-6mg from induced $C88^{BirA/bio-Ldb1}$ cells or 15mg from non-induced $C88^{BirA/bio-Ldb1}$ cells were adjusted to 200mM KCl and 0.3% NP40 with 10mM KCl buffer (10mM KCl, 10mM HEPES-KOH, 1,5mM $MgCl_2$, 25% glycerol, 0,75% NP40, 2mM PMSF) prior to overnight incubation with blocked beads at 4°C. The beads were washed 6 times 5 min each in washing solution (150/200mM KCl, 10mM HEPES-KOH pH 7.9, 1.5mM $MgCl_2$, 0.2mM EDTA, 0.3% NP40, 0.2mM PMSF) at RT. Pulled down proteins were processed and analyzed by mass spectrometry as described.

Immunoprecipitations and Western blot analysis

Immunoprecipitations were performed as described, Western blots as in. Bio-Ldb1 was detected using a 1/10000 dilution of Streptavidin-Horeseraddish peroxidase (HRP) conjugate (NEN).

Antibodies

Monoclonal rat antibodies against Mtgr1 and Lyl1 were produced by Absea. Antibody against E2-2 was obtained from Abcam (ab2233), all others from Santa Cruz Biotechnology: α-Ldb1 (sc-11198), α-Lmo4 (sc-11121, sc-22833), α-Eto-2 (sc-9741), α-Runx1 (sc-8563), α-Cdk9 (sc-484), α-HEB (sc-357), α-E2A (sc-349), α-Gata1-N6 (sc-265).

Chromatinimmunoprecipitations

Fixation, lysis of cells and sonication of chromatin were performed as described.

Primers for real time-PCR were as in. Primers for the Gata1 HS and negative control sequences were:

Gata1 HS-3.5 Rev: CCGGGTTGAAGCGTCTTCT

Gata1 HS-3.5 For: TCAGGGAAGGATCCAAGGAA

Gata1 Negrev: TGCCGCTTGCCTTTGTAAG
Gata1 Negfor: CACTAGCAGCTGGGTGGGTTA

Immunohistochemistry

E9.5 FvB mouse embryos were fixed in 2% paraformaldehyde/PBS for 2hrs at RT. After overnight equilibration in 20% Sucrose/PBS at 4^{O}C embryos were orientated, quick frozen in Tissue Tek (Sakura Finetek).

Immunohistochemistry was essentially carried out as described in [32]. All animal experiments were carried out according to the Dutch Welfare of Animals Act.

Zebrafish maintenance and Morpholino oligo injections

WT zebrafish were kept and staged according to . ATG morpholinos (Gene-Tools) were derived from the genebank cDNA of ldb1 (NM_131313, 5'-GCCCACGTCTCGGTCCAGCATGGTG-3'), tcf4 (NM_131259, 5'-AGCTGCGGCATTTTTCCCGAGGAGC-3'), e12 (X76997, 5'-GGTGCCC ACCGTCGCCATCCTGATC-3'), cdk9 (BC055634, 5'-CGACGCCATCGTAGTATTTGGACAT-3'), lmo4 (NM_177984, 5'-AGCTTTCCACACGACTGTTCACCAT-3'). For identification of the zebrafish Eto family homologues xblastn searches of the EST database using the murine protein sequences of Eto-2, Eto and Mtgr1 were performed. ATG-MOs were designed against the hypothetical translational start sites in EST AF164710 (5'- AACATGACGGTTGGAACTCTGGTT-3'), the orthologue of Eto-2 and EST CD053087 (5'-GACCGGGCATGGCAGGAACCTTCT-3') that showed similarity to the least conserved ETO family member Mtgr1.
All morpholinos were dissolved in water to a concentration of 1mM and injected at three doses (0.1nl, 0.5nl and 1.2nl) into zebrafish embryos at the 2-8 cell stage. As an injection control rhodamin-dextrane was added to a concentration of 10% vol/vol before use.

Whole mount in situ hybridization

Digoxigenin-UTP (Roche) labeled antisense and sense RNA probes against the zebrafish orthologues of embryonic beta globin, βE1 and runx1 (M. Gering) were synthesized from linearised plasmids using T3 and T7 RNA polymerases. In situ hybridization was performed as described by . Probes were detected by incubation with alkaline phophatase (AP) coupled anti-Digoxigenin antibody (Roche) and color reaction with Fast Red (Roche) for α-E1 and BM Purple (Roche) for runx1.

Results

Generation of bio-Ldb1 cells

In order to identify Ldb1 interacting partners, C88^{BirA} MEL cells expressing the *Escherichia coli* BirA protein-biotin ligase[30] were stably transfected with pEV-Neo-bio-Ldb1. Induction of MEL cells to differentiate upregulates β-globin. Since BirA and bio-Ldb1 are under the control of human β-globin LCR sequences, these C88$^{BirA/bioLdb1}$ cells also upregulate the biotin ligase and bio-tagged Ldb1 upon induction (Figure 1 A).

Sixteen C88$^{BIR/bioLdb1}$ clones were isolated and induced to differentiate. Nuclear extracts were tested for presence of the fusion protein using an α-Ldb1 antibody and Streptavidin-HRP (Figure 1B). Non-induced clone #3F4 expresses low levels of bio-Ldb1 and, upon induction, upregulates it to levels comparable to endogenous Ldb1 (Figure 1C). This allowed us to compare Ldb1 complexes before and after terminal differentiation.

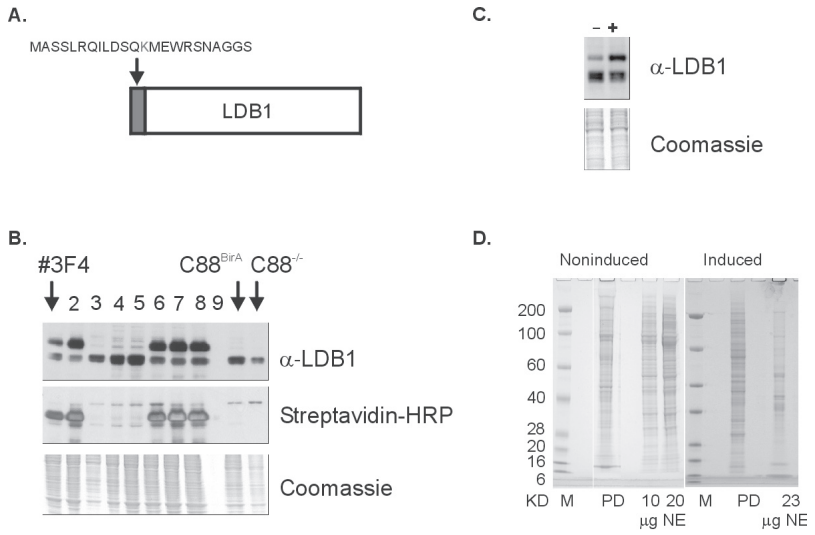


Figure 1 Ldb1 biotinylation and streptavidin pulldown.

A. Schematic representation of bio-Ldb1. The 23 amino acid sequence recognized by BirA and a triple Haemagglutinin tag are fused in tandem to the amino-terminal end of the ATG-less Ldb-. B. Expression of Ldb1 and bio-Ldb1 in nuclear extracts of induced C88$^{BirA/bio-Ldb1}$, C88$^{-/BirA}$ and C88$^{-/-}$ cells. Eight of 16 transfectants are shown. Three clones (lanes 3-5) did not express bio-Ldb1, lane 9 not loaded. Clone #3F4, (arrow) was chosen for further experiments. Lanes on the right are C88$^{-/BirA}$ and C88$^{-/-}$ controls. C. Ldb1 and bio-Ldb1 expression in equal amounts (see Coomassie stained gel) of nuclear extracts of non-induced (-) and induced (+) #3F4 cells. Expression of endogenous Ldb1 is reduced in induced cells. D. PAGE of proteins bound to bio-Ldb1; PD: pulled down proteins; NE: untreated nuclear extract. Pull downs of non-induced and induced cell extracts washed at lower stringency conditions are shown.

Identification of Ldb1 interaction partners

Nuclear extracts prepared from non-induced and induced #3F4 cells were incubated with streptavidin coated paramagnetic beads and separated by PAGE[30,31] (Figure 1D). Proteins were trypsin digested, eluted and analysed by mass spectrometry (LC-MSMS). Table 1 shows the pulled down proteins when washed at the same level of stringency (150mM salt). An

Table I

Proteins pulled down	C88-/BirA	Nonind. lo str.	Ind. lo str.	Ind. hi str.
Transcription factors				
LIM only proteins				
Lmo2	-	+	-	+
Lmo4	-	-	+	+
Zinc finger proteins				
Gata1	-	-	+	+
Basic helix-loop-helix				
Tal1	-	+	+	+
E2A	-	+	+	+
Lyl1	-	+	+	+
HEB	-	+	+	+
E2-2*	-	+	+	+
ETO-family				
Eto-2	-	+	+	+
Mtgr1	-	+	-	+
Runt domain				
Runx1	-	-	+	+
SSDP				
Ssdp2	-	+	+	+
Ssdp3	-	+	+	+
RIKENcDNA1210001E11(Ssdp4)	-	+	+	+
Cell cycle proteins/kinase				
Cdk9	-	+	-	-

Table 1: Proteins identified by LC-MSMS in bio-ldb1 pulldown experiments.
The lower amount of identified proteins and their corresponding peptides in the lower stringency pulldown experiment of induced cells (ind.1) is due to the lower amount of input for the LC-MSMS analysis. Proteins marked with an asterisk were not tested in immunoprecipiations.

additional pulldown with nuclear extract from induced #3F4 cells was performed under higher stringency washing conditions (200mM) with very similar results.

The screen for interacting proteins was validated by the fact that known Ldb1 partners (Lmo2, Tal1, Gata1 and E2A)[18] were readily identified. We also found 3 proteins of the Ssdp family, members of which interact with Ldb1 in HeLa cells and in Drososphila[33-35]. In addition, a substantial number of novel (potential) interaction partners of different functional classes was found, including transcription (co)factors, cell cycle proteins, chromatin remodelling and DNA repair proteins. In agreement with our previous results[30,31], a number of endogenously biotinylated proteins and background proteins were also identified (data not shown).

Some of the previously identified binding partners of Ldb1 (Tal1, E2A), the bHLH

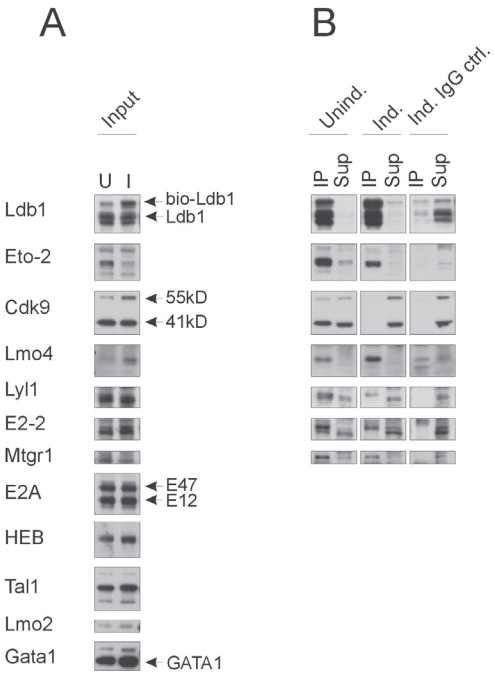


Figure 2 Analysis of Ldb1 interacting protein complexes.

A. Protein levels of Eto-2, Cdk9 and Lmo4 change with induction of MEL cell differentiation. Western Blot analysis of non-induced (U) and induced (I) MEL cell nuclear extracts. Input lanes indicate levels of tested proteins in untreated nuclear extracts diluted to the same concentration as in the IP experiments. Proteins immunoprecipitating with Ldb1 (IP lane) and supernatant (Sup) are shown and labeled accordingly. The isoforms of E2A and Cdk9 and bio-Ldb1 are indicated.

B. Changes in protein levels are reflected in the amounts in which they interact with Ldb1. Extracts were incubated with α-Ldb1 antibodies and equal amounts of immunoprecipitated (IP) and Supernatant (Sup) loaded. Less Eto-2 but more Lmo4 co-precipitate with Ldb1. Even though upregulated, Cdk9 (55kD form) does not interact with Ldb1 in induced cells. The IgG controls were performed with nuclear extracts of noninduced (not shown) and induced (shown here) bio-Ldb1 cells with a non-specific, isotype matched antibody.

proteins HEB, E2-2, Lyl1 and the novel interacting protein Eto-2 were detected in all three analyses. Eto-2, an ortholog of the *D. melanogaster* gene *nervy,* was the most abundant protein in two of the three LC-MSMS outputs. Mtgr1, another member of the ETO protein family, was also found identified by MS analysis of non-induced and induced cell extracts. Additionally, Ssdp2, Ssdp3 and a protein homologous to human Ssdp4 were identified in all MS analyses, but not analysed further. The known partner Lmo2, present in extracts from non-induced and induced cells (Figure 2), was only detected in one of the induced nuclear extracts. Lmo4, Gata1 and Runx1 were present in the MS analyses of induced cells only, although it should be noted that Gata1 may be absent due to the fact that only very few peptides are detectable in our MS analysis [31]. In contrast to the proteins mentioned above, the cell cycle protein Cdk9 was only found in nuclear extracts of non-induced cells. Proteins of different chromatin remodeling complexes were also identified only in induced cell extracts. With the exception of BRG/brm-associated factor 53A none of these were found under both stringency washing conditions.

To determine the expression levels of the binding partners found in the MS analysis, western blots of equal amounts of nuclear extracts of noninduced and induced MEL cells were carried out. This showed that the levels of all binding partners except for the newly identified interacting proteins Eto-2, Cdk9 and Lmo4, did not change significantly (Fig.2A, input lanes). Lmo4 and Cdk9 increased with induction whereas there was considerably less Eto-2 in induced extracts (Figure 2A, input Lmo4, Cdk9 and Eto-2 panels). Interestingly the much less abundant 55kD isoform of Cdk9[36] is upregulated with induction, whilst the 41kD isoform of Cdk9 is present at the same levels in non-induced and induced extracts (Figure 2A, input Cdk9 panels); a phenomenon also seen in differentiating macrophages[37].

Identification of different complexes

To confirm the interactions found in the MS analysis, immunoprecipitations of equal amounts of nuclear extracts from induced and non-induced C88 (not shown) and #3F4 cells (Figure 2) were performed with an α-Ldb1 antibody. The α-Ldb1 antibody depleted the extracts of non-induced cells almost completely of endogenous and bio-Ldb1, whereas a very low amount (<5%) remained in the supernatant of the induced cell extract (Figure 2B, Ldb1 panel). Lmo2, Tal1, the two E2A isoforms E12 and E47, HEB and Gata1, precipitated equally with Ldb1 before and after induction (not shown). Consistent with its lower level in induced cell extract, less Eto-2 precipitated with Ldb1 from induced cell extracts when compared to that of non-induced cells (Figure 2B, Eto-2 panels). Mtgr1, the other identified Eto-family member, was also enriched to a lesser extent in induced cells (Figure 2B, Mtgr1 panels). As expected, more Lmo4 precipitated with Ldb1 in induced cells when compared to non-induced cells (Figure 2B, Lmo4 panels), while the amount of co-precipitated Lmo2 did not change (not shown). These results suggest that the newly identified partner Lmo4 may play a more important role at later stages of erythroid differentiation. Most notably however, the two isoforms of Cdk9 immunoprecipiated with Ldb1 only in non-induced cells (Figure 2B, Cdk9 panels). Since Cdk9 is involved in cell cycle progression[38], the interaction between Ldb1 and Cdk9, which appears to be exclusive to noninduced, proliferating MEL cells, may link the formation of a complex containing Ldb1 and Cdk9 to the maintenance of the proliferative state (see also below). Although their level is unchanged , there was less co-precipitation of E2-2 and Lyl1 with either α-Ldb1 (Figure 2B) or α-Eto-2 (not shown) in induced cells.

In order to gain a better understanding of some of the Ldb1 containing complexes and to determine whether Ldb1 binding partners bind to Ldb1 in the absence of Eto-2 and vice versa we performed sequential IP experiments: We first depleted nuclear extracts of either Ldb1 or Eto-2 with their respective antibodies and then incubated the supernatants with α-Eto-2 and α-Ldb1 respectively (Fig. 3). IPs were then analysed for Ldb1 interacting proteins identified and validated in the single IP experiments (not shown and see Figure 3). It should be noted that we have not set out to characterise all the possible complexes that may be formed by every single Ldb1 interacting partner identified in our MS analysis, a task which is outside the scope of this paper.

As noted above, the α-Ldb1 antibody (almost) completely depletes Ldb1 protein from the extract (Figure 2B and 3 Ldb1 panel). A precipitation with α-Eto-2 antibody brings down high amounts of Ldb1, indicating that a high proportion of complexes contain both Ldb1 and Eto-2. However, α-Eto-2 did not deplete the extract of Ldb1 protein since more Ldb1 protein is precipitated with α-Ldb1 from the α-Eto-2 treated supernatant (Figure 3 - Ldb1 panel). Conversely, precipitation of Ldb1 with α-Ldb1 antibody brings down large amounts of Eto-2 (Figure 2B, 3 - Eto-2 panels) but does not deplete Eto-2 completely from the extract since there is more Eto-2 left in the supernatant that can be precipitated with α-Eto-2 antibody (Figure 3 – Eto-2 panels). Taken together, these results suggest that there are at least three complexes, one containing both Eto-2 and Ldb1, one that contains Ldb1 but not Eto-2 and one that contains Eto-2 but not Ldb1 (suppl. Figure 1).

To investigate the binding behaviour of Cdk9, the only protein that does not interact with Ldb1 upon induction of differentiation, we tested the sequential IPs for the presence of Cdk9 protein. α-Ldb1 antibody precipitated Cdk9 with a substantial amount left in the supernatant which was not precipitable by α-Eto-2 antibody (Figure 3 – Cdk9 panels). In reverse, incubation of nuclear extracts with α-Eto-2 antibody also precipitated Cdk9 only partially, the remainder

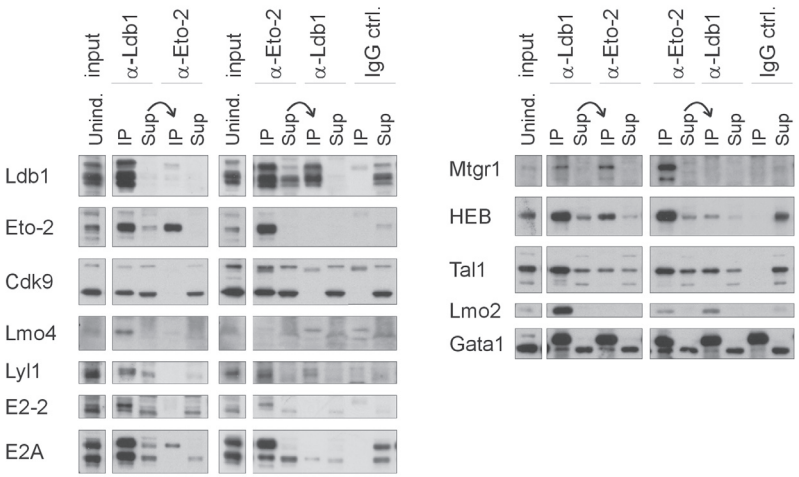


Figure 3. Sequential immunprecipitations to investigate Ldb1 and Eto-2 containing complexes.

Nuclear extracts of noninduced bio-Ldb1 cells were first depleted of either Ldb1 or Eto-2 with their respective antibodies. The supernatants were then incubated with α-Eto-2 or α-Ldb1, respectively. The second supernatant was also loaded to determine which proteins do not interact with either Eto-2 or Ldb1. IgG lanes are control IPs carried out with a non-specific, isotype matched antibody.

also not being bound to Ldb1 (Figure 3 – Cdk9 panels). We conclude that Cdk9 requires both Ldb1 and Eto-2 for its interaction with either of the two proteins. E2-2 and Lyl1 exhibited a similar interaction behaviour (Figure3). with the exception that Lyl1 is also bound to Ldb1 alone (suppl. Figure 1).

To test whether the Ldb1-Lmo2 interaction requires Eto-2, we first depleted Eto-2 from the extract and then carried out a Ldb1 IP testing for Lmo2 (Figure 3– Lmo2 panels): α-Eto-2 brought down only some Lmo2. Since all of Lmo2 is in complex with Ldb1 (see above) we conclude that there is a complex containing Lmo2-Ldb1-Eto-2 and that the remaining Lmo2 left in the supernatant after Eto-2 depletion is bound to Ldb1 but not Eto-2. Lmo4 basically behaves similarly but shows an important quantitative difference. There is little Lmo4 before differentiation which increases several fold after induction of differentiation (figure 2B).

The two isoforms of E2A show a different binding behaviour with regards to Ldb1 and Eto-2: Both the larger E47 and smaller E12 isoforms are precipitated equally by Ldb1, albeit incompletely. The remaining E47, but not E12 is bound by Eto-2 (Figure 3 – E2A panels). Conversely, α-Eto-2 completely depletes the extracts of the E47 isoform but binds only some E12 (Figure 3 – E2A panels). Some of this remaining E12 is precipitable by Ldb1. Taking into account the Ldb-1/E2A data from above this indicates the presence of a complex containing Ldb1/Eto-2/E47/E12, two further complexes composed of at least Eto-2/E47 or Ldb1/E12 and “free” E12 that can participate in other complex formation.

Mtgr1 was the second member of the Eto family identified in our MS analysis. α-Ldb1 antibody precipitates some Mtgr1, but not all, with a substantial amount left in the supernatant that is all precipitated with an α-Eto-2 antibody (Figure 3 – Mtgr1 panels). Precipitation with

α-Eto-2 antibody however completely depletes Mtgr1 from the nuclear extracts indicating that all Mtgr1 is complexed with Eto-2 (Figure 3 – Mtgr1 panels). We conclude that there are at least two complexes containing Mtgr1: Firstly, one complex containing Mtgr1/Eto-2/Ldb1 and that the Mtgr1/Ldb1 interaction requires Eto-2. Secondly, a further complex containing Mtgr1/Eto-2 but not Ldb1.

Tal1 also forms several complexes, with and without either Ldb1 or Eto-2. α-Ldb1 antibody precipitates most but not all Tal1 from the nuclear extract, part of the remaining Tal1 is pulled down with Eto-2 (Figure 3 – Tal1 panels). Testing first with α-Eto-2 shows that Eto-2 brings down Tal1 but not all, some of which is precipitable with an α-Ldb1 antibody (Figure 3 – Tal1 panels). This indicates that there are at least four Tal1 containing complexes: Eto-2/Tal1/Ldb1, Ldb1/Tal1 Eto-2/Tal1 and "free" Tal1.

HEB also forms complexes that contain either Ldb1 and/or Eto-2. Both α-Ldb1 and α-Eto-2 precipitate large amounts of HEB indicating that HEB forms complexes with Ldb1 or Eto-2, and probably a complex containing all three of these proteins. A second precipitation with either α-Eto-2 or α-Ldb1 respectively precipitates more HEB, but again not completely, with small amounts remaining in the supernatant after this second immunoprecipitation step (Figure 2C – A and B HEB panels). Therefore, HEB forms at least three complexes: HEB/Ldb1, HEB/Eto-2 and "free" HEB and possibly HEB/Ldb1/Eto-2. When an α-HEB IP is carried out it does precipitate all the partners (suppl. Table 1), including Lyl1, Cdk9 and E2-2, suggesting that it is part of a large Ldb1/Eto-2 complex (suppl. Figure 1) or perhaps a smaller one containing Ldb1 and Eto-2.

Finally, Gata1, a transcription factor which we have previously shown to participate in a multitude of complexes[39], also forms separate complexes with either Eto-2 or Ldb1. Treatment of extracts with α-Ldb1 antibody brings down small amounts of Gata1 with the majority of Gata1 left in the supernatant. Further precipitation with α-Eto-2 provides evidence of an Eto-2-Gata1 interaction separate from Ldb1 (Figure 3 – Gata1 panels). In reverse, and in accordance with the previous IP, α-Eto-2 antibody precipitates Gata1 but incompletely. Very little Gata1 is then precipitable with an α-Ldb1 antibody indicating that possibly most of the Eto-2/Gata1 complex also contains Ldb1 (Figure 3 - Gata1 panels). We conclude that Gata1 forms at least four complexes: Gata1/Ldb1, Gata1/Eto-2, Gata1/Ldb1/Eto-2 and "free" Gata1, which we know is participating in other complexes.

Single IP's using antibodies for the Ldb1 interacting proteins confirmed all the pairwise interactions described above (not shown).

Although it is difficult to distinguish the complexes from each other, the data strongly suggest that there are several subcomplexes formed by Ldb1 and its interaction partners that can form larger, functional complexes (possibly via the homodimersation of Ldb1). Taking these data together we can distinguish two large subcomplexes with either Ldb1 or Eto-2 and one large complex containing both Ldb1 and Eto-2 (suppl. Figure 1). Upon differentiation association with Cdk9 is lost and the level of Eto-2 is substantially decreased, while the amount of Lmo4 is increased several fold. The reduction in association is also reflected in the reduced co-imunoprecipiation of E2-2 and Lyl1 with α-Ldb1 and α-Eto-2 in induced cells. In addition these complexes appear to interact with Runx1, but we have as yet not been able to characterize this interaction due to the poor quality of antibody.

Chromatin immunoprecipitations

To confirm Eto-2 and Ldb1 are bound to chromatin at specific regulatory sites we carried out chromatin immunoprecipitations using a number of erythroid genes: *Gata1*, *GPA*, *Myb*, *Myc* and *Eklf* for which the binding of Gata1 complexes is known (Figure 4)[31,40]. Among the *Gata1* gene regulatory regions the HS-3.5 is known to bind the Gata1/Tal1/Ldb1 complex, whereas the DNaseI hypersensitive site in the gene (Gata1 IE) does not[41]. The GATA-E box sites in the *GPA* and *Eklf* promoters are also known targets of Gata1/Ldb1[20,31]. The *c-myb and c-myc* genes were tested because they bind a Gata1/Gfi-1b complex[31]. In all cases we find an enrichment for *Gata1* (as expected) when compared to non GATA site negative control fragments (not shown; [31]). In contrast, Eto-2 and Ldb1 were bound to *Gata1*, *Eklf* and *GPA* (Figure 3), but not to *c-myc and c-myb* (not shown). The ratio of Eto-2/Ldb1 binding to the three elements decreases during differentiation in MEL cells (Figure 3) in accordance with the fact that there is less Ldb1/Eto-2 complex (see Figure 2). Gata1, GPA and Eklf proteins are expressed late in erythroid differentiation (the -3.5 HS of the *Gata1* gene also regulates *Gata1* expression in megakaryocytic cells[42,43] whereas *c-myc* and c-*myb* are downregulated. The decrease of Eto-2 levels in late erythroid cells suggests that Eto-2 adds a repressive function to the Gata1/Tal1/Ldb1 complex to repress late erythroid genes early during differentiation. In contrast Eto-2 would not be essential for Gfi-1b repressed genes such as c-*myb* and c-*myc*[31].

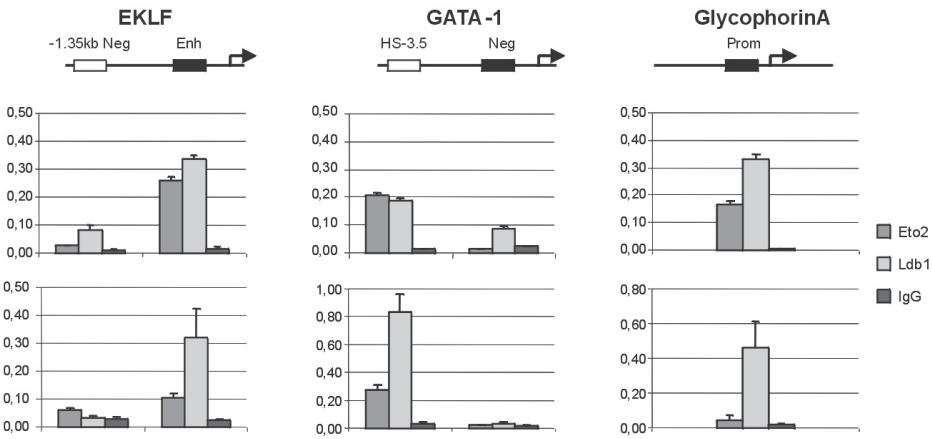


Figure 4: ChIP of ldb-1 and Eto-2. Top: Localisation of the upstream Hypersensitive sites in the Eklf, Gata1 and glycophorin promoter.
Middle row: Bar graphs of the relative enrichment of sequences immunoprecipitated by Eto-2 (grey), Ldb1 (light grey) and the IgG control (dark grey) in non-induced MEL cells. All values were normalised to a GAPDH control. Bottom row: Same as middle row for induced MEL cells.

Eto-2, Cdk9 and Lmo4 are required for definitive hematopoiesis in a zebrafish model system.

Ldb1 and its constitutive binding partners Lmo2 and Tal1 are essential for embryonic blood formation[29,44,45]. We next asked whether the newly identified Ldb1 interacting partners E2A, Cdk9, Eto-2, Lmo4 and Mtgr1 are required for hematopoietic development. Since the genetic regulation of embryonic and definitive hematopoiesis is highly conserved between zebrafish

and mammals, we tested the role of these proteins by ATG-Morpholino (ATG-MO) mediated inhibition of mRNA translation in zebrafish embryos[46]. In zebrafish, primitive erythrocytes expressing embryonic hemoglobin derive from the intraembryonic intermediate cell mass (ICM) and start circulating at 24 hours post fertilisation (hpf). Shortly thereafter, with development of the definitive hematopoietic system, ventrally located flk1 positive precursors of the dorsal aorta start to express runx1. Runx1 is also expressed in primitive erythrocytes, the olfactory epithelium, Rohon-Beard neurons[47] and the anterior paraxial mesoderm.
ATG-MOs targeted against the zebrafish orthologues of Cdk9, Eto-2, Lmo4, Ldb1, and E12 were injected at three increasing doses into 2-8 cell stage embryos. To test the effects of these morpholinos on the embryonic and definitive hematopoeitic system we analysed embryos after onset of blood circulation by in situ hybridisation with probes against embryonic β-globin and runx1, respectively (Figure 5).

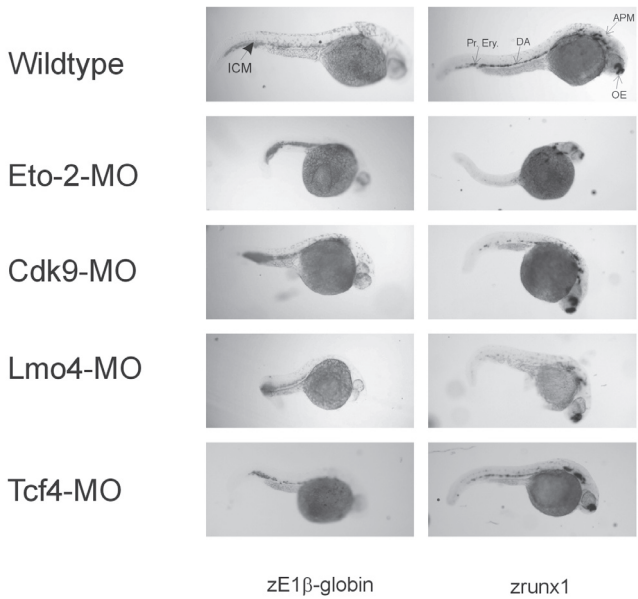


Figure 5 Analysis of zebrafish embryos after morpholino injections.
All pictures were taken at the same magnification. Top panels: non injected WT control embryos (30hpf) stained for embryonic β-globin (left column) and runx1 (right column). The intermediate cell mass (ICM), dorsal aorta (DA), primitive erythrocytes (Pr.Ery.), anterior paraxial mesoderm (APM.) and olfactory epithelium (OE.) are indicated. The globin signal is red, runx1 signal is blue. The second row of panels are representative embryos injected with 1pmol eto-2-MO (32hpf), the third row with 1pmol cdk9-MO (28hpf), the fourth row with 1 pmol lmo4-MO (30hpf) and the fifth row with 0.5pmol of the control tcf4-MO (30hpf).

Ldb1-MO injected embryos displayed variable phenotypes at all three doses, including deformation of the body axis, dysmorphic somites, abnormal tail morphology, hematopoietic defects resembling that of the mouse (A. Hostert, unpublished) and necrosis in the brain (not shown). The severity of defects increased with increasing dose of injected Ldb1-MO. As a negative control, we used a morpholino targeted against the zebrafish orthologue of *Tcf7/2*.

Tcf7/2 is not expressed in hematopoietic tissues and when deleted has a specific effect outside the mouse hematopoietic system, i.e. only in the mouse in the intestinal epithelium (Korinek et al., 1998). As expected, no effects of the z-tcf4-MO on the hematopoietic system were observed in the injected zebrafish at any dose, but embryos injected with the highest dose displayed some slight tail abnormalities.

The cdk9-MO had no effect on embryonic β-globin expression and expression of runx1 in the primitive erythrocytes located in the posterior ICM, the olfactory epithelium and the anterior paraxial mesoderm. However, it had a severe effect on definitive erythopoiesis. 0.5 pmol cdk9-MO showed a clear reduction of runx1 signal in the dorsal aorta precursor population (n=10/18) which decreased further in embryos injected with 1 pmol (n=8/10). Surprisingly the effect appears to be specific to the hematopoietic system although cdk9 is expressed in many tissues[48].

Embryos injected with the eto-2-MO had a similar but distinct phenotype. The reduction of runx1 expression in embryos injected with 0.5 pmol (n=13/16) and 1 pmol of eto-2 MO (n=5/7) was more severe when compared to the cdk9-MO: runx1 in the dorsal aorta was either almost or completely abolished. In addition, and different from the cdk-9-MO embryos, runx1 expression in primitive erythrocytes was severely reduced or completely absent. Embryonic β-globin expression was normal in all eto-2-MO injected embryos, however, some primitive erythrocytes in the caudal region were located laterally to the midline as opposed to their location in wild type embryos. This is similar to the effect observed when sonic hedgehog signaling is inhibited, again suggesting that eto-2 may play a role in the response to extracellullar signals. Embryos treated with the eto-2 MO were also reduced in size. Injection of the highest dose of the mtgr1-MO did not affect embryonic hematopoiesis. Some reduction of runx1 expression in dorsal aorta precursor cells was observed (not shown).

Embryos treated with the lmo4-MO were comparable to cdk9 and eto-2. Expression of embryonic γ-globin was normal, whereas reduced levels of runx1 were observed in the dorsal aorta at 0.5 pmol MO (n=8/13) and decreased further with lmo4-MO injected at 1 pmol. In addition, some of the treated embryos appeared to have brain or neural tube abnormalities, which are the cause of perinatal death of the correponding knock out mouse[49-51]. These mice showed no defects in the hematopoietic system, however, only half of the homozygous *Lmo4* null mutants were born. The other half died around E9 of gestation, possibly due to a hematopoietic phenotype.

We conclude that the newly identified Ldb1 interaction partners eto-2, cdk9 and lmo4 are essential for definitive erythropoiesis, whereas Mtgr1 plays a less critical role It is noteworthy in this context that Eto-2 (in zebrafish, this paper) and most of its constitutive binding partners, namely HEB and E47 (in mouse[52]) and Mtgr1 (in zebrafish, this paper), are not required for embryonic hematopoiesis.

Ldb1 interacting partners are expressed in the same cells in the para-aortic splanchnopleura of the early mouse embryo.

The results obtained for the novel Ldb1 interaction partners in the zebrafish suggest that they would be expressed in the early mouse embryo at stages prior to the “birth” of the definitive hematopoietic stem cells in the AGM[1]. We therefore performed immunohistochemistry and immunofluoresence on E9.5 embryo sections (Figure 6). Immunohistochemical analysis with α-Ldb1, α-E2A, α-Lmo2, α-Gata1, α-Eto-2 and α-Cdk9 showed that all are expressed in the para-aortic splanchnopleura (P-Sp), the region destined to

contribute to the AGM (Figure 6A, B).

To further determine the expression pattern of the interacting proteins and to confirm that cells within the P-Sp co-express these proteins, we performed in situ immunofluorescence experiments on the cryosections. We found that cells positive for Gata1 also expressed Eto-2 and Ldb-1 (Figure 6C). Cells expressing Ldb-1 were also positive for Runx1 (Figure 6C), E2A and Lmo2 (not shown). The observation that the newly identified interaction partners of Ldb1 are co-expressed with Ldb1, Gata1, E2A and Lmo2 in the pre-AGM cells of the P-Sp strongly suggests that they form a complex that may be essential for the induction of the definitive hematopoietic system. The fact that Eto-2 is predominantly in the cytoplasmic at that stage (Figure 6C, Column 3 Panels Eto-2 and Merge) suggests that it may respond to extra-cellullar signals at the pre-hematopoietic stage. At later stages (fetal liver) it is abundant in the nucleus but its cytoplasmic expression is very weak. (not shown).

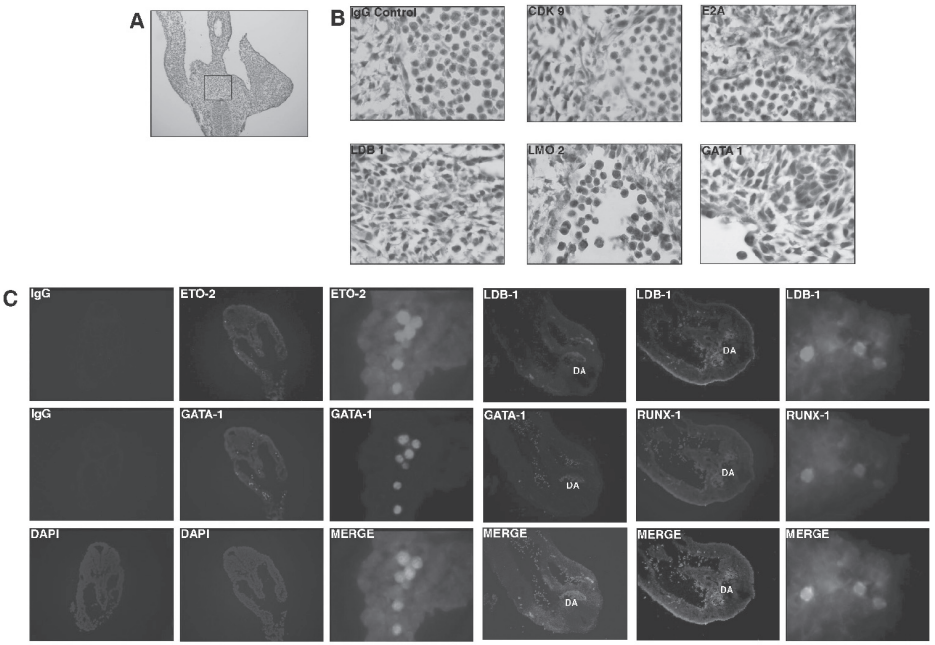


Figure 6 Ldb1 interacting partners are expressed in the P-Sp.
A. Hematoxylin stained 10μm transversal cryosection of a D9.5 FvB embryo. The area of the P-Sp is indicated by the box. B. High magnification (1000 X) of the areas such as shown in (A). Expression of Cdk9, E2A, Ldb-1, Lmo2 and Gata1 was detected with specific antibodies and visualised with DAB. IgG panel shows the background staining with an unspecific IgG antibody. C. Immunfluorescence analysis of E9.5 embryos with α-Eto-2, α-Ldb1, α-Runx1 and α-Gata1 antibodies. Specific staining is seen for both Eto-2 (Green) and Gata1 (Red) antibodies, compared to the IgG control. Third column of panels show an enlargement of part of the section shown in the second column. Merge is a superposition of images of the Eto-2 and Gata1 detection. The third column shows cells that express both Eto-2 and Gata1 appear as yellow. The fourth column shows that Ldb-1 (Green) and Gata1 (red) expressing cells are located in the dorsal aorta (DA) region; cells positive for Ldb1 are also positive for Gata1. Ldb1 (Green) and Runx-1 (Red) expression overlap in E9.5 embryos.

Discussion

Ldb1 forms dynamic complexes during erythroid differentiation

By using a systems biology approach we identified all known and new binding partners of Ldb1 and determined their importance for hematopoietic development. The important implication of this work is that the analysis of proteins co-expressed and interacting with each other in a late mature cell type are already implicated at the earliest stages of (in this case blood) development.

The Ldb1 proteome has a number of interesting interactions: 1. with Eto-2 (and Mtgr1). 2. with a large number of different bHLH proteins. 3. with Cdk9 and 4. with the Ssdp proteins although this latter interaction was not analysed further. Recently we have shown that Gata1 forms at least five clearly identifiable protein complexes[31]. We did not observe such distinct complexes for Ldb1. Clearly Ldb1 forms a core complex with the known partners Gata1, Tal1, Lmo2 and E2A and the newly identified partner Eto-2 (Figure 6). Eto-2 also forms complexes with E2A and Tal1 without Ldb1. In the sequential IPs we can distinguish between the preferential binding partners of Eto-2 and Ldb1 and deduce the existence of a higher order complex whose formation is favored in proliferating cells. At the same time it is difficult to separate groups of interacting proteins from each other by IPs, especially if a protein interacts in different combinations. Hence we are in the process of purifying the different complexes. With induction, levels of Eto-2 decrease and formation of the large complex is lower (Figure 7). Cdk9, which has been linked to cell cycle progression[38], would no longer be part of the complex. This may explain why the cells stop proliferating upon differentiation. At the same time the levels of Lmo4 increase, possibly replacing Lmo2/Eto-2 and leading to an activation of transcription of genes expressed after terminal differentiation.

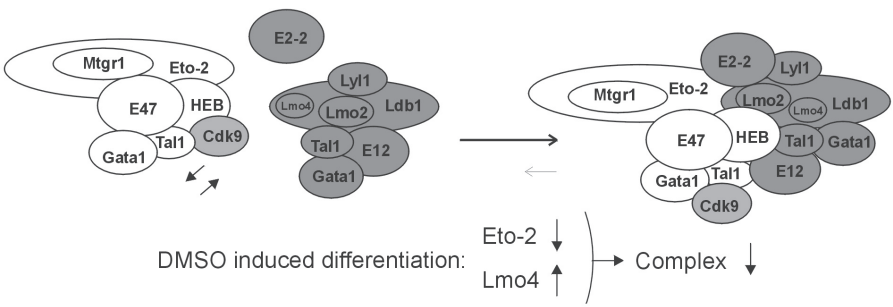


Figure 7 Model of Ldb1 complexes. Based on the interaction of Ldb1 and Eto-2 and their respective binding partners.
The horizontal arrows indicate that the balance of interaction is towards the large complex in proliferating non-induced cells. Upon the induction of differentiation and termination of proliferation the level of Eto-2 drops while the level of Lmo4 rises, hence the equilibrium would shift towards the smaller complexes. The presence of several DNA binding proteins in a single complex may explain the role of Ldb1 as a facilitator of long range interactions.

Noteworthy is the presence of a large number of DNA binding proteins within the complex, particularly the presence of at least five bHLH proteins and the zinc finger transcription factor Gata1. It is possible that such a complex may very well be involved in the establishment and/or facilitation of long range interactions, processes in which Ldb1 has been implicated [57].

Specifically, in *D. melanogaster* Chip was identified in a screen for factors involved in the long range gene activation of the *cut* gene. Chip was proposed to bridge the Pannier (GATA) and Achaete/Scute (bHLH) complexes causing the intervening DNA to loop out bringing DNA control elements into close proximity . The erythroid Ldb-1-Lmo2-Tal1-E2A-Gata1 complex may have a similar role. It was recently shown that the complex binds to multiple sites in the α-globin locus[21]. We envisage that these complexes interact and promote long range interactions also in other gene loci, eg. in the β-globin locus ACH[10].

The Eto family members Eto-2 and Mtgr1 are thought to be repressors by binding the NCor/Sin3A/HDAC1 complex . Eto-2 was the most abundant protein in the MS analysis, suggesting it is a direct (and crucial) binding partner of Ldb1 (confirmed by immunoprecipitations). We also observed that Eto-2 and its family member Mtgr1 interact with each other. Moreover Eto-2 appears to be the bridging factor for Ldb1 to interact with Mtgr1 (Fig. 2). Eto proteins and the *D. melanogaster* orthologue *nervy* have four highly conserved protein interaction domains . The DNA binding of the Eto-2 complexes we describe probably occurs through its bHLH binding partners and/or Gata1.

Interestingly, with induction Eto-2 and the Eto-2/Ldb1 complex decrease, while the level of Lmo4 increases. Thus the level of the large Ldb1-Eto-2 complex drops to be replaced by one with Lmo4. Complementary results were obtained by Goardon et al. (personal communication) through characterization of Tal1 complexes. As indicated by the ChIP experiments the repressive role of Eto-2 may very well explain how late erythroid genes become activated. An alternative or complementary explanation may be post-translational protein modifications of the Eto-2 complex members, a possibilty that can not be excluded at present.

A good candidate to carry out modifications and change specific interactions is Cdk9. This CDC2 ortholog precipitated differentially with Ldb1 in proliferating and differentiating MEL cells. The invariance of the main isoform independent of the cell cycle stage has been reported. Cdk9 is thought to have two functions: Regulating RNA polymerase II by phosphorylating its C-terminal domain[66-71]. Secondly, regulating the cell cycle. A “knockdown” of the *D. melanogaster* orthologue of Cdk9 causes an arrest at the G1 to S transition , congruent with the in vitro data that the human Cdk9 orthologue phosphorylates Rb . Interestingly, Gata1 has been found to interact with Rb *in vitro* . We detect an interaction between Cdk9 and Ldb1 only in non-induced, proliferating MEL cells, suggesting Ldb1 and its partners might be involved in transcriptional control of the cell cycle through the dual function of Cdk9. Putative target genes could be E2F family members expressed in hematopoietic cells. Cdk9 also interacts with BRG1 and STAT3 to activate transcription of the cell cycle inhibitor p21waf1 suggesting Cdk9 has different functions in the cell cycle depending on its partners. In this context it is interesting that we detect Eto-2 protein in the cytoplasm of hematopoietic stem cell precursors (Figure 4) and of developing neurons (N. Meier, unpublished). Eto-2 may respond to extracellular signals by translocating from the cytoplasm into the nucleus, whilst it is downregulated in terminally differentiating cells. This suggests that Eto-2 plays a key role in the birth of the definitive HSCs and is required for the subsequent phases of expansion of the different lineages. It then needs to be downregulated to allow the activation of late genes such as *Eklf* for terminal differentiation.

Late erythroid differentiation complexes and early hematopoiesis

It is remarkable that the analysis of interacting proteomes of Gata1 and Ldb1 (this paper) in late erythroid cells has resulted in the identification of a number of proteins essential

for early hematopoiesis (as found in the zebrafish experiments), in particular Lmo4, Cdk9 and Eto-2.

Lmo4 expression in mice has been detected at E9 from the caudal region of the dorsal and lateral paraxial mesoderm up to the direct vicinity of the dorsal aorta, suggesting that Lmo4 could play a role in HSC formation . It is also upregulated in late T cell differentiation similar to what is observed in MEL cells. In Xenopus Xlmo4 and Gata-2 act synergistically in ventral mesoderm formation. However an Xlmo4-MO did not prevent ventral mesoderm formation . In zebrafish Lmo4 is expressed at gastrulation but not during ventral mesoderm formation . In accordance with this, we found embryonic hematopoiesis in lmo4-MO treated zebrafish embryos to be normal. We show the expression of lmo4 at later stages to be important for the formation of definitive HSCs in zebrafish. Whether the same phenotype is observed in mice remains to be determined.

MO mediated repression of eto-2 had the most drastic phenotype in zebrafish – absence of runx1 expression in embryonic erythrocytes and lack of definitive hematopoiesis, suggesting it is essential for runx-1 expression. Cdk9-MO treated fish lacked definitive cells. However in embryonic cells runx1 expression was detected, suggesting it does not regulate runx-1 directly.

The fact that these factors are already co-expressed in the same cells in the murine P-Sp before the generation of the definitive HSCs in the AGM suggests that they are also essential for definitive hematopoiesis in the mouse. Eto-2 and Cdk9 may be more important in this process than Gata1, because loss of Gata1 is not deleterious to these very early stages (with cytoplasmic Eto-2). This in turn suggests that Tal1, Ldb1 and Lmo2 are dominant at this stage. Because all of the Ldb1 bound Gata1 is complexed with Eto-2, we postulate that when Eto-2 translocates from the cytoplasm to the nucleus, the balance of interactions is changed causing the transition to a more proliferative Gata1 driven phase followed by terminal differentiation.

Acknowledgements

We are grateful to Drs. S. Orkin, R. Baer and G. Gill for the α-Lmo2, α-Tal1 and α–Ldb1 antibodies. N.M. was supported by a DFG Fellowship, A.H. was supported by a Marie Curie Fellowship. This work is supported by the NWO (NL), the NIH and the EU (CiO).

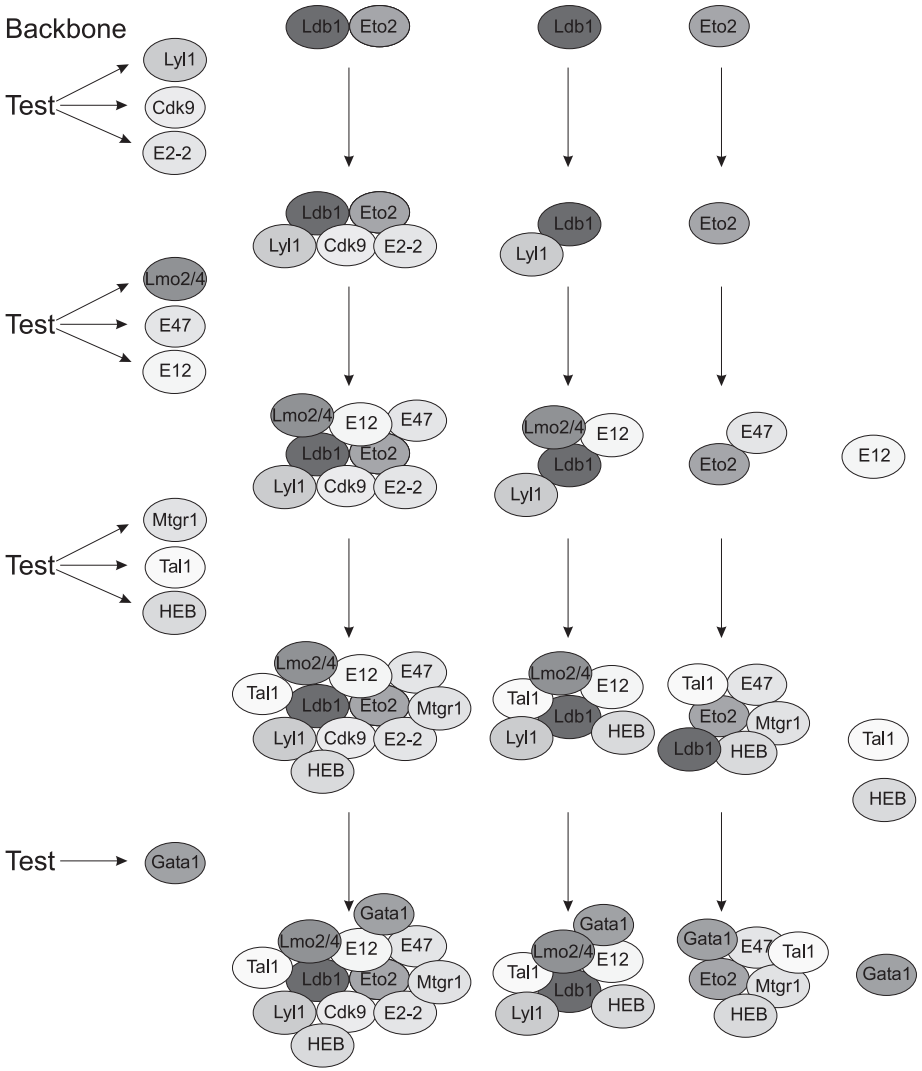
References

1. Durand C & Dzierzak, E. Embryonic beginnings of adult hematopoietic stem cells. *Haematologica* **90**, 100-108 (2005).
2. de Bruijn MF, Speck NA, Peeters MC & Dzierzak, E. Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo. *EMBO J* **19**, 2465-2474 (2000).
3. Yoder MC et al. Characterization of definitive lymphohematopoietic stem cells in the day 9 murine yolk sac. *Immunity* **7**, 335-344 (1997).
4. Cumano A, Ferraz JC, Klaine M, Di Santo JP & Godin, I. Intraembryonic, but not yolk sac hematopoietic precursors, isolated before circulation, provide long-term multilineage reconstitution. *Immunity* **15**, 477-485 (2001).
5. Medvinsky A & Dzierzak, E. Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell* **86**, 897-906 (1996).
6. de Bruijn MF, Speck NA, Peeters MC & E, D. Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo. *EMBO J* **19**, 2465-2474 (2000).
7. Ottersbach K & Dzierzak, E. The murine placenta contains hematopoietic stem cells within the vascular labyrinth region. *Dev Cell* **8**, 377-387 (2005).
8. Gekas C, Dieterlen-Lievre F, Orkin SH & Mikkola, H. The placenta is a niche for hematopoietic stem cells. *Dev Cell* **8**, 365 - 375 (2005).
9. Cantor AB & Orkin, S. Hematopoietic development: a balancing act. *Curr Opin Genet Dev* **11**, 513-519 (2001).
10. Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. & de Laat, W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell Biol* **10**, 1453-1465 (2002).
11. Palstra RJ et al. The beta-globin nuclear compartment in development and erythroid differentiation. *Nat Genet* **35**, 190-194 (2003).
12. Patrinos GP et al. Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. *Genes Dev* **18**, 1495-1509 (2004).
13. Tolhuis B, Palstra RJ, Splinter E, Grosveld F & de Laat, W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell Biol* **10**, 1453-1465 (2002).
14. Morcillo P, Rosen C, Baylies MK & Dorsett, D. Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in Drosophila. *Genes Dev* **11**, 2729-2740 (1997).
15. Torigoi E et al. Chip interacts with diverse homeodomain proteins and potentiates bicoid activity in vivo. *Proc Natl Acad Sci U S A* **97**, 2686-2691 (2000).
16. Agulnick AD, T. M., Breen JJ, Tanaka T, Dawid IB, Westphal H. Interactions of the LIM-domain-binding factor Ldb1 with LIM homeodomain proteins. *Nature* **384**, 270-272 (1996).
17. Breen JJ, Agulnick AD, Westphal H & Dawid, I. Interactions between LIM domains and the LIM domain-binding protein Ldb1. *J Biol Chem* **273**, 4712-4717 (1998).
18. Wadman IA et al. The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA1 and Ldb1/NLI proteins. *EMBO J* **16**, 3145-3157 (1997).
19. Brand M et al. Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics. *Nat Struct Mol Biol* **11**, 73-80 (2004).
20. Lahlil R, Lecuyer E, Herblot S & Hoang, T. SCL assembles a multifactorial complex that determines glycophorin A expression. *Mol Cell Biol* **24**, 1439-1452 (2004).
21. Anguita E et al. Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA1 and GATA-2. *EMBO J* **23**, 2841-2852 (2004).
22. Fernandez-Funez P, Lu CH, Rincon-Limas DE, Garcia-Bellido A & Botas, J. The relative expression amounts of apterous and its co-factor dLdb/Chip are critical for dorso-ventral compartmentalization in the Drosophila wing. *EMBO J* **17**, 6846-6853 (1998).
23. Rincon-Limas DE, Lu CH, Canal I & Botas, J. The level of DLDB/CHIP controls the activity of the LIM homeodomain protein apterous: evidence for a functional tetramer complex in vivo. *EMBO J* **19**, 2602-2614 (2000).

24. Milan M, Diaz-Benjumea FJ & Cohen, S. Beadex encodes an LMO protein that regulates Apterous LIM-homeodomain activity in Drosophila wing development: a model for LMO oncogene function. *Genes Dev* **12**, 2912-2920 (1998).
25. Mukhopadhyay M et al. Functional ablation of the mouse Ldb1 gene results in severe patterning defects during gastrulation. *Development* **130**, 495-505 (2003).
26. Warren AJ et al. The oncogenic cysteine-rich LIM domain protein rbtn2 is essential for erythroid development. *Cell* **78**, 45-57 (1994).
27. Robb L et al. Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene. *Proc Natl Acad Sci U S A* **92**, 7075-7079 (1995).
28. Porcher C et al. The T cell leukemia oncoprotein SCL/Tal1 is essential for development of all hematopoietic lineages. *Cell* **86**, 47-57 (1996).
29. Shivdasani, R. A., Mayer, E.L., Orkin, S.H. Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. *Nature* **373**, 432-434 (1995).
30. de Boer E et al. Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc Natl Acad Sci USA* **100**, 7480-7485 (2003).
31. Rodriguez P et al. GATA1 forms distinct activating and repressive complexes in erythroid cells. *EMBO J* **24**, 2354-2366 (2005).
32. de Bruijn, M. F. et al. Hematopoietic stem cells localize to the endothelial cell layer in the midgestation mouse aorta. *Immunity* **16**, 673-683 (2002).
33. Chen L et al. Ssdp proteins interact with the LIM-domain-binding protein Ldb1 to regulate development. *Proc Natl Acad Sci U S A* **99**, 14320-14325 (2002).
34. van Meyel DJ, Thomas JB & Agulnick, A. Ssdp proteins bind to LIM-interacting co-factors and regulate the activity of LIM-homeodomain protein complexes in vivo. *Development* **130**, 1915-1925 (2003).
35. Nishioka N et al. Ssdp1 regulates head morphogenesis of mouse embryos by activating the Lim1-Ldb1 complex. *Development* **132**, 2535-2546 (2005).
36. Shore SM, Byers SA, Maury W & Price, D. Identification of a novel isoform of Cdk9. *Gene* **307**, 175-182 (2003).
37. Liu H & Herrmann, C. Differential localization and expression of the Cdk9 42k and 55k isoforms. *J Cell Physiol* **203**, 251-260 (2005).
38. Bettencourt-Dias M et al. Genome-wide survey of protein kinases required for cell cycle progression. *Nature* **432**, 980-987 (2004).
39. Rodriguez, P. et al. GATA1 forms distinct activating and repressive complexes in erythroid cells. *EMBO J* **24**, 2354-2366 (2005).
40. Rodriguez P, B. E., Krijgsveld J, Kolodziej KE, Guyot B, Heck AJ, Vyas P, de Boer E, Grosveld F, Strouboulis J. GATA1 forms distinct activating and repressive complexes in erythroid cells. *EMBO J* **24**, 2354-2366 (2005).
41. Guyot B, Valverde-Garduno V, Porcher C & Vyas, P. Deletion of the major GATA1 enhancer HS 1 does not affect eosinophil GATA1 expression and eosinophil differentiation. *Blood* **104**, 89-91 (2004).
42. McDevitt MA, Fujiwara Y, Shivdasani RA & Orkin, S. An upstream, DNase I hypersensitive region of the hematopoietic-expressed transcription factor GATA1 gene confers developmental specificity in transgenic mice. *Proc Natl Acad Sci U S A* **94**, 7976-7981 (1997).
43. Onodera K et al. GATA1 transcription is controlled by distinct regulatory mechanisms during primitive and definitive erythropoiesis. *Proc Natl Acad Sci U S A* **94**, 4487-4492 (1997).
44. Warren, A. J. et al. The oncogenic cysteine-rich LIM domain protein rbtn2 is essential for erythroid development. *Cell* **78**, 45-57 (1994).
45. Robb, L. et al. Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene. *Proc Natl Acad Sci U S A* **92**, 7075-7079 (1995).
46. Nasevicius A & Ekker, S. Effective targeted gene knockdown in zebrafish. *Nat Genet* **26**, 216-220 (2000).
47. Kalev-Zylinska ML et al. Runx1 is required for zebrafish blood and vessel development and expression of a human RUNX1-CBF2T1 transgene advances a model for studies of leukemogenesis. *Development* **129**, 2015-2030 (2002).

48. Bagella L et al. Cloning of murine CDK9/PITALRE and its tissue-specific expression in development. *J Cell Physiol* **177**, 206-213 (1998).
49. Hahm K, S. E., Fujiwara Y, Lindeman GJ, Visvader JE, Orkin SH. Defective neural tube closure and anteroposterior patterning in mice lacking the LIM protein LMO4 or its interacting partner Deaf-1. *Mol Cell Biol* **24**, 2074-2082 (2004).
50. Tse E et al. Null mutation of the Lmo4 gene or a combined null mutation of the Lmo1/Lmo3 genes causes perinatal lethality, and Lmo4 controls neural tube development in mice. *Mol Cell Biol* **24**, 2063-2073 (2004).
51. Lee SK et al. The LIM domain-only protein LMO4 is required for neural tube closure. *Mol Cell Neurosci* **28**, 205-214 (2005).
52. Zhuang, Y., Cheng, P., Weintraub, H. B-lymphocyte development is regulated by the combined dosage of three basic helix-loop-helix genes, E2A, E2-2, and HEB. *Mol Cell Biol* **16**, 2898-2905 (1996).
53. Berman J, Hsu K & Look, A. Zebrafish as a model organism for blood diseases. *Br J Haematol* **123**, 568-576 (2003).
54. Nasevicius A, E., SC. Effective targeted gene knockdown in zebrafish. *Nat Genet* **26**, 216-220 (2000).
55. Gering M & Patient, R. Hedgehog signaling is required for adult blood stem cell formation in zebrafish embryos. *Dev Cell* **8**, 389-400 (2005).
56. Korinek V et al. Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nat Genet* **19**, 379-383 (1998).
57. Morcillo, P., Rosen, C., Baylies, M. K. & Dorsett, D. Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in Drosophila. *Genes Dev* **11**, 2729-2740 (1997).
58. Ramain P et al. Interactions between chip and the achaete/scute-daughterless heterodimers are required for pannier-driven proneural patterning. *Mol Cell Biol* **6**, 781-790 (2000).
59. Wang J, Hoshino T, Redner RL, Kajigaya S & Liu, J. ETO, fusion partner in t(8;21) acute myeloid leukemia, represses transcription by interaction with the human N-CoR/mSin3/HDAC1 complex. *Proc Natl Acad Sci U S A* **95**, 10860-10865 (1998).
60. Gelmetti V et al. Aberrant recruitment of the nuclear receptor corepressor-histone deacetylase complex by the acute myeloid leukemia fusion partner ETO. *Mol Cell Biol* **18**, 7185-7191 (1998).
61. Lutterbach B et al. ETO, a target of t(8;21) in acute leukemia, interacts with the N-CoR and mSin3 corepressors. *Mol Cell Biol* **18**, 7176-7184 (1998).
62. Davis JN, McGhee L & Meyers, S. The ETO (MTG8) gene family. *Gene* **303**, 1-10 (2003).
63. Davis JN, Williams BJ, Herron JT, Galiano FJ & Meyers, S. ETO-2, a new member of the ETO-family of nuclear proteins. *Oncogene* **18**, 1375-1383 (1999).
64. Graña X et al. PITALRE, a nuclear CDC2-related protein kinase that phosphorylates the retinoblastoma protein in vitro. *Proc Natl Acad Sci U S A* **91**, 3834-3838 (1994).
65. Garriga J et al. CDK9 is constitutively expressed throughout the cell cycle, and its steady-state expression is independent of SKP2. *Mol Cell Biol* **23**, 5165-5173 (2003).
66. Marshall, N. F., Peng, J., Xie, Z. & Price, D. H. Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J Biol Chem* **271**, 27176-27183 (1996).
67. Zhu, Y. et al. Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro. *Genes Dev* **11**, 2622-2632 (1997).
68. Wei P, Garber ME, Fang SM, Fischer WH & Jones, K. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding toTAR RNA. *Cell* **92**, 451-462 (1998).
69. Marshall NF, Peng J, Xie Z & Price, D. Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J Biol Chem* **271**, 27176-27183 (1996).
70. Zhu Y et al. Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro. *Genes Dev* **11**, 2622-2632 (1997).
71. Mancebo HS et al. P-TEFb kinase is required for HIV Tat transcriptional activation in vivo and in vitro. *Genes Dev* **11**, 2633-2644 (1997).
72. Whyatt DJ et al. The level of the tissue-specific factor GATA1 affects the cell-cycle machinery. *Genes Funct* **1**, 11-24 (1997).

73. Giraud S, Hurlstone A, Avril S & Coqueret, O. Implication of BRG1 and cdk9 in the STAT3-mediated activation of the p21waf1 gene. *Oncogene* **23**, 7391-7398 (2004).
74. Kenny DA, Jurata LW, Saga Y & Gill, G. Identification and characterization of LMO4, an LMO gene with a novel pattern of expression during embryogenesis. *Proc Natl Acad Sci U S A* **95**, 11257-11262 (1999).
75. de la Calle-Mustienes E et al. Xenopus Xlmo4 is a GATA cofactor during ventral mesoderm formation and regulates Ldb1 availability at the dorsal mesoderm and the neural plate. *Dev Biol* **264**, 564-581 (2003).
76. Lane ME, Runko AP, Roy NM & Sagerström, C. Dynamic expression and regulation by Fgf8 and Pou2 of the zebrafish LIM-only gene, lmo4. *Gene Expr Patterns* **2**, 207-211 (2002).
77. Antoniou, M. in *Methods Mol Biol, Vol.7: Gene Transfer and Expression Protocols* (ed. Murray, E. J.) 421-434 (The Humana Press Inc., NJ, 1991).
78. Andrews NC & Faller, D. A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. *Nucleic Acids Res* **19**, 2499 (1991).
79. Westerfield, M. *The Zebrafish Book: A Guide for the Laboratory Use of Zebrafish (Brachydanio rerio).* (University of Oregon Press, Eugene, Oregon, USA, 1993).
80. Quinkertz, A. & Campos-Ortega, J. A new beta-globin gene from the zebrafish, betaE1, and its pattern of transcription during embryogenesis. *Dev Genes Evol* **209**, 126-131 (1999).
81. Jowett T, Y. Y. Double fluorescent in situ hybridization to zebrafish embryos. *Trends Genet* **12**, 387-389 (1996).



Supplementary Figure 1: Ldb-1 forms independant complexes in erythroid cells.
Proteins tested in sequential immunodepletion are depicted on the left.

Chapter 6:

Discussion

After the complete sequencing of the human genome as well as the genomes of other species, the focus of the molecular biology has shifted towards the study of gene function and of the protein complement of the cell, the proteome. Comparison of the sequenced genomes has revealed a small difference in gene numbers between organisms. For example, the Drosophila and human genomes contain 12,000 and 25,000 genes, respectively. This difference in gene numbers increases the complexity of the proteome of the organisms. In studying proteome complexity, or proteomics, powerful but also simple biochemical techniques are required for protein complex isolation and identification, for example in elucidating transcriptional regulation in higher vertebrates. Efficient protein isolation from crude extracts requires simple, high-affinity purification techniques. In addition, sensitive mass spectrometry methods that identify peptide sequences have become an integral part of protein complex identification strategies.

Biotinylation tagging

In this thesis, we demonstrated that biotinylation tagging is a very efficient approach to isolate and characterize protein complexes. We applied this technique in mammalian cells to isolate, in one step, several protein complexes even of low abundance. As described in this thesis, GATA-1 and Ldb1 complexes were isolated in erythroid cells. One of the main advantages of the single step biotin/streptavidin purification system is the very high affinity binding of biotin to avidin. This reduces the risk of protein partners dissociating from complexes during manipulations as is often found with conventional multi-step protein purification protocols. In addition, this method allows a more direct and rational approach for protein complex purification in describing most of the partners (repressing and activating), compared to previous studies describing similar functions but fragmented in different approaches in several studies, as it was shown for TAL-1 for example[1,2]. Biotinylation tagging also potentially offers an advantage in ChIP experiments due to a higher enrichment compared to antibodies, presumably due its very high affinity for streptavidin (E. Katsantoni, E. de Boer and K. Kolodziej, pers. comm. and Viens 2003). In addition, the use of biotinylation tagging/streptavidin in ChIP could alleviate the variability observed in these assays using different antibodies. Furthermore, biotinylation tagging of specific transcription factors can also be applied in ChIP-on-chip approaches (i.e. using ChIPed DNA to hybridize to genomic arrays), with the aim of identifying transcriptional gene target networks. Thus, biotinylation tagging has the potential to serve as a uniform platform for characterizing the protein complexes as well as gene targets of any given (tagged) transcription factor.

Recently, biotinylation tagging was also successfully used in diverse applications on a genome-wide scale. For example, a biotin-tagged H3.3 histone variant was used to describe the dynamics of its incorporation in chromatin[3]. In addition to its use in isolating nuclear proteins, the cell surface protein epidermal growth factor receptor (EGFR) was also shown to be efficiently biotinylated[4]. This approach therefore can be used to isolate cell populations. Another biological application of biotin-tagging a cell surface protein is based on the use of streptavidin-labeled quantum dots (QD). QD are nano-particles that allow protein visualization *in vivo*[5]. Furthermore, another large-scale application of the biotinylation tagging system is to isolate all proteins that are covalently modified *in vivo* by ubiquitin or SUMO by biotin-tagging these modifiers.

Lastly, the biotinylation tagging method has been improved at different levels. Higher expression of the bacterial BirA biotin ligase in mammalian cells was obtained after mammalian codon optimization[6]. Along the same lines, a shorter sequence of the peptide tag has been used successfully[7], thus reducing the risk of interference with functionality of the protein, even

though we have not seen so far any differences in the properties of tagged GATA-1 or Ldb1. The application of biotinylation tagging in the characterization of transcription factor complexes in erythroid cells, and particularly for GATA-1, has provided novel insight as to the molecular basis of their functions, as discussed further below.

What's new with GATA-1 function: regulation of transcription programs

Several lines of evidence have implicated GATA-1 as a key regulator of megakaryocytic/erythroid lineage (Meg/E) differentiation and of their associated transcription programs[8-10]. This has been demonstrated by the ectopic expression of GATA-1 in progenitor and non-Meg/E cells[11,12]. Significantly, GATA-1 expression in primary multipotential progenitors such as hematopoietic stem cells (HSC), common lymphoid progenitor (CLP) and common myeloid progenitor (CMP) was shown to induce expression of a subset of megakaryocytic and erythroid genes[13]. The activation of the Meg/E program was accompanied by the suppression of alternative lineage-affiliated programs. For example, CLP reprogramming by GATA-1 towards a Meg/E fate was characterized by the repression of genes associated with the lymphoid transcription program[13]. GATA-1 has also been shown to directly repress the GATA-2 gene in terminal erythroid differentiation[14]. Considering the essential GATA-2 functions in HSC and progenitor proliferation[15], this finding implicates GATA-1 in the repression of early multipotential transcription programs. Taken together, these observations suggest that GATA-1 functions during differentiation to concomitantly activate the Meg/E program and repress transcription programs associated with hematopoietic multipotentiality and alternative (non-Meg/E) lineages.

FOG-1 has been implicated as a key co-factor in these GATA-1 functions. FOG-1 functions were revealed by gene knockout studies, but also by the ectopic expression of FOG-1 in *in vitro* differentiated eosinophilic cells where FOG-1 is not normally expressed. Thus, FOG-1 expression led to the reprogramming of these cells towards an earlier, less differentiated state with the concomitant repression of eosinophilic markers such as EOS47 or MBP[16,17]. Considering the important GATA-1 functions in eosinophilic differentiation[18], it is of interest that the expression of FOG-1 in these cells appears to dominantly divert GATA-1 away from its normal functions. In addition, the repression of the GATA-2 gene has also been shown to require an interaction between GATA-1 and FOG-1. Thus FOG-1 is implicated as an important co-factor in the suppression of early transcription programs[19].

Our data on the GATA-1/FOG-1/MeCP1 complex provide additional insight as to the molecular basis of the GATA-1/FOG-1 functions. We showed that the recruitment by GATA-1 and FOG-1 of the MeCP1 repressive complex to the GATA-2 locus is responsible for its repression in terminal erythroid differentiation (Chapter 4). In addition, the detection by ChIP assays of the GATA-1/FOG-1/MeCP1 complex bound *in vivo* to repressed genes associated with non-Meg/E lineages, such as the eosinophilic MBP and IL-5Rα and the myeloid PU.1 genes, implicate this complex in the silencing of non-Meg/E lineage programs. Thus, we have identified the MeCP1 complex as an important partner in mediating the repressive GATA-1 and FOG-1 functions in erythroid differentiation and we suggest that the GATA-1/FOG-1/MeCP1 complex plays a major role in the suppression of multipotential and alternative lineage transcription programs, which represent an integral part of GATA-1 function (see also below). In providing more definitive evidence for a role for the MeCP1 complex in erythropoiesis, (or for any other newly identified protein partner, for that matter) one would need to carry out gene knockdown or conditional knockout experiments. In fact, we did employ RNAi to knockdown selected components of the MeCP1 complex, such as Mi2β, MTA2 and others. However, this approach proved inconclusive

since cells transfected with these constructs either died or showed no effect (A. Swiatek, J. Campbell and J. Strouboulis, pers. communication). We also analyzed fetal liver cells from MBD2 knockout embryos for defects in erythroid cell differentiation and maturation. MBD2 knockout mice are essentially viable and fertile[20]. We did not detect any significant differences in erythroid proliferation or differentiation between wild type, heterozygous and homozygous MBD2 knockout embryos (data not shown). However, it should be noted that MBD2 knockout mice still express a truncated form of the MBD2 protein, suggesting that the remaining protein product could still be functional. Since MBD3 has been previously shown to play key roles in the assembly and properties of the MeCP1/NuRD complex[21] and the MBD3 gene knockout is early embryonic lethal[20], suggesting essential functions in development, we are now investigating a conditional knockout of the MBD3 gene for defects in erythropoiesis. Another approach that was used successfully to ascribe functions in hematopoiesis of newly identified partners of key transcription factors was based on the use of morpholinos for Ldb1 and its partners in Zebrafish (Chapter 5). This approach is based on the inhibition of mRNA translation in Zebrafish embryos. However, this may not be suitable in the functional analysis of more general co-factors or complexes, such as MeCP1/NuRD, due to pleiotropic effects (i.e. morpholinos or antisense). Nevertheless, ChIP assays combined with the characterization of protein complexes provide invaluable insight into transcription factor function.

The observation that the ectopic expression of FOG-1 by itself in progenitors such as HSCs, CLPs and CMPs, does not induce a Meg/E program[13] suggests that FOG-1 is found downstream of GATA-1 in eliciting the aforementioned repressive functions. This is supported by the observations on the ectopic FOG-1 expression and reprogramming of the GATA-1-expressing eosinophilic cells[17] and, more importantly, by the observation that GATA-1 expression in CLPs leads to an increase in FOG-1 expression. In agreement with these observations, it was recently shown that the FOG-1 promoter was indeed bound by GATA-1 in erythroid cells *in vivo*[22]. Thus, a key event in Meg/E differentiation would be the upregulation/stabilization of GATA-1 expression, which upregulates the expression of FOG-1 that in turn synergies with GATA-1 in repressive or activating functions complexes.

The stable binding of the GATA-1/FOG-1/MeCP1 complex to known repressed genes in differentiated erythroid cells implies that these genes are actively suppressed, even after the cell has made a fate decision to become erythroid. We can extrapolate that this complex stays bound to target genes during cell differentiation. These observations provide a clear demonstration of how GATA-1 regulates gene expression at the chromatin level and suggest that even after chromatin remodeling (i.e. closing down) has occurred, the transcription factor remains bound to DNA. Data from Kim and colleagues have also shown that Ikaros, an essential transcription factor for lymphocyte development, interacts with the NuRD complex in differentiated T cells and propose that the complex remains stably bound to DNA throughout the cell cycle[23]. In addition, previous observations on the repressed GATA-2 locus showed that GATA-1 and FOG-1 remained bound even after a wave of histone deacetylation had occurred across the locus[14]. These observations suggest that the shutting down of alternative lineage transcription programs may not be irreversible in differentiated hematopoietic cells. Thus, cell reprogramming could be explained if the “resident” transcription factor/chromatin remodeling complexes are competed off the genes by ectopically expressed key transcription factors and co-factors. For example, in eosinophils, where GATA-1 interacts with C/EBPβ and FOG-1 is not expressed, the forced expression of FOG-1 inhibits expression of eosinophilic genes[17]. The expression of FOG-1 displaces the preformed GATA-1 complexes in eosinophils, implying that the interaction

between FOG-1 and GATA-1 must be dominant. In light of our data, we hypothesize that GATA-1 interacts in eosinophils with ectopic FOG-1 and associates with MeCP1 to (initially) repress the eosinophilic program and promote Meg/E differentiation. It would therefore be of interest to look at the FOG-1/MeCP1 occupancy of the EOS47 and MBP promoters in eosinophilic cells following FOG-1 expression. It would also be interesting to investigate the mechanism by which FOG-1 remains silent in eosinophils. Considering that GATA-1 binds to the FOG-1 promoter in erythroid cells[22] we would speculate that GATA-1 also mediates FOG-1 repression in eosinophils, perhaps through its interactions with C/EBP factors. This could be tested by ChIP experiments in the FOG-1 locus in eosinophils isolated from mice.

What's new with GATA-1 function: activation versus repression

The characterization of GATA-1 complexes provides a clear example of how a key transcription factor can carry out both repressive and activating functions. The GATA-1/FOG-1 interactions provide an interesting case in point. For example, whereas we have provided clear examples of the repressive functions of the GATA-1/FOG-1/MeCP1 complex, some of the activating functions of GATA-1 could take place in the context of GATA-1/FOG-1 interactions that are independent of the MeCP1 complex. ChIP assays have previously shown GATA-1/FOG-1 to bind *in vivo* to active genes such as the α-globin locus and the GATA-1 gene itself[19,24]. Several reports have shown that disruption of the GATA-1 and FOG-1 interaction down-regulates erythroid gene expression such as α and β globin, Band 3, DC11 and HD2[10,22,25]. During erythropoiesis, the β-major gene is activated only after protein synthesis of FOG-1 and not immediately upon onset of GATA-1 expression[22]. In addition, the authors show a direct binding of GATA-1 to the FOG-1 locus. These data show that some of the activating functions of GATA-1 in erythroid cells require FOG-1. So we can also imagine a GATA-1/FOG-1 complex involved in activating genes that dictate a specific cell fate, with GATA-1 playing a dominant role. Importantly, this complex would be active in the same cells in parallel to the repressive GATA-1/FOG-1/MeCP1 complex. Given the FOG-1 size fractionation profile on a superose 6 column (Chapter 3), we can expect additional protein partners in a GATA-1/FOG-1 complex that would synergize in transcriptional activating functions. Significantly, we (Chapter 3) and Blobel and colleagues have shown that only the large isoform of FOG-1 interacts with MeCP1, thus providing a mechanism for the formation of repressive or activating GATA-1/FOG-1 complexes[26] (see also below).

In addition to the GATA-1/FOG-1 activating complex, we have provided an example of a different activating GATA-1 complex in association with TAL-1/Ldb1 binding to the enhancer of the EKLF gene, the expression of which is essential for erythropoiesis and globin gene activation[27]. Several lines of evidence have previously shown GATA-1 to be part of a pentameric complex containing TAL-1, Ldb1, Lmo2 and E2A[28]. We have not investigated the other partners of the pentameric complex, but it is possible that Ldb1, E2A and Lmo2 are also present on the EKLF enhancer. However, it is also of note that expression of a TAL-1 mutant unable to bind DNA in erythroid cells does not affect expression of genes such as EKLF or GATA-1[29]. It would thus be of interest to determine whether EKLF is activated in the erythroid specific TAL-1 knockout.

A critical question arising from these studies is how these complexes that contain shared protein partners can exist at the same time in the cell, and how they are regulated? DNA sequence at regulatory elements, specific protein partners and protein posttranslational

modifications (PTMs) are the three key regulatory levels that can determine the formation and binding of a protein complex and its effects on transcription. Our work provides examples of how DNA sequences and specific protein partners can regulate the function of a protein complex. For example, ChIP analysis of the GATA-2 locus showed binding of two distinct GATA-1 complexes, one that contains the repressive MeCP1 complex binding to the -2.8 kb region and another GATA-1 complex binding to the -3.9 kb region that does not include MeCP1 ([30] and Chapter 4). Interestingly, the -2.8kb and -3.9kb GATA-1 binding sites are DNase I hypersensitive when the GATA-2 locus is active, but once GATA-1 binds (and represses GATA-2) the -2.8kb site loses DNase I sensitivity. The -3.9kb HS remains sensitive. This suggests that the binding of MeCP1 to the -2.8 kb modifies the chromatin structure into a closed, less sensitive conformation, also characterized by a wave of histone deacetylation[14]. Taken together, these data support a model whereby GATA-1 can interact in the context of two different complexes binding at the same time on different sites in the same locus.

DNA sequence analysis of the GATA-2 locus revealed that sequences adjacent to the -2.8kb region (that binds GATA-1/FOG-1/MeCP1) and to the -3.9kb region (that binds GATA-1) differ in consensus binding sites. For example, the -3.9kb region contains multiple SP1 binding sites. Interestingly, it was shown that SP1 binding sites were necessary to enhance transcription of genes such as TAL-1 and c-kit[31,32]. In addition, Sp1 binding motifs occur frequently together with GATA motifs in hematopoietic regulatory elements and GATA-1 has been reported to interact directly with Sp1 by EMSA and co-transfections[33]. These observations suggest that SP1 binds with GATA-1 at the -3.9kb element, thus preventing the binding/spreading of the GATA-1/FOG-1/MeCP1 complex. This is supported by the observations of Bresnick and colleagues who showed that during GATA-2 repression by GATA-1 the levels of histone H3 acetylation were greatly reduced at the -2.8kb region, whereas deacetylation at the -3.9kb region was more moderate[14]. To provide insight into the regulation of the different GATA-1 complexes binding to the GATA-2 locus, mutations of either of the GATA binding sites at -3.9kb and -2.8kb should be made, or these sequences should be swapped and assayed for their effects on GATA-2 expression.

Another example of the importance of DNA sequence in the regulation of protein complexes binding is the GATA-1 locus itself. GATA-1 was shown to activate its own transcription[34]. However, the down regulation of GATA-1 expression was shown to be necessary for terminal erythroid differentiation[35]. ChIP analysis of the GATA-1 regulatory sequences showed that the DNase I HS located at the -3.5 kb region of the GATA-1 locus is bound by the pentameric complex including GATA-1 and TAL-1. Whereas GATA-1 binds also to the erythroid IE promoter and the HS located at the +3.5 kb region, the TAL-1 complex was not detected[36]. This shows that different protein complexes regulate the GATA-1 locus at the same time, highlighting again the importance of the combination of interacting partners to define the expression pattern of a gene. Thus, the underlying DNA sequences are likely to be an important determining factor in the binding specificity of the GATA-1 protein complexes.

A second level for regulating protein complex formation can be through the proteins themselves. Our work from GATA-1 and Ldb1 complexes purification provide several examples. FOG-1 is detected as two isoforms of about 110 and 140 kDa. We and the Blobel group showed that only the larger FOG-1 isoform that contains a N-terminal 45aa domain interacts with the MeCP1complex, suggesting that it is sufficient for binding the NuRD/MeCP1 complex (Chapter 3 and [26]). In addition, a similar mechanism was observed in the formation of Ldb1 complexes, in which Eto-2, a novel Ldb1 partner, preferentially interacts with the larger isoform

of E47 (Chapter 5) which could influence transcriptional activation.

The presence or absence of a protein partner in the complex could also influence the activity of the complex. Data obtained from the Ldb1 purification demonstrate that Ldb1 forms multiprotein complexes with Cdk9 in non-induced MEL cells but not in induced MEL cells. Similarly, Ldb1 forms protein complexes with Eto2, which include GATA-1. However, the ratio of Eto-2 in Ldb1 complexes changes during erythroid differentiation, suggesting that Ldb1 complexes are dynamic during differentiation. As the Eto-2/Ldb1 ratio decreases in complexes bound to activated genes, we suggest that Eto-2 is having a repressing function on TAL-1 target genes that diminishes with erythroid differentiation.

A third regulatory level relies on PTMs that could influence transcription factor complex formation. GATA-1, and a number of other transcription factors, have been shown to be modified by acetylation, phosphorylation and sumoylation. For example, acetylation of GATA-1 by CBP/P300 was linked to increased transcription activity[37,38]. However, the molecular mechanism remains an open question. While Boyes and colleagues have shown that acetylation of chicken GATA-1 increased its affinity for binding to DNA, Hung and colleagues found no such increase with acetylated murine GATA-1[39]. Nonetheless posttranslational modifications can affect the properties of proteins by most likely altering their three-dimensional conformation, and thus subsequently influencing the binding of other protein partners. For example, phosphorylation of serine 26 and 178 in the GATA-1 protein is critical for interactions of GATA-1 with Lmo2[40]. In addition, GATA-1 was shown to be sumoylated in mammalian cells[41]. However, no correlation between GATA-1 sumoylation and GATA-1 function has yet been found. Nevertheless, interaction with PIASy ligase was shown to repress GATA-1 transcriptional activity, albeit independently of sumoylation[41].

A striking observation from our studies on the characterization of GATA-1 and Ldb1 complexes in erythroid cells is the higher abundance of repressive complexes compared to activating complexes. This observation suggests that activating complexes would be more labile. Recent studies have linked ubiquitination-mediated degradation of transcriptional activators to their activating functions[42]. This suggests a very fast turnover of the activating proteins. However, this mechanism alone cannot explain our results. We could not identify the co-factors of the activating transcription factor complexes, but we did detect by ChIP of formaldehyde crosslinked chromatin a stable binding of the activating factors, suggesting that the transcription factors are not degraded. Taken together, these results suggest a higher degree of complexity in the regulation gene activation, for which the mechanism(s) remain to be explained.

What's new with GATA-1 function: a key regulator of cell cycle progression

GATA-1 binds to retinoblastoma and regulates expression of cell cycle related genes such as c-myb and c-myc[43,44]. During cell cycle progression, c-Myc promotes the G1/S phase transition by inducing cyclin E-CDK2, which results in an increase in active E2F transcription complexes that promote cell cycle progression. Conversely, repression of c-myc gene expression is required for G1 arrest[45,46]. The binding of GATA-1 with Gfi-1b to the c-myc and c-myb promoters provides clues as to how GATA-1 suppresses cell proliferation in differentiating MEL cells through G1 arrest, and explains part of the GATA-1 function in terminal differentiation. Our data agree with recent microarray data obtained in G1E-ER4 cells suggesting that the induced GATA-1 expression represses the c-myc and c-myb genes in terminal erythroid differentiation[10,44].

In addition, genes coding for cyclin-dependent kinases (cdk), such as cdk2, 4 and 6 that are functionally associated with cell cycle progression, were inhibited by GATA-1 expression, whereas expression of kinases inhibitors such as p18 and p27 were induced[44]. Our data also provide insight as to Gfi-1b functions, since Gfi-1b was also shown to regulate the cell cycle in myeloid cells[47]. Our data also provide clues as to how GATA-1 overexpression leads to a failure in the G1 arrest in erythroid cells[43]. We speculate that GATA-1 overexpressing mice and cells exhibit high level of expression of the c-myc and c-myb genes, that would lead to the failure of G1 arrest. Interestingly, the forced expression of Gfi-1b in myeloid cells blocks G1 arrest and prevents the down-regulation of c-myc and c-myb thus inhibiting myeloid cell differentiation[47]. In the light of previous and our results, we propose that GATA-1 exerts two levels of regulation of the cell cycle, one by suppressing c-myc and c-myb genes by direct binding to their promoters, and a second by direct binding to Rb.

The mechanism by which GATA-1 and Gfi-1b mediate repression is unclear. Gfi-1b contains a N-terminal SNAG domain that has been implicated in recruiting the chromatin modifying Sin3A/HDAC1 repressive complex[48]. In addition, deletion of the SNAG domain abolishes Gfi-1b-induced erythroid maturation as well as myeloid differentiation, showing that it is an essential functional domain[47,49]. Thus, it is possible that the GATA-1/Gfi-1b complex we purified contains additional co-factors that modulate their repressing function, e.g. at the chromatin structure level. However, due to the low abundance of the GATA-1/Gfi-1b complex we were not able to identify them. It is possible that part of the GATA-1 complexes we identified interacting with HDACs may also involve Gfi-1b. These interactions would be in addition to the GATA-1 interactions with MeCP1. In fact, the superose 6 gel filtration profiles of Gfi-1b and GATA-1 support this hypothesis, as Gfi-1b and GATA-1 are eluted in a higher molecular weight range than expected for a complex formed only by GATA-1/Gfi-1b. This would provide evidence for another large protein complex involving GATA-1 in gene repressing function.

GATA-1 and GATA-2

The N-terminal zinc finger (NF) is the main GATA-1 protein-protein interaction domain (Our work and [10,26,39]). In contrast to the DNA binding C-terminal zinc finger domain (CF), the NF domain is not as well conserved between the different members of the GATA family of transcription factors. Thus, differences in NF sequences could be responsible for differentially mediating interactions with protein partners that are specific to individual GATA factors. This may account for the fact that the expression of the hematopoietic GATA-2 transcription factor only partially rescues the GATA-1 low phenotype[50]. Rescue experiments in GATA-1.05 mice with GATA-2 expressed under the 3.9kb regulatory sequences of GATA-1 (shown to be sufficient to drive similar pattern of expression as wild type GATA-1) showed a rescue of embryonic lethality. However, mice developed severe anemia in the adult age[50]. To provide insight into the lack of the complete rescue by GATA-2 of the GATA-1 null phenotype, the GATA-2 protein complex was purified from erythroid cells using biotinylation tagging (H. Braun, unpublished data). Analysis of isolated complexes showed that GATA-2 interacts with some, but not all, of the GATA-1 partners. Whereas GATA-2 was found to interact with FOG-1 and members of the MeCP1/NuRD complex, transcription factors such as TAL-1, Ldb1, Lmo2 and Gfi-1b were absent from the GATA-2 pull-down. A number of novel GATA-2 partners were also identified which do not overlap with GATA-1. Thus, GATA-1 and GATA-2 interact with overlapping but also distinct partners. This can in part explain why GATA-2 does not completely compensate for loss of GATA-1 in mice. Our data support the notion that GATA-2 and GATA-1 have at least some

overlapping functions.

Analysis of mice expressing either a GATA-1 mutant unable to bind FOG-1, or a combination of GATA-1 and GATA-2 mutants that are unable to bind FOG-1 show different degree of severity in megakaryocytic defects[51]. If both factors cannot interact with FOG-1, the megakaryocytic lineage is completely absent, whereas in the presence of GATA-2 only the maturation of megakaryocytes is impaired. In conclusion, although GATA-1 and GATA-2 recognize a similar DNA binding sequence and share partly overlapping expression profiles, they play unique as well as overlapping roles *in vivo* reflected by this protein interactions.

GATA-1 complex interactions: re-examination of knockout phenotypes

Our findings on the distinct GATA-1 complexes give some insight on the phenotypes observed with the knockouts of key GATA-1 interacting partners, such as FOG-1, Gfi-1b, TAL-1 and Ldb1. One striking observation that emerges from the analysis of the knockouts models for GATA-1, FOG-1, Gfi-1b and TAL-1 is the close similarity they present in erythroid and megakaryocytic differentiation[9,52-55]. However, some phenotypic differences suggest that these transcription factors possess specific functions in erythropoiesis. We have suggested a direct function for GATA-1 in the suppression of cell proliferation based on our findings that a GATA-1/Gfi-1b complex binds to the repressed c-myc and c-myb genes in differentiated erythroid cells. The GATA-1 gene knockout leads to a complete block of erythroid maturation at the pro-erythroblast stage due to apoptosis, whereas erythroid progenitors in Gfi-1b mutants continue to proliferate[8,53,55]. We can speculate that in the absence of Gfi-1b protein, the c-myc and c-myb genes would not be repressed and erythroid progenitors would therefore continue to proliferate. These data would also suggest that Gfi-1b functions in erythropoiesis do not overlap with the GATA-1 anti-apoptotic functions. However, it is of note that erythroid progenitors in Gfi-1b mutants are blocked at the BFU-E stage. This phenotype resembles that of TAL-1 conditional knockout that also leads to a block of erythroid differentiation at a similar stage[54,56]. These observations suggest an essential role for Gfi-1b and TAL-1 that is GATA-1 independent early in erythroid differentiation. It is also possible that the two transcription factors Gfi-1b and TAL-1 are involved, in parallel, in a common pathway, as it was shown that the Drosophila Gfi-1b ortholog *senseless* interacts with bHLH proteins in sensory organ development[57].

FOG-1 knockout embryos exhibit a severe, but partial, block of erythroid differentiation at the pro-erythroblast stage. Cells survive for longer time during erythroid maturation as compared to the GATA-1 knockout[52]. It is remarkable that despite the similar phenotypes between the Gfi-1b and the FOG-1 knockouts in the erythroid and megakayocytic lineages, we have not seen any direct interactions between these two factors in erythroid cells. This raises the possibility that FOG-1 and Gfi-1b are involved in distinct but parallel pathways or in common pathways. As erythroid defects occur earlier in Gfi-1b mutants than in FOG-1 mutants, we can think that FOG-1 is temporally playing a role downstream of Gfi-1b.

Future directions

As a conclusion, we can picture GATA-1 as a orchestra conductor. It interacts with many partners recruiting effector molecules for the modification of chromatin structure in the proper temporal and spatial gene regulation of several target genes (Figure 1 and 2). The regulation of the distribution of the co-factors between GATA-1 complexes remains to be understood. However, our data suggest that GATA-1 uses chromatin remodeling complexes such as MeCP1 to repress genes early during erythroid differentiation and remain bound

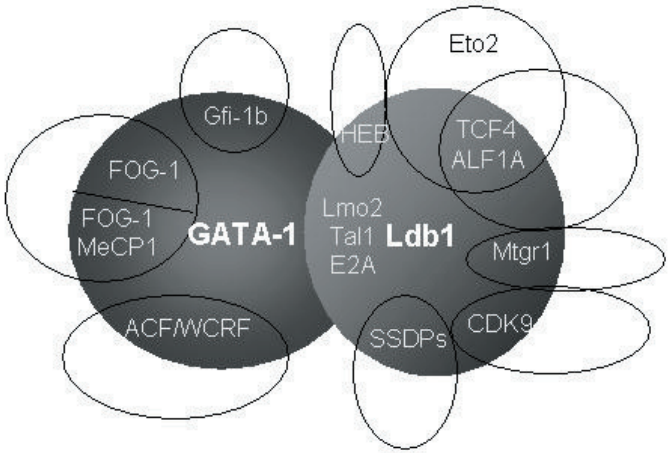


Figure 1: GATA-1 and Ldb-1 protein partners in erythroid cells.
Proteins partners (white circles) of biotinylated GATA-1 and Ldb-1 (grey circles) form distinct protein complexes. In addition, they are involved in other protein-protein interactions that are GATA-1 or Ldb1 independent (non-overlapping circles).

at a late stage of terminal erythroid differentiation. Other GATA-1 complexes such as the GATA-1/FOG-1 and the GATA-1/Gfi-1b need further characterization to identify the co-factors that are involved in the transcriptional activity regulation. We have shown that the GATA-1/FOG-1/MeCP1 complex was able to repress in erythroid cells eosinophilic and myeloid lineage-affiliated genes. It would be interesting to investigate the regulation of more classes of genes in erythroid differentiation (e.g. the repression of megakaryocytic genes) to know whether the function of the GATA-1/FOG-1/MeCP1 complex would be broader than the one we propose in our study. In addition, as GATA-1 is also expressed in megakaryocytic cells with FOG-1, it would be interesting to know whether the GATA-1/FOG-1/MeCP1 complex exists in megakaryocytes and whether it represses erythroid genes. We would also speculate that in megakaryocytic cells eosinophilic and myeloid programs would also have to be switched off in a similar way as we have proposed in erythroid cells.

GATA-1/FOG-1 interactions are very important in Meg/E differentiation. Thus, a very interesting question is how does GATA-1 regulate the differentiation of megakaryocytic versus erythroid cells from the Meg/E progenitors, taking into account the considerable overlap in GATA-1 interactions (i.e. with FOG-1 or Gfi-1b) between these two lineages. This point raises the question as to what are the differences in GATA-1/FOG-1 and GATA-1/Gfi-1b complexes between erythroid and megakaryocytic cells, stressing again the importance of identifying protein partners in order to understand transcriptional regulation. In order to address these points, the GATA-1 complex purification in megakaryocytes using the biotinylation tagging method is presently under way (P. Vyas, Oxford).

In order to provide a broader picture of the protein network in hematopoiesis, a larger number of transcription factors and co-factors need to be tagged. We can therefore anticipate the reconstitution of a map of protein-protein inter-connections as has been described in yeast[58]. In addition to the distinct GATA-1 complexes in erythroid cells, purification of the protein complexes of the GATA-1 partner Ldb1 has also revealed the presence of distinct complexes with or without GATA-1 (Figure 1). These data expand on our knowledge of transcription factor

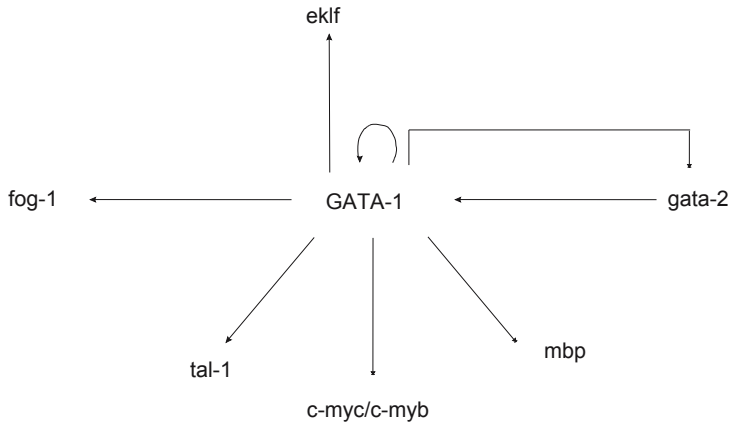


Figure 2: GATA-1 protein regulates several target genes in erythroid cells.
EKLF, FOG-1, TAL-1 and GATA-1 gene expression are activated by the transcription factor GATA-1. In contrast, GATA-2, c-Myc, c-Myb and MBP gene expression are suppressed. GATA-2 activates also GATA-1 gene.

complexes and provide more clues as to how transcription factors interact in hematopoiesis. For example, purification of FOG-1 protein complexes would give a better understanding of the FOG-1 functions that are GATA-1 independent, in erythroid but also megakaryocytic cells. FOG-1 was described to interact with TACC3 (Transforming Acidic Coiled Coil)[59]. This protein was not found in our GATA-1 purification, suggesting that FOG-1, like Ldb1, interacts with proteins independently of GATA-1.

The dynamics of protein complexes are an important aspect in the regulation of gene expression. We have seen that Ldb1 can form different complexes at different stages of erythroid differentiation. Similarly, we have preliminary results showing that GATA-1 complexes are also dynamic during erythroid cell differentiation. Thus, it would be interesting to purify GATA-1 complexes at an earlier stage of differentiation to identify potentially novel GATA-1 partners. It is also interesting to determine the relative abundance of GATA-1 complexes for example between non-differentiated and differentiated MEL cells to identify function of GATA-1 at an earlier stage of erythroid differentiation. To do so, GATA-1 associated proteins isolated from MEL cells at the two different stages of differentiation can be mixed after labeling one of the two populations of proteins and the proteins analyzed by Mass Spectrometry. A similar approach based on Isotope-coded Affinity Tag (ICAT) has already been used to purify NF-E2/p45 protein complexes[60].

Lastly, we can envisage developing a ChIP-on-chip approach based on the biotinylation method to address the question of gene target network regulated by the different transcription factor complexes. For instance, DNA sequences bound by GATA-1 and Ldb1 can be isolated form ChIPed material and hybridized with DNA promoter arrays to determine the sequences that are common and distinct to the two factors, thus providing more insight on their molecular functions in hematopoiesis.

References

1. Huang, S., Qiu, U., Stein, R. & Brandt, S. p300 functions as a transcriptional coactivator for the TAL1/SCL oncoprotein. *Oncogene* **18**, 4958-67 (1999).
2. Huang, S. & Brandt, S. mSin3A regulates murine erythroleukemia cell differentiation through association with the TAL1 (or SCL) transcription factor. *Mol Cell Biol.* **20**, 2248-59 (2000).
3. Mito, Y., Henikoff, J. & Henikoff., S. Genome-scale profiling of histone H3.3 replacement patterns. *Nat Genet* **37**, 1090-7 (2005).
4. Chen, I., Howarth, M., Lin, W. & Ting., A. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nature methods* **2**, 99-104 (2005).
5. Howarth, M., Takao, K., Hayashi, Y. & Ting., A. Targeting quantum dots to surface proteins in living cells with biotin ligase. *Proc Natl Acad Sci U S A* **102**, 7583-8 (2005).
6. Mechold, U., Gilbert, C. & Ogryzko, V. Codon optimization of the BirA enzyme gene leads to higher expression and an improved efficiency of biotinylation of target proteins in mammalian cells. *Journal of Biotechnology*, 245-9 (2005).
7. Driegen, S. et al. A generic tool for biotinylation of tagged proteins in transgenic mice. *Transgenic Res.* **14**, 477-82 (2005).
8. Weiss, M., Keller, G. & Orkin, S. Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. *Genes Dev* **8**, 1184-97 (1994).
9. Pevny, L. et al. Development of hematopoietic cells lacking transcription factor GATA-1. *Development* **121**, 163-72 (1995).
10. Crispino, J., Lodish, M., MacKay, J. & Orkin, S. Use of altered specificity mutants to probe a specific protein-protein interaction in differentiation: the GATA-1:FOG complex. *Mol Cell* **3**, 219-28 (1999).
11. Visvader, J., Elefanty, A., Strasser, A. & Adams, J. GATA-1 but not SCL induces megakaryocytic differentiation in an early myeloid line. *Embo J* **11**, 4557-64 (1992).
12. Heyworth, C., Pearson, S., May, G. & Enver, T. Transcription factor-mediated lineage switching reveals plasticity in primary committed progenitor cells. *Embo J* **21**, 3770-81 (2002).
13. Iwasaki, H. et al. GATA-1 converts lymphoid and myelomonocytic progenitors into the megakaryocyte/erythrocyte lineages. *Immunity* **19**, 451-62 (2003).
14. Grass, J. A. et al. GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. *Proc Natl Acad Sci U S A* **100**, 8811-6 (2003).
15. Tsai, F. & Orkin, S. Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. *Blood* **89**, 3636-43 (1997).
16. Yamaguchi, Y., Nishio, H., Kishi, K., Ackerman, S. & T., T. S. C/EBPbeta and GATA-1 synergistically regulate activity of the eosinophil granule major basic protein promoter: implication for C/EBPbeta activity in eosinophil gene expression. *Blood* **94**, 1429-39 (1999).
17. Querfurth, E. et al. Antagonism between C/EBPbeta and FOG in eosinophil lineage commitment of multipotent hematopoietic progenitors. *Genes Dev* **14**, 2515-25 (2000).
18. Yu, C. et al. Targeted deletion of a high-affinity GATA-binding site in the GATA-1 promoter leads to selective loss of the eosinophil lineage in vivo. *J Exp Med* **195**, 1387-95 (2002).
19. Pal, S. et al. Coregulator-dependent facilitation of chromatin occupancy by GATA-1. *Proc Natl Acad Sci U S A* **101**, 980-5 (2004).
20. Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V. A. & Bird, A. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes Dev* **15**, 710-23 (2001).
21. Zhang, Y. & Reinberg, D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev* **15**, 2343-60 (2001).
22. Welch, J. J. et al. Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood* **104**, 3136-47 (2004).
23. Kim, J. et al. Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes. *Immunity* **10**, 345-55 (1999).

24. Anguita, E. et al. Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. *Embo J* **23**, 2841-52 (2004).
25. Letting, D. L., Chen, Y. Y., Rakowski, C., Reedy, S. & Blobel, G. A. Context-dependent regulation of GATA-1 by friend of GATA-1. *Proc Natl Acad Sci U S A* **101**, 476-81 (2004).
26. Hong W, N. M., Chen YY, Kori R, Vakoc CR, Rakowski C, Blobel GA. FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. *Embo J* **24**, 2367-78 (2005).
27. Wijgerde, M. et al. The role of EKLF in human beta-globin gene competition. *Genes Dev* **10**, 2894-902 (1996).
28. Osada, H., Grutz, G., Axelson, H., Forster, A. & Rabbitts, T. H. Association of erythroid transcription factors: complexes involving the LIM protein RBTN2 and the zinc-finger protein GATA1. *Proc Natl Acad Sci U S A* **92**, 9585-9 (1995).
29. Porcher, C., Liao, E. C., Fujiwara, Y., Zon, L. I. & Orkin, S. H. Specification of hematopoietic and vascular development by the bHLH transcription factor SCL without direct DNA binding. *Development* **126**, 4603-15 (1999).
30. Martowicz, M., Grass, J., Boyer, M., Guend, H. & Bresnick, E. Dynamic GATA factor interplay at a multicomponent regulatory region of the GATA-2 locus. *J Biol Chem* **280**, 1724-32 (2005).
31. Lecointe, N. et al. GATA-and SP1-binding sites are required for the full activity of the tissue-specific promoter of the tal-1 gene. *Oncogene* **9**, 2623-32 (1994).
32. Lecuyer, E. et al. The SCL complex regulates c-kit expression in hematopoietic cells through functional interaction with Sp1. *Blood* **100**, 2430-40 (2002).
33. Merika, M. & Orkin, S. Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Kruppel family proteins Sp1 and EKLF. *Mol Cell Biol* **15**, 2437-47 (1995).
34. Vyas, P. et al. Different sequence requirements for expression in erythroid and megakaryocytic cells within a regulatory element upstream of the GATA-1 gene. *Development* **126**, 2799-811 (1999).
35. Whyatt, D. et al. An intrinsic but cell-nonautonomous defect in GATA-1-overexpressing mouse erythroid cells. *Nature* **406**, 519-24 (2000).
36. Valverde-Garduno, V. et al. Differences in the chromatin structure and cis-element organization of the human and mouse GATA-1 loci: implications for cis-element identification. *Blood* **104**, 3106-16 (2004).
37. Blobel, G. A., Nakajima, T., Eckner, R., Montminy, M. & Orkin, S. H. CREB-binding protein cooperates with transcription factor GATA-1 and is required for erythroid differentiation. *Proc Natl Acad Sci U S A* **95**, 2061-6 (1998).
38. Boyes, J., Byfield, P., Nakatani, Y. & Ogryzko, V. Regulation of activity of the transcription factor GATA-1 by acetylation. *Nature* **396**, 594-8 (1998).
39. Hung, H. L., Lau, J., Kim, A. Y., Weiss, M. J. & Blobel, G. A. CREB-Binding protein acetylates hematopoietic transcription factor GATA-1 at functionally important sites. *Mol Cell Biol* **19**, 3496-505 (1999).
40. Towatari, M., Ciro, M., Ottolenghi, S., Tzuzuki, S. & T, E. Involvement of mitogen-activated protein kinase in the cytokine-regulated phosphorylation of transcription factor GATA-1. *Hematol J* **5**, 262-72 (2004).
41. Collavin, L. et al. Modification of the erythroid transcription factor GATA-1 by SUMO-1. *Proc Natl Acad Sci U S A* **101**, 8870-5 (2004).
42. Lipford, J., Smith, G. & Deshaies, R. A putative stimulatory role for activator turnover in gene expression. *Nature* **438**, 113-6 (2005).
43. Whyatt, D. et al. The level of the tissue-specific factor GATA-1 affects the cell-cycle machinery. *Genes Funct* **1**, 11-24 (1997).
44. Rylski, M. et al. GATA-1-mediated proliferation arrest during erythroid maturation. *Mol Cell Biol* **23**, 5031-42 (2003).
45. Ho, J., Ma, W., Mao, D. & Benchimol., S. p53-Dependent Transcriptional Repression of c-myc Is Required for G1 Cell Cycle Arrest. *Mol Cell Biol.* **25**, 7423-31 (2005).
46. Arsura, M. & Sonenshein, G. E. in *Transcription Factors: Normal and Malignant Development of Blood Cells* (eds. Ravid, K. & Licht, J. D.) 521-549 (Wiley-Liss, Inc., New York, 2001).

47. Tong, B. et al. The Gfi-1B proto-oncoprotein represses p21WAF1 and inhibits myeloid cell differentiation. *Mol Cell Biol* **18**, 2462-73 (1998).
48. Peinado, H., Ballestar, E., Esteller, M. & Cano., A. Snail mediates E-cadherin repression by the recruitment of the Sin3A/histone deacetylase 1 (HDAC1)/HDAC2 complex. *Mol Cell Biol.* **24**, 306-19 (2004).
49. Garcon, L. et al. Gfi-1B plays a critical role in terminal differentiation of normal and transformed erythroid progenitor cells. *Blood* **105**, 1448-55 (2005).
50. Takahashi, S. et al. GATA factor transgenes under GATA-1 locus control rescue germline GATA-1 mutant deficiencies. *Blood* **96**, 910-6 (2000).
51. Chang, A. N. et al. GATA-factor dependence of the multitype zinc-finger protein FOG-1 for its essential role in megakaryopoiesis. *Proc Natl Acad Sci U S A* **99**, 9237-42 (2002).
52. Tsang, A. P., Fujiwara, Y., Hom, D. B. & Orkin, S. H. Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG. *Genes Dev* **12**, 1176-88 (1998).
53. Fujiwara, Y., Browne, C. P., Cunniff, K., Goff, S. C. & Orkin, S. H. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc Natl Acad Sci U S A* **93**, 12355-8 (1996).
54. Hall, M. A. et al. The critical regulator of embryonic hematopoiesis, SCL, is vital in the adult for megakaryopoiesis, erythropoiesis, and lineage choice in CFU-S12. *Proc Natl Acad Sci U S A* **100**, 992-7 (2003).
55. Saleque, S., Cameron, S. & Orkin, S. H. The zinc-finger proto-oncogene Gfi-1b is essential for development of the erythroid and megakaryocytic lineages. *Genes Dev* **16**, 301-6 (2002).
56. Mikkola, H. K. et al. Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene. *Nature* **421**, 547-51 (2003).
57. Nolo, R., Abbott, L. & Bellen, H. Senseless, a Zn finger transcription factor, is necessary and sufficient for sensory organ development in Drosophila. *Cell* **102**, 349-62 (2000).
58. Gavin AC, B. M., Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M, Hofert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier MA, Copley RR, Edelmann A, Querfurth E, Rybin V, Drewes G, Raida M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G, Superti-Furga G. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141-7 (2002).
59. Garriga-Canut, M. & Orkin, S. Transforming acidic coiled-coil protein (TACC3) controls Friend of GATA-1 (FOG-1) subcellular localization and regulates the association between GATA-1 and FOG-1 during hematopoiesis. *J Biol Chem* **279**, 23597-605 (2005).
60. Brand, M. et al. Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics. *Nat Struct Mol Biol* **11**, 73-80 (2004).

Appendix

Isolation of transcription factor complexes by in vivo biotinylation tagging and direct binding to streptavidin beads

Patrick Rodriguez, Harald Braun, Katarzyna E. Kolodziej, Ernie de Boer, Jennifer Campbell, Edgar Bonte, Frank Grosveld, Sjaak Philipsen and John Strouboulis

Book Chapter - In Press -
Methods in Molecular Biology
Gene Mapping, Discovery, and Expression
Methods and Protocols

Department of Cell Biology, Erasmus Medical Center, P.O. Box 1738, 3000DR Rotterdam, The Netherlands

Isolation of transcription factor complexes by in vivo biotinylation tagging and direct binding to streptavidin beads

Abstract

Efficient tagging methodologies are an integral aspect of protein complex characterization by proteomic approaches. Due to biotin's very high affinity for avidin and streptavidin, biotinylation tagging offers an attractive approach for the efficient purification of protein complexes. The very high affinity of the biotin/(strept)avidin system also offers the potential for the single-step capture of lower abundance protein complexes, such as transcription factor complexes. The identification of short peptide tags that are efficiently biotinylated by the bacterial BirA biotin ligase, led to an approach for the single-step purification of transcription factor complexes by specific in vivo biotinylation tagging. A short sequence tag fused N-terminally to the transcription factor of interest is very efficiently biotinylated by BirA co-expressed in the same cells, as was demonstrated by the tagging of the essential hematopoietic transcription factor GATA-1. The direct binding to streptavidin of biotinylated GATA-1 in nuclear extracts resulted in the single-step capture of the tagged factor and associated proteins, which were eluted and identified by mass spectrometry. This led to the characterization of several distinct GATA-1 complexes with other transcription factors and chromatin remodeling co-factors, which are involved in activation and repression of gene targets. Thus, BirA-mediated tagging is an efficient approach for the direct capture and characterization of transcription factor complexes.

Key words: biotinylation tagging; BirA; transcription factors; chromatin; mass spectrometry; size fractionation; GATA-1

Introduction

Completion of the sequencing of an ever-increasing number of genomes has led to a shift of focus towards the characterization of the protein complement of cells, i.e. the proteome. A key aspect of proteomic analysis is the development of simple methodologies for the efficient isolation of protein complexes for peptide analysis and identification by powerful mass spectrometric approaches. This is particularly challenging for the analysis of nuclear proteins involved in transcriptional regulation such as transcription factors and their chromatin associated co-factors due to their relatively lower abundance, the different parallel functions that they execute (e.g. activation and repression involving different partners) and the often transient nature of their interactions. Transcription factor purification approaches involving several pre-purification steps are laborious and costly and most likely result in the isolation of only the most abundant of the protein complexes formed by the factor. We describe here the application of in vivo biotinylation tagging as a simple approach for the efficient direct purification of transcription factor complexes from crude nuclear extracts (1).

Biotin is a naturally occurring cofactor essential for certain metabolic enzymes such as carboxylases. Specific protein-biotin ligases are responsible for covalently attaching biotin to these enzymes. The key to using biotinylation lies in the fact that biotinylated substrates can be bound very tightly by the naturally occurring proteins avidin and streptavidin ($K_d = 10^{-15}$), a fact that has been widely exploited in many affinity-based biochemical applications. In addition, in vivo biotinylation tagging offers a number of advantages for protein purification purposes. Firstly, there are few naturally biotinylated proteins (mostly cytoplasmic and mitochondrial) ensuring that non-specific background binding remains low. Secondly, the very high affinity of (strept)avidin for biotin allows high stringencies to be employed during purification without fear of losing binding of the tagged protein.

The biotinylation tagging approach described here is based on previous work on the screening of a combinatorial synthetic peptide library for efficient biotinylation by the bacterial BirA biotin ligase (2). This led to the identification of a number of short sequence tags that can be very efficiently biotinylated in vitro (2, 3). Such tags were subsequently utilized for the efficient in vivo biotinylation of tagged proteins in bacterial cells through the co-expression of BirA (4, 5). We have applied this approach in mammalian cells and demonstrated its efficiency in specifically biotinylating nuclear proteins in cultured cells and transgenic mice through the co-expression of the BirA biotin ligase together with the tagged protein (Figure 1A) (1). Our work is focused primarily on the biotinylation tagging of hematopoietic transcription factors in erythroid cells. Most of our work to date has been carried out with GATA-1, a critical transcription factor for erythroid cell differentiation. We have been able to very efficiently biotinylate GATA-1 in cultured mouse erythroleukemic cells (Figure 1B and C; (1)) leading to the isolation and characterization of GATA-1 protein complexes by direct binding of nuclear extracts to streptavidin beads. Using this approach we identified a number of GATA-1 complexes, containing other essential hematopoietic transcription factors (FOG-1, Gfi-1b and TAL-1) and chromatin remodeling and modification complexes (MeCP1 and ACF/WCRF). These complexes were implicated in the transcriptional activation and repression of different subsets of target genes (6). Thus, biotinylation tagging has proven to be a very efficient method for the single-step purification and characterization of transcription factor complexes. It should also be noted that we have no evidence so far that biotinylation tagging adversely affects the physiological properties of the tagged protein (1).

In this chapter we describe protocols for the binding of nuclear extracts expressing a specific biotin-tagged protein to streptavidin beads and the preparation of the eluted material for analysis by mass spectrometry. We do not provide protocols for the stable transfection of cultured cells as these will vary depending on the cell line/type used in each case. We routinely prepare large scale nuclear extracts from a few liters of cultured cells, we test for the presence of the biotin tagged protein in high molecular weight complex(es) by gel filtration using a Superose 6 column and then carry out the binding of the tagged factor to streptavidin paramagnetic beads. We normally check the efficiency of the biotin tagging and the binding to streptavidin beads by testing the nuclear extract (input), the bound material (eluate)

and the flowthrough (unbound) by Western blotting using first an antibody against the tagged protein followed, by streptavidin-HRP conjugate on the same blot. In this way the fraction of the tagged protein that becomes biotinylated in vivo and subsequently captured by the streptavidin beads can be determined. As shown in Figure 1, for GATA-1 the biotinylation efficiency and capture by the beads is nearly 100%. The proteins eluted from the beads are fractionated by SDS-PAGE, the gel is stained and photographed (Figure 4). The gel lane with the fractionated proteins is excised and cut into small pieces (or gel plugs) along its entire length. The gel plugs are then processed for protein identification by mass spectrometry. The following sections describe in detail all of these techniques. We also provide an overview of the background binding in experiments using nuclear extracts and the specific enrichment observed when purifying biotin tagged transcription factors (Figure 4). Lastly, we provide protocols for the size fractionation of nuclear extracts using a preparative Superose 6 column and for the cleavage of proteins bound to streptavidin beads using TEV protease (Figure 3). Both approaches are presented with the aim of reducing the background in protein purification by biotinylation tagging. Given the increase in the potential applications of biotinylation tagging for protein purification, the description of such protocols may prove a useful resource.

Materials

2.1 Cell culture

1. Dulbecco's Modified Eagle's Medium (DMEM) (Cambrex Bio Science, Belgium), supplemented with 10 % Fetal Bovine Serum (FBS, Hyclone, Belgium).
2. Penicillin (used at 100u/ml final concentration) and streptomycin (used at 100μg/ml final concentration), stored at –20°C (100x stock from Cambrex Bio Science, Belgium).
3. 100% Dimethylsulfoxide (DMSO, Merck, Germany), used at 2% final concentration.
4. Neomycin (Gibco-BRL, UK), stock prepared as 100mg/ml in phosphate buffered saline (PBS, see below), filter-sterilized, aliquoted, stored at –20°C and used at 400μg/ml final concentration.
5. Puromycin (Sigma, St. Louis, MO), 1000x stock aliquoted and stored as above and added to 1μg/ml final concentration.

2.2 Nuclear extract preparation

1. Dulbecco's Phosphate Buffered Saline (PBS, Cambrex Bio Science, Belgium)
2. Protease inhibitors: Complete (Roche, Germany). Use 1 tablet for 50ml of solution
3. Cell resuspension buffer: 2.2 M sucrose in 10 mM HEPES-KOH, pH 7.9, 25 mM KCl, 0.15 mM Spermine, 0.5 mM Spermidine, 1 mM EDTA (with protease inhibitors added as above).
4. Standard household blender with rotating blades for homogenizing cells.
5. Nuclear lysis buffer: 10 mM HEPES, pH 7.9, 100 mM KCl, 3 mM $MgCl_2$, 0.1 mM EDTA, 20% glycerol (with protease inhibitors added as above).
6. Coomassie Plus Assay reagent (Pierce, IL).
7. Protein standards: dilutions of 0, 100, 200, 300, 400 and 500 μg/ml of chicken egg albumin (Sigma-Aldrich, MO) in ddH_2O prepared from a 20 mg/ml stock solution. 10 μl of each standard diluted in 1 ml (final volume) of ddH_2O is used for obtaining a standard curve.
8. Spectrophotometer: Ultraspec II (LKB Biochrom), cuvettes (Sterna, Germany).
9. Conductivity meter: Philips PW 9526.
10. Standards for determining salt concentration: 100, 200 300, 400 mM KCl diluted in ddH_2O from a 1M KCl stock solution. 10 μl of each standard are diluted in 1 ml (final volume) of ddH_2O for obtaining a standard curve.

2.3 Size fractionation by Superose 6 gel filtration

1. Superose 6 analytical grade column: HR 10/30 with a total bed volume of 24 ml. Preparative grade column: XK 50/600 with a bed height of 30.3 cm, total bed volume of 589 mL, both purchased form Amersham Biosciences (UK). Range of separation of 5,000 to 5,000,000 Da. 20% ethanol is used as preservative. The column is connected to an AKTA FLPC system (Amersham Biosciences).
2. Gel filtration calibration kit: dextran blue and high molecular weight standards (Amersham Biosciences, UK).
3. Column running buffer: 20mM HEPES, 0.5mM EGTA, 1mM $MgCl_2$, 200mM KCl, 10% glycerol. All buffers used for gel filtration column should be filtered prior to use.
4. 100% Trichloroacetic acid (TCA, Sigma-Aldrich, MO).

2.4 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

1. Pre-cast NuPAGE 4-12% Bis-Tris gel (Invitrogen, UK).
2. Gel electrophoresis buffer: 1 X MOPS buffer diluted from 20 X stock solution and NuPAGE antioxidant (both from Invitrogen, UK).
3. Sample loading buffer, final concentration: 62.5 mM Tris-HCl pH 6.8, 25% glycerol (v/v from 100% stock, Sigma-Aldrich, St. Louis, MO), 2% SDS (v/v from 20% stock in ddH_2O), 0.01% bromophenol bl (w/v, Sigma-Aldrich, St. Louis, MO), 5 % β-mercaptoethanol (v/v, 100% stock, Merck, Germany). Can be prepared as a 4X stock solution.
4. Broad range pre-stained SDS-PAGE molecular weight standards (BioRad, Hercules, CA).
5. SimplyBlue Safestain gel staining solution (Invitrogen, UK).

2.5 Western blotting

1. ProTran nitrocellulose membrane (Schleicher & Schuell, Germany).
2. Gel-blotting paper (Schleicher & Schuell, Germany).

3. Blotting buffer: 25 mM Tris (made directly from the solid), 192 mM Glycine (made directly from the solid), 20% methanol.
4. Tris-Buffered Saline (TBS): 10 mM Tris-HCl, pH 7.4, 150 mM NaCl. Can be prepared as a 10X stock and stored at room temperature.
5. Blocking buffer: 5% Bovine Serum Albumin (BSA, Roche, Germany) in 1X TBS, prepared fresh.
6. Washing buffer 1x TBS adjusted to 0.5M NaCl (using a 5M NaCl stock solution), 0.3% Triton X-100 (Sigma-Aldrich, MO).
7. Primary and secondary antibody dilution: in blocking buffer with 0.2% NP-40.
8. Secondary antibodies: anti-rabbit (1/50000 dilution), anti-mouse (1/15000 dilution) from Amersham Biosciences (UK), anti-rat (1/3000 dilution) and anti goat (1/4000 dilution) from DakoCytomation (Denmark). All antibodies are purchased as horseradish peroxidase (HRP) conjugates.
9. Enhanced Chemiluminescence (ECL) kit (Amersham Biosciences, UK).
10. Bio-Max MR film (Kodak, Rochester, NY).

2.6 Streptavidin binding

1. Streptavidin paramagnetic beads (Dynabeads M280, Dynal, Sweden).
2. Chicken Egg Albumin (Sigma-Aldrich, MO).
3. Binding buffer: 1X TBS, 0.3% 0.3% NP-40 (Nonidet-40, Sigma-Aldrich, MO).
4. HENG buffer: 10 mM HEPES-KOH, pH 9, 1.5 mM $MgCl_2$, 0.25 mM EDTA, 20 % glycerol, 1mM PMSF (phenyl methyl sulfonyl fluoride, prepared as a 100x stock in ethanol and stored at –20°C).
5. Wash buffer: HENG buffer with 250 mM KCl and 0.3% NP-40.
6. Elution buffer: 1X sample loading buffer.
7. Magnets: Dynal MPC-1 and MPC-S magnets, for large and small volumes respectively (Dynal, Sweden).

2.7 TEV protease cleavage

1. Nuclear extract dilution buffer: 20mM Tris-HCl pH7.5, 0.45% NP-40.
2. Tobacco Etch Virus (TEV) protease (AcTEV Protease, Invitrogen, Scotland).

2.8 Sample preparation for mass spectrometry

1. Trypsin, sequencing grade (Roche, Germany). 10x stock made by dissolving lyophilized powder in 1mM HCl to 100ng/μl final concentration.
2. 50mM Ammonium bicarbonate (Sigma-Aldrich, MO). Dissolved in ddH_2O and filter sterilized.
3. 100% Acetonitrile (Sigma-Aldrich, MO).
4. 100% Formic acid (Sigma-Aldrich, MO).
5. Gel slice destaining solution: 25mM ammonium bicarbonate in 50% acetonitrile prepared by mixing equal volumes of the stock solutions given above.
6. 50mM iodoacetamide (Sigma-Aldrich, MO) prepared in 50mM ammonium bicarbonate.
7. 6.5mM DTT (dithiothreitol, (Sigma-Aldrich, MO) prepared in 50mM ammonium bicarbonate.

3. Methods

3.1 Cell culture

1. Mouse Erythroleukemia (MEL) cells are Friend virus transformed erythroid progenitors arrested at the proerythroblast stage of differentiation (7). MEL cells are cultured in DMEM supplemented with 10% FBS at 37°C and 5% CO_2. Cells are grown to a maximum of $2x10^6$ cells/ml and routinely split to a density of $5x10^4$ cells/ml. Cells appear to be semi-adherent and rounded with a smooth surface (*see* **Note 1**).
2. For induction, cells are diluted to $5x10^5$ cells/ml and cultured in DMEM with 2% DMSO (v/v) for at least 3 days. Cells become smaller but remain round and when pelleted appear pink/red due to hemoglobinization.

3.2 Nuclear extract preparation

1. Cells are harvested in 1 liter centrifuge bottles by centrifugation at 640 x *g* for 40 min at 4°C in a Beckman J4 centrifuge and washed once with 100 ml of ice-cold PBS. Resuspended cells are transferred into 50ml Falcon tubes and re-pelleted in an Eppendorf 5810R benchtop centrifuge at 2540 x *g* for 10 min at 4°C. The supernatant is discarded.
2. The cell pellet is gently resuspended by pipetting up and down in 200 ml of cell resuspension buffer with protease inhibitors added. Cells are equilibrated to the new osmotic conditions for 20 min on ice.
3. Cells are lysed in a blender using a single 30 second pulse at setting 3 (*see* **Note 2**). Excessive foaming should be avoided.
4. Lysis efficiency is checked under the microscope by staining a 10 μl aliquot with an equal volume of Unna stain (Methylgreen-Pyronin). Nuclei appear blue whereas intact cells appear with a blue nucleus surrounded by a non-stained cytoplasm. Optimal lysis should result in more than 90% of the nuclei appearing free of cytoplasm.
5. Nuclei are pelleted by ultracentrifugation at 141,000 x *g* using the SW28 rotor for 2 h at 4°C. A clean white pellet corresponding to the nuclei should be visible at the bottom of the tube. The top layer (cellular debris/cytoplasm) is discarded.
6. Nuclei are resuspended in 15 ml of nuclear lysis buffer and proteins are extracted by the drop wise addition of a 3.3M KCl solution with gentle agitation on ice, until the final concentration is ~350-400mM (*see* **Note 3**). Nuclear lysis and protein extraction are allowed to proceed by incubating on ice for 20 min. Two phases should be visible: one is clear and represents the soluble nuclear extract fraction whereas the other phase appears viscous and represents the insoluble fraction of mostly chromatin fragments.
7. Insoluble material is removed by ultracentrifugation at 300,000 x *g* using the SW50.1 rotor for 1 h at 4°C. The supernatant (soluble nuclear extract, approximately 17-18ml) is aliquoted in 1-5ml aliquots.
8. A small aliquot is used to measure protein concentration using the Bradford method. First, a standard curve is obtained by adding 200μl of Bradford reagent to 800 μl of each of the chicken egg albumin standards, incubating for 5 min at room temperature and measuring the absorbance at 595nm in a spectrophotometer. The absorbance of dilutions of the aliquot of the nuclear extract are measured in the same way. The concentration of the nuclear extract is determined using the standard curve, taking the dilution of the sample into account.
9. The salt concentration of the solution is determined by measuring the conductance of the sample with a conductivity meter. First, a calibration curve of conductance is obtained using the KCl dilutions as standards. The conductivity of a diluted aliquot of the nuclear extract is then measured and its approximate salt concentration is estimated using the calibration curve.
10. Nuclear extracts are snap frozen in liquid nitrogen and stored at –70°C (*see* **Note 4**).

3.3 Analytical Superose 6 gel filtration (*see* **Note 5**)

1. An aliquot of nuclear extract is thawed on ice and centrifuged for 5 min at full speed using a microcentrifuge at 4°C. The volume of extract loaded on the column should not exceed 1% of the column volume.
2. The pump and the column are equilibrated

with column running buffer. The running program is set up as follows: 100 μl/min flow rate; sample volume loop 200μl; fraction volume 500 μl; elution length 1 column volume of running buffer; alarm pressure set at 0.5 MPa (*see* **Notes 6 and 7**). After each run the column is washed with 2 column volumes of running buffer.

3. The calibration standards are run through the column to establish the elution volume of protein complexes and of free protein monomers (*see* **Notes 8-10**).
4. Collected fractions are concentrated by trichloroacetic acid precipitation (TCA), as follows: 125 μl of cold 100 % TCA are added to each 500 μl fraction, mixed well and incubated on ice for 30 min.
5. Proteins are pelleted by centrifugation at full speed in a microcentrifuge at 4^0C for 20 min.
6. The supernatants are discarded and the pellets washed with at least 500μl 1% ice cold TCA (in ddH_2O) and re-centrifuged as above.
7. The supernatants are discarded again and the pellets are washed with ice-cold acetone and re-centrifuged as above.
8. The pellets are air dried (on ice) and re-suspended in 50 μl of sample loading buffer (*see* **Note 11**). The fractionation patterns of specific proteins are determined by SDS-PAGE and Western blotting, as described below. An example of a GATA-1 Superose 6 fractionation profile of nuclear extracts from MEL cells expressing biotin-tagged GATA-1, is shown in Figure 2. From this, it is clear that both endogenous and tagged GATA-1 elute in high molecular weight (>670kDa) fractions. The fractionation profile of tagged GATA-1 follows that of the endogenous GATA-1 protein.

3.4 Preparative gel filtration by Superose 6

1. Steps 1 and 3 are as above.
2. Running program: 4ml/min flow rate; sample volume loop 5ml; fraction volume 10ml; elution length 1.5 column volumes; alarm pressure set at 0.65Mpa (*see* **Note 12**).
3. Fractions are collected in 15 ml Falcon tubes and used for precipitation (see below) or for binding to streptavidin beads (section 3.6).
4. Proteins are precipitated with 20% TCA (2.5 ml of 100 % TCA are added to every 10ml fraction) and kept on ice for 1h. The tubes are centrifuged in an Eppendorf 5810R benchtop centrifuge at maximum speed (2540 x *g*) for 20 min at 4°C. The pellet is subsequently washed with ice-cold 1% TCA (in ddH_2O). At this step the pellet can be carefully resuspended and transferred into microfuge tubes. The samples are re-pelleted by spinning as above, or in a microfuge for 20 min, full speed at 4ºC. The pellets are washed with ice-cold acetone, centrifuged again as above and air dried on ice. The protein pellets are finally resuspended in 50μl of sample loading buffer and denatured by boiling before loading on an SDS-PAGE gel (section 3.8).

3.5 Binding to streptavidin beads

1. 5-10mg of nuclear extract is thawed on ice and diluted to 150mM KCl final concentration by the dropwise addition of ice-cold HENG with gentle shaking (*see* **Note 13**). NP-40 is adjusted to a 0.3% final concentration.
2. We routinely use 200 μl of resuspended streptavidin beads per 5mg of nuclear extract. The beads are blocked with 200ng/ml chicken egg albumin (CEA) in a 1ml final volume (made up with HENG buffer), for 1 hour at room temperature on a rotating platform.
3. The beads are immobilized using a magnetic rack and the blocking solution is removed. The beads are then resuspended in the diluted nuclear extract and incubated on a rotating platform 4°C for 2 hours to overnight.
4. The beads are immobilized on ice using the magnetic rack. The supernatant, corresponding to the unbound fraction or flowthrough, is collected and saved.
5. The beads are washed in 1 ml of washing buffer as follows: 2 quick rinses followed by 3 washes, 10 min each at room temperature on a rotating platform.
6. After the last wash, the beads are resuspended in 50 μl of sample buffer. Bound proteins are eluted by boiling the beads for 5 min.
7. The eluted material is fractionated by SDS-PAGE and processed for analysis by mass spectrometry as described below.

3.6 Binding of preparative Superose 6 fractions to streptavidin beads

1. After determining the fractionation profile of the protein(s) of interest by SDS-PAGE and Western blotting (as below), the peak fractions are pooled in a suitable sterile container (we conveniently use a sterilized glass measuring cylinder). All work is carried out on ice or in the cold room.
2. KCl and NP-40 concentrations are adjusted to 150mM and 0.3%, respectively (as above) with gentle mixing.
3. The diluted fractions are divided equally into separate 50ml Falcon tubes, so that tubes are not more than ¾ full, and resuspended streptavidin M280 beads (equilibrated and blocked as above) are added followed by overnight incubation at 4°C on a rotating platform. We use approximately 10µl of streptavidin M280 beads for every 10ml fraction.
4. The beads are immobilized using the Dynal MPC-1 magnet and the supernatant corresponding to the unbound fraction is removed and saved. The immobilized beads from each tube are resuspended in washing buffer and pooled by transferring to a microfuge tube.
5. The beads are washed 4-5 times at room temperature in 1ml washing buffer for 5-10 min each.
6. The bound proteins are eluted by boiling the beads in sample buffer (50µl of sample buffer per 20µl of immobilized beads). The eluted material is fractionated by SDS-PAGE and processed for analysis by mass spectrometry as described below.

3.7 TEV protease cleavage

With the aim of reducing the non-specific background observed in streptavidin binding experiments, we developed a modified version of the biotin tag for the N-terminal tagging of fusion proteins. This tag consists of a shorter (14aa) amino acid sequence than the one previously used but which is also very efficiently biotinylated by the BirA biotin ligase (3). The biotin tag sequence is followed by a 7aa cleavage site for the highly specific TEV-protease (Tobacco Etch Virus protease) (Figure 3A). In this way, the biotinylated protein and associated complexes can be specifically released from the streptavidin beads by cleaving off with the TEV protease (Figure 3B and C).

1. Immediately after the washing steps (step 5, section 3.5) the streptavidin paramagnetic beads are resuspended in 1xTBS/0.3% NP-40 buffer (95µl of buffer for every 50µl of resuspended beads used in the outset of the experiment).
2. 5-10% (v/v) of TEV-protease is added to the resuspended beads followed by incubation for 1-3h at 16°C with shaking (*see* **Note 14**).
3. The supernatant can be collected (it should contain the cleaved protein) and, if necessary, can be concentrated by TCA precipitation, as above.
4. The efficiency of protease cleavage can be monitored by testing an aliquot of the supernatant by SDS-PAGE and Western blotting (Figure 3B and C). The material that remains bound to the beads after TEV protease cleavage can be eluted by boiling in sample buffer and tested by SDS-PAGE. Successful cleavage results in the loss of the biotin-tag (as visualized by Streptavidin-HRP, Figure 3C) accompanied by a downshift in the size of the tagged protein thus resulting in faster migration by SDS-PAGE (as visualized by the tagged protein-specific antibody, Figure 3B).

3.8 SDS-PAGE

We preferably use NuPAGE pre-cast gels since they give clear and reproducible results in terms of resolution and sharpness of the protein bands. This if of particular importance if the gel is to be processed for mass spectrometry. There is also the added advantage of using gradient precast gels (e.g. 4-12%) for resolving proteins in a wide range of molecular weights. We use the Invitrogen electrophoresis system for running the NuPAGE gels.

1. Pre-cast gels are removed from the plastic envelope and rinsed in ddH_2O. The sticker near the bottom of the gel and the comb are removed carefully.
2. 800 ml of 1X MOPS electrophoresis buffer are prepared.
3. The gel is placed in the electrophoresis system (Invitrogen). The outside chamber is filled with 600 ml of 1X MOPS.
4. 500 µl of antioxidant are added to the remaining 200 ml of the 1X MOPS buffer,

mixed and used to fill the inner chamber.
5. The wells of the gel are rinsed twice with the running buffer.
6. The samples in sample loading buffer are boiled for 5 min to denature the proteins and loaded directly onto the gel. 5-10 μg of nuclear extract are loaded per lane.
7. The gel is electrophoresed at 200V constant voltage for 60-75 min.

3.9 Western blotting

Samples that have been separated by SDS-PAGE are electrophoretically transferred onto nitrocellulose membrane by "wet blotting". For protein transfer we use the Trans-Blot electrophoretic transfer cell (Bio-Rad, CA).

1. Four gel-blotting papers are cut to the size of the gel. A "sandwich" is set up consisting of a sponge, 2 pieces of blotting paper, the gel, the membrane, another 2 pieces of blotting paper and a sponge (*see* **Note 15**). The membrane is pre-wetted in water followed by transfer buffer.
2. The "sandwich" is placed in the transfer tank containing transfer buffer pre-chilled at 4^0C, such that the membrane is between the gel and the anode.
3. Blotting is carried out under constant amperage at 390 mA for 70 min in the cold room (*see* **Note 16**).
4. At the end of the transfer, the "sandwich" is disassembled and the membrane is rinsed in 1x TBS / 0.05% NP-40 (*see* **Note 17**).
5. The membrane is blocked at room temperature in freshly prepared blocking buffer for 1 hour on a rocking platform.
6. The blocking buffer is discarded, replaced by the primary antibody (in this case anti-GATA-1 N6 antibody diluted 1:5000) in blocking buffer / 0.2 % NP-40 and incubated overnight at 4°C on a rotating wheel.
7. The primary antibody is removed and the membrane washed three times with 50 ml of washing buffer, 15 min each wash at room temperature (*see* **Note 18**).
8. A freshly prepared secondary antibody (in this case anti-rat diluted 1:3000 in blocking buffer) is added to the membrane and incubated for one hour at room temperature on a rocking platform.
9. The secondary antibody is discarded and the membrane is washed as in step 8, followed by one wash for 5 min in 1xTBS/0.05% NP-40.
10. The membrane is lifted out of the wash buffer and excess liquid is removed by touching it on a clean tissue and placed in a clean tray such as a plastic weigh boat.
11. 3 ml of ECL solution (per filter) is prepared according to the manufacturer's instructions and immediately added to the membrane and shaken gently for 1 min to ensure an even coverage of the membrane by the liquid.
12. Excess liquid is again removed by touching the membrane on a clean tissue. The membrane is wrapped in cling wrap and exposed to film in an autoradiography cassette in a dark room, *as soon as possible*.

3.10 Preparation of samples for mass spectrometry (*see* **Note 19**)

The procedures described below are for the analysis by liquid chromatography-tandem mass spectrometry (LC-MS/MS) using a Q-Tof Ultima API mass spectrometer. The treatment of samples may vary depending on the type of analysis and the instrument used. It is best to consult with the mass spectrometry facility where the analysis is to be carried out for the processing of samples.

1. Following electrophoresis by SDS-PAGE the gel is stained overnight with Colloidal Blue, according to the manufacturer's instructions.
2. The gel is destained in several changes of ddH_2O until the background (i.e. the non-protein containing part of the gel) is completely destained. This usually takes several hours (i.e. more than 12 hours).
3. The destained gel is photographed to provide a record of the purification experiment.
4. 20-25 microfuge tubes are rinsed in 60% acetonitrile.
5. The entire lane is cut out lengthwise and divided into at least 20 gel slices. Each gel slice is placed in a separate tube.
6. Each gel slice is destained in 100μl of destaining solution (25mM ammonium bicarbonate in 50% acetonitrile) for 20-30min. This step is repeated until the gel slice becomes completely destained (usually 3-4 times). Alternatively, gel slices can be destained overnight at 4^oC.
7. Each gel slice is dehydrated in 100μl of

100% acetonitrile for 5-10min at room temperature. The plug become hard and white at this step.
8. The gel slices are reduced with freshly prepared 6.5mM DTT solution for 45-60min at 37°C.
9. The solution is discarded and proteins in the gel slices are alkylated by adding 100µl of 54 mM iodoacetamide solution and incubating for 60min at room temperature in the dark.
10. The solution is discarded and the gel slices are washed in 100µl of gel slice destaining solution for 15 min at room temperature. This step is repeated once more.
11. The washing solution is discarded and the gel slices are dried in 100µl of 100% acetonitrile for 10min. The solution is again discarded and the gel slices are dried at room temperature.
12. Proteins are in-gel digested in 15µl of 10 ng/µl modified trypsin at (diluted from the 100x stock in 50 mM ammonium bicarbonate) for 30 min on ice (*see* **Note 20**). 15 µl of 50 mM ammonium bicarbonate are added to the samples followed by overnight incubation at 37°C.
13. Samples are equilibrated to room temperature. 30µl of 2% acetonitrile in 0.1% formic acid are added to the samples and incubated at room temperature for 15 min. The samples are then vortexed briefly and sonicated for 1 minute.
14. The supernatants are collected in separate tubes and the remaining gel slices are treated with 30µl of 50% acetonitrile in 0.1% formic acid and incubated as above. Samples are again vortexed and sonicated as above and the supernatants are collected and pooled with the corresponding supernatants from step 14.
15. The samples are vacuum dried in a vacuum centrifuge for 45-60 minutes until they are dry.
16. The eluted peptides are now ready for analysis by mass spectrometry.

Figure 4A shows an example of a preparative binding of nuclear extracts from MEL cells expressing biotinylated GATA-1 (lane 3) and control extracts from cells expressing the BirA biotin ligase only (lane 5). The control binding experiment shows that background consists of a few strongly stained bands against a backdrop of more faintly staining bands. Analysis by mass spectrometry identified the strongly staining bands as corresponding to naturally biotinylated proteins such as carboxylases (1). The bulk of the remaining background proteins corresponded to abundant nuclear proteins such as splicing factors, proteins involved in ribosome biogenesis etc. (Figure 4B). The low background binding (<1% of the total) of transcription factors and chromatin remodeling and modification proteins is of note (Figure 4B). By contrast, analysis of the GATA-1 purification gel slice shows a large increase in the binding of transcription factors and chromatin remodeling and modification proteins, thus indicating specific co-purification with biotin-tagged GATA-1 (Figure 4C) (6). A number of these interactions have been validated by independent immunoprecipitation experiments using nuclear extracts from non-transfected MEL cells and shown to include essential hematopoietic transcription factors such as FOG-1, TAL-1 and Gfi-1b in addition to chromatin remodeling complexes such as MeCP1 and WCRF/ACF (6). In addition, the analysis of the GATA-1 binding experiment identified abundant chromatin associated proteins, such as topoisomerases, as background due to their indirect co-purification with GATA-1 by virtue of their association with chromatin. Thus, we have defined background in these experiments as consisting primarily of naturally biotinylated proteins, abundant nuclear proteins associated with RNA metabolism and ribosome biogenesis and abundant chromatin associated proteins that are indirectly co-purified with chromatin-bound transcription factors.

4. Future prospects

We have shown that biotinylation tagging is highly efficient in cultured cells (Figure 1) and transgenic mice (1) and we have used this approach to identify a number of different complexes formed by the essential hematopoietic transcription factor GATA-1 (6). Due to its efficiency and ease of application, biotinylation tagging offers the prospect of rapidly expanding the characterization of transcription factor complexes. For example, the biotinylation tagging of the hematopoietic transcription factor partners of GATA-1 and the characterization of their protein complexes will lead to the rapid elucidation of the distinct and overlapping transcriptional networks these factors regulate in hematopoiesis. Similarly, the biotinylation tagging of chromatin co-factors will lead to a better understanding of their interactions with tissue-specific transcription factors and the molecular basis of their functions (i.e. chromatin remodeling and modification in activation and repression). Furthermore, efforts in reducing the background along the lines described here (i.e. a pre-purification steps such as gel filtration or the use of protease cleavage) will help in further expanding the utility of biotinylation tagging, for example in preserving the native properties of complexes or in determining stoichiometries. The utility of biotinylation tagging will be further increased through the development of additional tools such as the recent derivation of a transgenic mouse strain that expresses BirA ubiquitously in all tissues (8), or the construction of a codon-optimized version of BirA for the efficient expression in mammalian cells (9). The recent description of the biotinylation of cell surface proteins (10) should also serve to expand the utility of this approach. Lastly, it should be noted that in vivo biotinylation tagging can also be employed (e.g. instead of antibodies) in all other applications involving an affinity purification or detection step, such as immunofluorescence (1), immunoprecipitations (1, 11) and chromatin immnoprecipitation (ChIP) assays (1, 12).

5. Notes

1. We routinely screen 12-20 stable transfected MEL cell clones by SDS-PAGE in order to select a clone that expresses the tagged protein at no more than 50% of the expression level of the endogenous protein. This is in order to ensure that the physiological interactions and functions of the protein of interest are not disturbed as a result of the overexpression of the tagged protein.
2. The specific lysis conditions will depend on the make of blender employed. It is recommended that conditions are optimized for cell density, length of lysis time and speed setting of the blender.
3. The final salt concentration is critical for the extraction of nuclear proteins.
4. We routinely obtain around 100 mg of nuclear extract from 4 liters of MEL cell culture at a density of $2x10^6$ cells/ml.
5. There are a large variety of column matrices commercially available for gel filtration, with each matrix having different optimal separation ranges and physicochemical properties (e.g. ability to withstand high pressure in the column). Thus, the choice of matrix will depend on the desired range of fractionation and the liquid chromatography operating system available to the user (e.g. FPLC or HPLC).
6. Users must also refer to the manufacturer's instructions and training for use of the column and the FPLC apparatus.
7. The resolution efficiency of new columns, expressed as the number of theoretical plates per meter of column under normal running conditions, should be tested first. This can be done by injecting a sample of acetone (5mg/mL) in ddH_2O water. Indicative efficiency for the analytical grade column is 11100 theoretical plates/m.
8. While loading the extract, care must be taken that no air bubbles enter the loop. Air bubbles as well as cell debris can damage the column bed.
9. Once a new column is installed, the void (V_0) volume is determined by the peak of elution of dextran blue. In order to further calibrate the column, a mixture of at least two proteins of known molecular weight should also be injected. Recommended standards: bovine serum albumin

(67kDa), thyroglobulin (669kDa), aldolase (158kDa).

10. If there is any suspicion that the column bed has been damaged, it is best to run the calibration standards again.
11. If the blue color of the sample loading buffer turns yellow, it is due to the protein sample being acidic which will also affect migration of the sample during SDS-PAGE. A few microliters of Tris-HCl pH9.0 are usually sufficient to neutralize the sample.
12. To avoid pressure build up the run can be started at a flow rate of 1ml/min. It is also better to inject the sample with the lower flow rate.
13. The concentration of 150mM KCl is critical for the efficient binding of biotinylated proteins to streptavidin beads. We have found that even modest increases in salt concentration severely affect binding efficiency.
14. Protease cleavage also works well with shorter incubation times (5-30min) and a broader temperature range (4-37^{o}C).
15. Avoid handling membrane directly, use gloves and forceps.
16. Under these transfer conditions, the temperature of the buffer can rise significantly and frothing may occur. This does not affect the transfer.
17. The gel can be stained after blotting in order to visualize residual proteins as a test for the efficiency of transfer as well as an indication of the amount of protein loaded per lane.
18. The primary antibody can be stored and re-used. Sodium azide is added to the antibody solution to a 0.02% final concentration and stored at 4^{0}C (sodium azide stock: 10% w/v in ddH_2O. **Caution: sodium azide is highly toxic**.
19. In order to reduce the risk of contaminating the samples for mass spectrometry, particularly with keratins, work is carried out in a hood wearing double gloves, a lab coat and always using sterile plasticware.
20. The volume of trypsin solution added will depend on the size of the gel slice. The volumes given above are for appr. 4x2mm gel slices. At this stage, gel slices should swell and little solution should remain visible.

References

1. de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F. and Strouboulis, J. (2003) Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc Natl Acad Sci U S A,* **100,** 7480-5.
2. Schatz, P.J. (1993) Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in Escherichia coli. *Biotechnology (N Y),* **11,** 1138-43.
3. Beckett, D., Kovaleva, E. and Schatz, P.J. (1999) A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci,* **8,** 921-9.
4. Smith, P.A., Tripp, B.C., DiBlasio-Smith, E.A., Lu, Z., LaVallie, E.R. and McCoy, J.M. (1998) A plasmid expression system for quantitative in vivo biotinylation of thioredoxin fusion proteins in Escherichia coli. *Nucleic Acids Res,* **26,** 1414-20.
5. Cull, M.G. and Schatz, P.J. (2000) Biotinylation of proteins in vivo and in vitro using small peptide tags. *Methods Enzymol,* **326,** 430-40.
6. Rodriguez, P., Bonte, E., Krijgsveld, J., Kolodziej, K.E., Guyot, B., Heck, A.J.R., Vyas, P., de Boer, E., Grosveld, F., Strouboulis, J. (2005) GATA-1 forms distinct activating and repressive complexes in erythroid cells. *EMBO J,* **in press**.
7. Singer, D., Cooper, M., Maniatis, G.M., Marks, P.A. and Rifkind, R.A. (1974) Erythropoietic differentiation in colonies of cells transformed by Friend virus. *Proc Natl Acad Sci U S A,* **71,** 2668-70.
8. Driegen, S., Ferreira, R., van Zon, A., Strouboulis, J., Jaegle, M., Grosveld, F., Philipsen, S., Meijer, D. (2005) A generic tool for biotinylation of tagged proteins in transgenic mice. *Transgenic Research,* **in press**.
9. Mechold, U., Gilbert, C. and Ogryzko, V. (2005) Codon optimization of the BirA enzyme gene leads to higher expression and an improved efficiency of biotinylation of target proteins in mammalian cells. *J Biotechnol,* **116,** 245-9.
10. Chen, I., Howarth, M., Lin, W. and Ting, A.Y. (2005) Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat Methods,* **2,** 99-104.
11. Koutsodontis, G. and Kardassis, D. (2004) Inhibition of p53-mediated transcriptional responses by mithramycin A. *Oncogene,* **23,** 9190-200.

12. Viens, A., Mechold, U., Lehrmann, H., Harel-Bellan, A. and Ogryzko, V. (2004) Use of protein biotinylation in vivo for chromatin immunoprecipitation. *Anal Biochem,* **325,** 68-76.thyroglobulin (669kDa), aldolase (158kDa).
10. If there is any suspicion that the column bed has been damaged, it is best to run the calibration standards again.
11. If the blue color of the sample loading buffer turns yellow, it is due to the protein sample being acidic which will also affect migration of the sample during SDS-PAGE. A few microliters of Tris-HCl pH9.0 are usually sufficient to neutralize the sample.
12. To avoid pressure build up the run can be started at a flow rate of 1ml/min. It is also better to inject the sample with the lower flow rate.
13. The concentration of 150mM KCl is critical for the efficient binding of biotinylated proteins to streptavidin beads. We have found that even modest increases in salt concentration severely affect binding efficiency.
14. Protease cleavage also works well with shorter incubation times (5-30min) and a broader temperature range (4-37^{o}C).
15. Avoid handling membrane directly, use gloves and forceps.
16. Under these transfer conditions, the temperature of the buffer can rise significantly and frothing may occur. This does not affect the transfer.
17. The gel can be stained after blotting in order to visualize residual proteins as a test for the efficiency of transfer as well as an indication of the amount of protein loaded per lane.
18. The primary antibody can be stored and re-used. Sodium azide is added to the antibody solution to a 0.02% final concentration and stored at 4^{0}C (sodium azide stock: 10% w/v in ddH_2O. **Caution: sodium azide is highly toxic**.
19. In order to reduce the risk of contaminating the samples for mass spectrometry, particularly with keratins, work is carried out in a hood wearing double gloves, a lab coat and always using sterile plasticware.
20. The volume of trypsin solution added will depend on the size of the gel slice. The volumes given above are for appr. 4x2mm gel slices. At this stage, gel slices should swell and little solution should remain visible.

Acknowledgements

We are indebted to Dr. Jeroen Krijgsveld (Utrecht University) and Dr. Jeroen Demmers (Erasmus Medical Center) for expert mass spectrometry analysis. Work in our laboratory has been supported by grants from the Dutch Research Organization (NWO), the European Union (grant HRPN-CT-2000-00078), the NIH (grant RO1 HL 073445-01) and the Netherlands Proteomic Center.

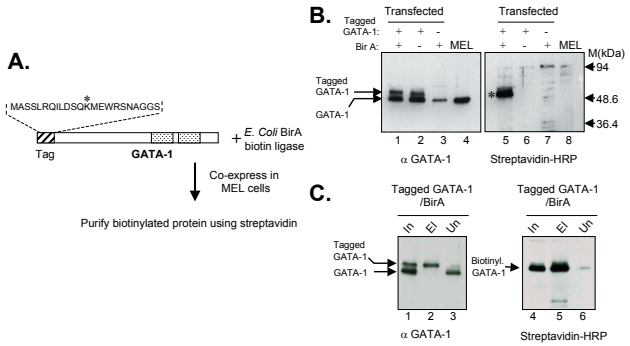


Figure 1: (A) Scheme for the specific biotinylation of tagged GATA-1 by BirA biotin ligase in mouse erythroleukemic (MEL) cells.

The sequence of the 23aa peptide tag fused to the N-terminus of GATA-1 is shown. The asterisk indicates the lysine residue that becomes specifically biotinylated by BirA. Speckled boxes indicate the positions of the two GATA-1 Zinc-fingers. (B) Biotinylation of tagged GATA-1 in MEL cells. Left panel: Western blot with an anti-GATA-1antibody to detect endogenous and tagged GATA-1 proteins. Right panel: Western blot of the same extracts with streptavidin-HRP conjugate to detect biotinylated GATA-1. Biotinylated GATA-1 (asterisk) is clearly visible in the right panel only in the lane of the double transfected cells. (C) Efficiency of GATA-1 biotinylation and binding to streptavidin beads. Left panel: Western blot using anti-GATA-1 antibody to detect binding of tagged GATA-1 to streptavidin beads Input and unbound material are shown in lanes 1 and 3. Right panel: the same filter stripped and re-probed with streptavidin-HRP to detect the binding of biotinylated GATA-1 to streptavidin beads (lane 5). Lane 6 shows that very little tagged GATA-1 remains unbound by streptavidin. In: Input (nuclear extract); El: Eluted material; Un: Unbound material. Reproduced with permission from (1). Copyright (2003) National Academy of Sciences, USA.

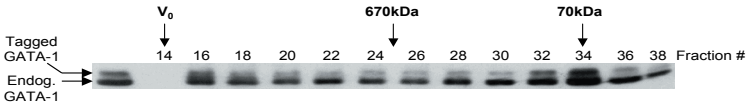


Figure 2: Superose 6 fractionation profiles of nuclear extracts from MEL cells expressing biotin-tagged GATA-1 detected by Western blotting.

The tagged GATA-1 protein is migrating with a slower mobility to that of the endogenous GATA-1 due to the extra tag sequences fused to it. The elution of the molecular weight markers is indicated at the top. V_0: void volume.

A.

GGCCGCCATC ATG GCT GGT GGC CTG AAT GAC ATC TTT GAG GCC CAG AAG ATC GAG TGG CAT GAG AAC CTG TAC TTC CAG GGA GCC
MET Ala Gly **Gly Leu Asn Asp Ile Phe Glu Ala Gln Lys Ile Glu Trp** His **Glu Asn Leu Tyr Phe Gln Gly** Ala
<---------->
Kozak sequence **Biotinylation tag** **TEV protease cleavage site**

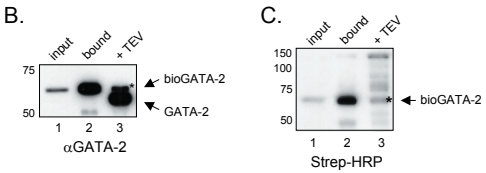


Figure 3: TEV protease cleavage of bioGATA-2 bound to streptavidin beads.
Panel (A): sequence and translation of the shorter biotinylation tag and the TEV protease cleavage site. Panels (B) and (C): Lane 1: nuclear extract from MEL cells expressing tagged GATA-2 (there is no endogenous GATA-2 expressed in MEL cells). Lane 2: biotinylated GATA-2 (bioGATA-2) bound to the beads. Lane 3: TEV protease cleavage of bound GATA-2. (A) Detection with anti-GATA-2 antibody. (B) Detection of the blot shown in panel A with streptavidin-HRP. TEV-cleavage of bioGATA-2 results in a downshift in size of the protein and loss of the biotin-tag (compare lanes 2 and 3 in panels A and B, respectively). Remaining uncleaved bioGATA-2 is indicated with an asterisk in panel B.

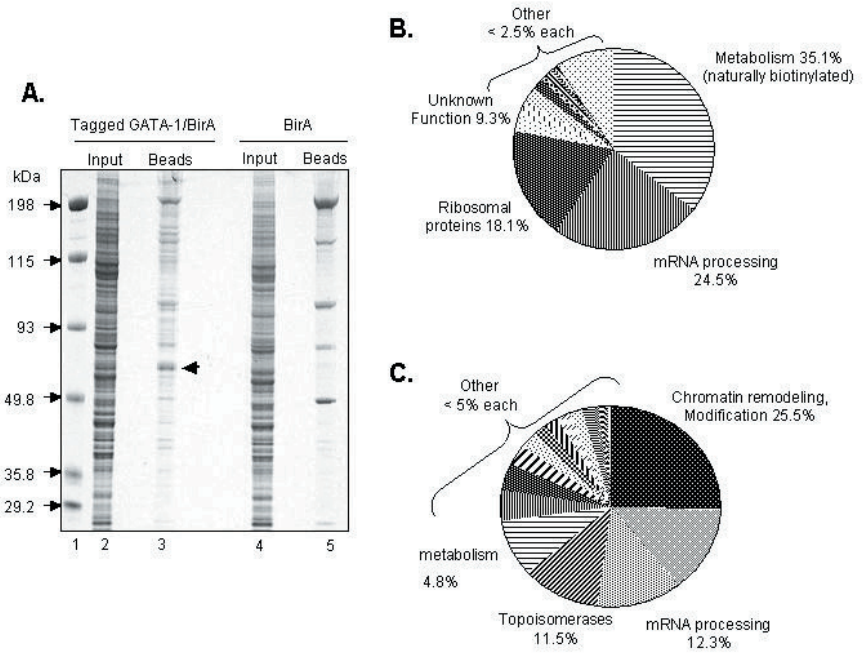


Figure 4: (A) Colloidal Blue-stained gel of a binding experiment of crude nuclear extracts to streptavidin beads.

Lane 1: Marker (M). Lane 2: input nuclear extract from tagged GATA-1/BirA double transfected cells. Lane 3: proteins eluted after binding to streptavidin beads. Lane 4: input nuclear extract from BirA transfected cells. Lane 5: proteins eluted after binding to streptavidin beads. Arrow in lane 3 indicates protein band containing biotinylated GATA-1, as determined by mass spectrometry. (B) Classification according to Gene Ontology criteria of proteins identified by mass spectrometry from the control experiment using extracts from cells expressing BirA (around 500 peptide sequences were identified). This represents the background binding. (C) Classification according to Gene Ontology criteria of proteins identified by mass spectrometry using extracts from cells expressing biotin tagged GATA-1 (more than 1000 peptide sequences were identified). A significant increase in the identification of chromatin remodelling proteins and transcription factors is clearly notable compared to the control experiment shown in (B). Fig. 4A reproduced with permission from (1). Copyright (2003) National Academy of Sciences, USA.

Summary

Samenvatting

We have adapted an *in vivo* biotinylation tagging method in mammalian cells in order to characterize interacting partners of the essential transcription factors GATA-1 and Ldb-1, with the aim of understanding the molecular mechanisms that these proteins regulate in erythroid cells. Using GATA-1 as a paradigm, we showed that biotin-tagged proteins can be pulled-down in a single purification step by binding to streptavidin beads, and their interacting partners identified by mass spectrometry. The biotinylation tagging technique was shown to be highly specific and very efficient as almost 100% of the tagged protein was biotinylated in mouse erythroleukemic (MEL) cells. Importantly, this method was also shown to work efficiently in transgenic mice. The identification of GATA-1 and of Ldb-1/Tal-1 partners led to the characterization of large protein complexes, containing transcription factors and/or co-factors, which carry out activating or repressing functions. For both transcription factors we identified a number of previously known partners, thus validating our purification approach, as well as novel interacting partners, such as the MeCP1 and ACF/WCRF complexes and the Gfi-1b transcription factor for GATA-1, or Eto-2 and cdk9 for Ldb1. Significantly, we showed GATA-1 to be part of at least five distinct complexes in differentiated MEL cells, all identified through the single step binding to streptavidin.

More specifically, GATA-1 interacts with the repressive MeCP1 complex, the interaction being mediated by FOG-1, a well-known GATA-1 co-factor. The GATA-1/FOG-1/MeCP1 complex was shown to carry histone deacetylase activity and to repress transcriptional activity in a transfection assay. Importantly, we showed that the GATA-2 gene, which is essential for early stages of hematopoiesis, is specifically repressed by the GATA-1/FOG-1/MeCP1 complex in differentiating erythroid cells. The promoters of the eosinophilic MBP and IL-5Rα genes or of the myeloid PU.1 gene also showed enrichment for a similar GATA-1/FOG-1/MeCP1 complex binding. Interestingly, in the case of the MBP gene, we found GATA-1 binding to the same promoter in eosinophils, where it is known to act as an activator together with C/EBP proteins. Taken together, these data suggest a role for the GATA-1/FOG-1/MeCP1 complex in the repression of early hematopoietic potential or of alternative cell fates in terminal erythroid differentiation.

GATA-1 forms a second complex with (at least) Gfi-1b, which was found to bind weakly to promoters of genes involved in cell proliferation, such as c-Myc and c-Myb. Microarray expression analysis of differentiating erythroid cells expressing an inducible form of GATA-1 has shown these genes to be repressed upon induction of GATA-1 expression, thus linking GATA-1 to cell cycle regulation in erythroid differentiation(Welch et al., 2004). A third GATA-1 complex involves the known association with the TAL-1 and Ldb1 transcription factors. This complex binds stably to the enhancer of the active EKLF erythroid gene, which contains E-box and GATA motifs. Finally, we also show GATA-1 to form a fourth distinct complex with FOG-1 independently of MeCP1 and a fifth complex with the chromatin remodeling ACF/WCRF complex.

In addition, we also purified from non-induced and induced MEL cells the complex for Ldb1, which is a ubiquitously expressed partner of the hematopoietic GATA-1 and TAL-1 transcription factors. Interestingly, purification of Ldb-1 in differentiated MEL cells revealed the presence of the pentameric complex that has been previously described and which included GATA-1, TAL-1, Ldb1, E2A and Lmo2. Novel partners were also identified and included Eto-2, Cdk9 and Lmo-4. Our work also suggested that Eto-2 provides a repressive role for TAL-1 target genes, and that repression is relieved by changes in the ratio of Eto-2/Ldb1 in the complex during erythroid differentiation. Ldb1 was also shown to form complexes independently of GATA-1. These data suggest that GATA-1 and Ldb1, in addition to their common functions via the complex GATA-1/TAL-1/Ldb-1, also carry different functions independently of each other in erythropoiesis.

Our data on GATA-1 and Ldb1 document the functional importance of transcription factors, co-factors and chromatin remodeling/modifying complexes and how antagonistic active or repressive functions of a single transcription factor can take place at the same time in the cell.

We hebben een in vivo biotinylatie tagging methode geschikt gemaakt voor zoogdiercellen om partners te kunnen karakteriseren van de essentiële transcriptiefactoren GATA-1 en Ldb-1, met het doel de moleculaire mechanismen te begrijpen die deze eiwitten in erytroïde cellen reguleren.We gebruiken GATA-1 als model en laten zien dat eiwitten met een biotine tag in één stap uit de oplossing kunnen worden gevist door binding aan streptavidinbollen en dat de samenwerkende partners door middel van massaspectometrie geanalyseerd kunnen worden. De biotinylatie tagging methode blijkt zeer specifiek en efficiënt want bijna 100% van de getagde eiwitten werd gebiotinyleerd in MEL (mouse erythro leukemic) cellen . Bovendien werkt deze methode ook efficient in transgene muizen. De identificatie van de GATA-1 en Ldb-1/Tal-1 partners heeft geleid tot de karakterisering van van grote eiwitcomplexen met transcriptiefactoren en/of co-factoren, die activerend of represserend werken. Voor beide transcriptiefactoren hebben we een aantal bekende partners gevonden, hetgeen onze zuiveringsmethode valideert, alsook onbekende partners zoals de MeCP1 en ACF/WCRF complexen en de Gfi-1b trancriptiefactor voor GATA-1 , of Eto-2 en cdk9 voor Ldb-1. Het is opvallend dat we laten zien dat GATA-1 tenminste deel uitmaakt van vijf verschillende complexen , allemaal geïdentificeerd door een enkele binding aan streptavidinebollen.

Bij nadere beschouwing zien we dat GATA-1 een interactie aangaat met het repressieve MeCP-1 complex, bewerkstelligd door FOG-1, een bekende cofactor van GATA-1. Het GATA-1/FOG-1/MeCP1 complex vertoont histon deacetylase activiteit en remt transcriptie in een transfectieproef . Bovendien laten we zien dat de expressie van het GATA-2 gen , belangrijk voor de vroege stadia van hematopoiese, specifiek wordt onderdrukt door het GATA-1/FOG-1/MeCP1 complex in differentiërende erytroïde cellen. De promoters van de eosinofiele genen MBP en IL-5R $\propto$ of van het myoloïde PU.1 gen laten ook verrijking zien van een soortgelijke GATA-1/FOG-1/MeCP1 complex binding. Interessant is dat we, in het geval van het MBP gen, GATA-1 binding vinden aan dezelfde promoter in eosinofiele cellen, waar het samen met C/EBP eiwitten een activator is. Samengenomen suggereren deze data een rol voor het GATA-1/FOG-1/MeCP1 complex in de repressie van vroeg hematopoietisch potentieel of in alternatieve bestemmingen van cellen in terminale erytroïde differentiatie.

GATA-1 vormt een tweede complex met (tenminste) Gfi-1b, dat zwak bindt aan promoters van genen die betrokken zijn bij celproliferatie, zoals c-Myc en c-Myb. Microarray expressie analyse van differentiërende erytroïde cellen die een induceerbare vorm van GATA-1 tot expressie brengen, toont repressie van deze genen aan na inductie van GATA-1 expressie. Dit laat zien dat GATA-1 een functie heeft bij de regulatie van de celcyclus in erytroïde differentiatie(Welch et al., 2004). Bij een derde GATA-1 complex is er sprake van betrokkenheid met de transcriptie factoren TAL-1 en Ldb-1. Dit complex bind stabiel aan de enhancer van het actieve erytroïde gen EKLF, die een E-box en GATA motieven bevat. Tenslotte tonen we een ,ander, vierde complex samen met FOG-1 maar onafhankelijk van MeCP1 aan en een vijfde complex samen met het chromatin remodeling complex ACF/WCRF.

Tevens hebben we, uit niet geïnduceerde en geïnduceerde MEL cellen, het complex voor Ldb-1 gezuiverd . Dit is een alomtegenwoordige partner van de hematopoiëtische transcriptiefactoren GATA-1 en TAL-1. De zuivering van Ldb-1 uit gedifferentiëerde MEL cellen laat het pentamere complex zien met GATA-1, TAL-1, Ldb-1, E2A en Lmo-4. Bovendien werden nieuwe partners gevonden waaronder Eto-2, Cdk9 and Lmo-4. Ons werk suggereert dat Eto-2 een negatieve functie heeft voor genen die een target voor TAL-1 zijn, en dat repressie wordt opgeheven door veranderingen in de verhouding tussen Eto-2 en Lbd-1 in het complex tijdens de erytroïde differentiatie. We laten ook zien dat Ldb-1 complexen vormt waarbij GATA-niet betrokken is. Deze resultaten suggereren dat GATA-1 en Ldb-1, behalve hun gezamenlijke functie via het complex GATA-1/TAL-1/Ldb-1, onafhankelijk van elkaar ook andere functies hebben in erythropoiëse.

Onze gegevens over GATA-1 en Lbd-1 bevestigen het functionele belang van transcriptie factoren, co-factoren en chromatin remodeling/modifying complexen en hoe tegenstrijdige activerende of repressieve functies van een enkele transcriptiefactor gelijktijdig in de cel kunnen plaatsvinden.

Curriculum vitae

Name:	Patrick Jean-François Rodriguez
Born:	27 June 1976 - Nîmes - France

➢ 1994-1996:	B.T.S. Biochemistry - Lycée Saliège - Toulouse France
➢ 1996-1999:	Insititut Universitaire Professionalise (I.U.P.) Master degree - Bioengineering - with honours -
➢ 2000:	DEA molecular biology - Strasbourg - France

Scientific research experience

➢ 2000-2005:	PhD student in the department of Cell Biology, Erasmus MC, Prof. F. Grosveld and Dr. J. Strouboulis Characterization of transcription factor complexes in hematopoiesis
➢ 1999-2000:	Transgene/I.G.B.M.C. (S. Braun/Prof J. L. Mandel) - *Strasbourg - France* - Improvement of transgene expression in non-viral vectors applied for gene therapy
➢ 1999:	Transgene (S. Braun) - Biochemical and immunological study for clinical trial in gene therapy for Duchenne Dystrophy
➢ 1998:	Biotechnology Center - (Prof F. Wurm) - Ecole Polytechnique de Lausanne - *Switzerland* - Production of recombinant antibodies in CHO cells
➢ 1997:	SANOFI Recherche - (N. Vita) - Measurement of intracellular calcium induced by neurotensin
➢ 1995-1996:	Cellular Biology Laboratory (Prof. E. Hollande) - Paul Sabatier University - *Toulouse - France* - Characterization and study of protein expression in cystic fibrosis

List of publications

Braun S, Thioudellet C, **Rodriguez P**, Ali-Hadji D, Perraud F, Accart N, Balloul JM, Halluard C, Acres B, Cavallini B, Pavirani A. Immune rejection of human dystrophin following intramuscular injections of naked DNA in mdx mice. Gene Ther. 2000 Sep;7(17):1447-57.

de Boer E, **Rodriguez P**, Bonte E, Krijgsveld J, Katsantoni E, Heck A, Grosveld F and Strouboulis J. Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. Proc Natl Acad Sci U S A. 2003 Jun 24;100(13):7480-5.

Rodriguez P, Bonte E, Krijgsveld J, Kolodziej KE, Guyot B, Heck AJR, Vyas P, de Boer E, Grosveld F and Strouboulis J.GATA-1 forms distinct activating and repressive complexes in erythroid cells. EMBO, 2005, Jul 6; (24): 2354-66.

Rodriguez P, Braun H, Kolodziej K. E., de Boer E, Campbell J, Bonte E, Grosveld F, Philipsen S and Strouboulis J. Isolation of transcription factor complexes by in vivo biotinylation tagging and direct binding to streptavidin beads. Methods Mol Biol Chapter. In press, 2005.

Grosveld F, **Rodriguez P**, Meier N, Kpric S, Pourfarzad F, Papadopoulos P, Kolodziej K, Patrinos P, Hostert A, and Strouboulis J. Isolation and characterization of hematopoietic transcription factor complexes by in vivo biotinylation tagging and mass spectrometry. Ann NY Acad Sci. In press, 2005.

Goardon N, Lambert J.A, **Rodriguez P**, Herblot S, Thibault P, Dumenil D, Strouboulis J, Romeo PH, and Hoang T. ETO2 coordinates cellular proliferation and differentiation during erythropoiesis. EMBO, In press, 2005.

Schuh AH, Tipping AJ, Clark AJ, Hamlett I, Guyot B, Iborra FJ, **Rodriguez P**, Strouboulis J, Enver T, Vyas P, Porcher C. ETO-2 associates with SCL in erythroid cells and megakaryocytes and provides repressor functions in erythropoiesis. Mol. Cell Biol., 2005 Dec, 25(23), 10235-50.

Meier N, Krpic S, **Rodriguez P**, Strouboulis J, Monti M, Krijgsveld J, Gering M, Patient R, Hostert A, Grosveld F. Novel binding partners of LDB1 are required for hematopoietic development. Submitted.

Acknowledgments

I would like to thank Prof. Frank Grosveld for giving me the opportunity to study for my PhD thesis in the Department of Cell Biology in Rotterdam. You have been very open, supportive and encouraging over the past 5 years. I appreciate the critical discussions we have had and the way you guided me to deepen my scientific understanding. Next, I wish to express my thanks to John Strouboulis, my co-promoter, for being very motivating and positive throughout my PhD. I learnt an interactive and dynamic way of carrying out my research, even during the frustrating periods. Thank you for the trust you had and what you taught me. Socially, I enjoyed your repertoire of restaurants in Rotterdam!

I acknowledge the members of the committee, Sjaak Philipsen, Elaine Dzierzak, Peter Verrijzer, Dies Meijer, Paul-Henri Romeo, and Marieke von Lindern for reading my thesis. Merci also to Paul-Henri Romeo with whom we had a fruitful collaboration, and I would appreciate to have a chance for more collaborations in the future. To Marieke von Lindern for enthusiastic discussions and your availability for scientific advice and expertise.

Most of the work presented in this thesis would not have been possible without Mass Spec technology. The collaboration with Jeroen Krijgsveld has been very pleasant and effective. His input has been very important in obtaining many results. I also want to thank the MS team in Rotterdam.

During my thesis I had the opportunity to exchange views with many scientists. Firstly, I wish to thank Paresh Vyas, Catherine Porcher and their teams for all the discussions and help they provided. Good luck to Isla Hamlett and Boris Guyot in their research. I had a lot of fun in Oxford! I would like to mention “the Kleinwalsertal meeting” that I very much appreciated. So thank you Sjaak and Guntram Suske for organizing such events.

Many thanks to the members (past and present) of lab 706 with whom I shared so much: Katy (also known as Katy Kato), Edgar, Tiago, Marianne, Javier, Vesna, Alireza and Jennifer. To my Paranimf Ernie, I enjoyed your ever-present good mood that contributed to an excellent atmosphere in the lab. Thank you to Albert who introduced me to Rotterdam. Carmen, you were here in the lab for just a couple of months but you contributed so much to my social life and I admire your speed of working! Good luck to the new members Debbie and Raymond. I really enjoyed my time in a such international lab. So, again, thank you all.

I thank all people I have interacted and had fun with from different labs (Niels, Sjaak, Wouter, Dubi and Dies) to name but a few. My social life in Rotterdam would have been different without you. It was always great to have a “boys night out” with our favorite spare ribs. Special thanks to Arnie for sharing “only one beer” in Sijf and to Miyata...who had one less! Arnie, good luck in your new job as a pilot. Thank you Helen.

Many thanks also to the secretaries Marike and Jasperina (who I kept teasing) for all the efforts and help. You have solved many problems that made my integration faster. Thank you also to the computer team, as well as Melle and her colleagues.

Finally, I wish to thank all of you who have been involved directly or indirectly in the completion of my thesis that I consider as a big step in my life. Thank you to Caroline for also being my Paranimf.

Un dernier mot pour exprimer ma reconnaissance à mes parents et à mon frère. Je les remercie d’avoir toujours été présents dans l’accomplissement de mes objectifs, dont celui de ma thèse aujourd’hui.

Patrick R.