ISAS - INTERNATIONAL SCHOOL
FOR ADVANCED STUDIES

Purification of NFE-3: a candidate factor involved in the β-globin gene switching

Thesis submitted for the Degree of
Doctor Philosophiae

Candidate: Stefania Bottardi

Supervisor: Prof. Claudio Santoro

Academic Year: 1996/97

SISSA - SCUOLA
INTERNAZIONALE
SUPERIORE
DI STUDI AVANZATI
TRIESTE
Strada Costiera 11

TRIESTE

ABSTRACT

The transcriptional regulation of the globin gene family in the erythroid lineage of hematopoietic cells provides an excellent system in which to study tissue and developmental stage-specific gene expression.

The human β-globin gene locus is located on the short arm of chromosome 11. It is composed of five linked active globin genes (ε, $^{g}\gamma$, $^{a}\gamma$, δ, β) that are arranged in the order of their developmental expression, and an upstream regulatory region, the Locus Control Region (LCR), in about 65 kb of DNA.

The proper sequential expression of each β-globin gene relies upon the integrity of *cis*-regulatory sequences located both proximally to each gene and distally within the LCR.

Such dynamic binary association is likely to be accomplished via protein-protein interaction between *trans*-factors bound to specific targets in the regulatory sequences (gene promoters and LCR), in a sort of a flip-flop mechanism.

In particular, the ratio of γ- to β-gene expression would depend on the specific affinities between their promoters and the LCR, reflecting the milieu of nuclear regulatory proteins within the cell.

The Hereditary Persistence of Fetal Hemoglobin (HPFH) is a genetically inherited condition characterized by the constitutive expression of the fetal γ-globin genes in adult life.

Natural non-deletional HPFH mutations within the distal CCAAT-box of the human γ-globin gene promoters affect the binding of several factors *in vitro*.

Among the nuclear factors capable to bind this region, NFE-3 is the only one whose binding is consistently decreased with all different HPFH mutations so far observed: the -117 G->A, the -114 C->T and the Δ13 deletion in the $^{a}\gamma$ promoter, and the -110 A->C in the $^{g}\gamma$ promoter.

NFE-3 was originally described as a factor present in extracts of erythroleukemic cell nuclei (K562, MEL, HEL) and capable to bind exclusively the distal and the unique CCAAT-boxes of the human γ- and ε-globin gene promoters.

In this work I have further characterized the NFE-3 binding activity from K562 cells and I have purified the proteins responsible for such an activity

An accurate analysis was carried out to differentiate NFE-3 from the other CCAAT-binding proteins (in particular from the ubiquitously expressed CP1/NF-Y) and to highlight its complex biochemical nature. In addition, it was observed that

both YY1 and Ku proteins coelute with NFE-3 and their correlation with NFE-3 activity was also investigated.

Purification of NFE-3 to near homogeneity, reveals that two polypeptides of 40- and 120-KDa segregate with its *in vitro* binding activity and several lines of evidence establish that these two proteins are the major components of such an activity.

Both polypeptides are actually under investigation by mass-mass spectrometry analysis.

The cloning of the cDNAs encoding for the 40-KDa and the 120-KDa proteins might be a further step toward a deeper understanding of the composite and fascinating world of β-globin gene regulation.

Part of this work has been published in the following articles.

Apezteguia, I., Calligaris, R., **Bottardi, S.**, and Santoro, C. (1994) Expression, purification and functional characterization of the two zinc-finger domain of the human GATA-1 *Prot. Expr. Purif.* 5, 101-107.

Bottardi, S., and Santoro, C. (1995) CTF/NF-1 binds the Stage Selector Element of the human γ-globin gene promoter *Bioch. Bioph. Res. Comm.* 215, 874-880.

Ronchi, A., **Bottardi, S.**, Mazzucchelli, C., Ottolenghi, S., and Santoro, C. (1995) Differential binding of the NFE3 and CP1/NFY transcription factors to the human γ and ε globin CCAAT boxes *J. Biol. Chem.* 270, 21934-21941.

Bottardi, S., Calligaris, R., Cogoi, S., Apezteguia, I., and Santoro, C. (1995) Alternative translation initiation site usage results in two functionally distinct forms of the GATA-1 transcription factor *Proc. Natl. Acad. Sci. USA* 92, 11598-11602.

INDEX

CHAPTER 1

INTRODUCTION

In higher eukaryotes the embryologic development and tissue differentiation are achieved through the execution of a genetic program that mainly determines which genes have to be transcribed and their rate of transcription in a particular cell type and in response to multiple physiological signals.

Over the past decade, innumerable studies carried out on hundreds of genes have highlighted the various mechanisms involved in the regulation of gene expression. The principles that we deduce from all these data can be broadly outlined as follows. The transcription of a gene depends on the presence of arrayed multiple *cis*-acting regulatory sequences. The activity of such sequences is exerted through the interaction with certain *trans*-acting factors. Multiple physiological signals and structural components influence the combinatorial interactions between *cis*- and *trans*-acting elements. Only certain combinations of complexes can functionally interplay with the RNA polymerase machinery. However, a lot has still to be understood on the interplay network that involves all the above mechanisms in the fine tuning of gene expression in those complex biological processes leading to cellular differentiation and development.

To this respect, the human β-globin gene family constitutes an ideal model of study.

1.1 The Human β-globin Gene Locus

The human β-globin gene locus consists of five linked genes (ε, $^{g}\gamma$, $^{a}\gamma$, δ, β) and a regulatory region, the Locus Control Region (LCR), spanning about 65 kb on the short arm of chromosome 11 (1, 2). The β-like globin genes are arranged in the order of their developmental expression: 5'-ε-$^{g}\gamma$-$^{a}\gamma$-δ-β-3' (Fig. 1).

The ε-globin gene is expressed in early embryo, in cells forming the blood islands in the yolk sac. The expression of γ-globin genes is limited at the fetal life and occurs in the fetal liver red blood cells. δ- and β-globin genes are expressed throughout adult life in the erythropoietic cells of the bone marrow.

The differential expression of globin genes during embryologic development is accomplished at the transcriptional level and is referred to as hemoglobin switching (3) (Fig. 2).

Globin genes in mice have a similar pattern of expression. The embryonic β-like globin gene (β^{h1}) is expressed in the blood islands of yolk sac from day of gestation 8 to 12. Then the embryonic to fetal/adult switch occurs: from day 12 on, the fetal/adult β-like globin gene (β^{maj}) is expressed in the liver till birth and then in the spleen and bone marrow (5, 6). Moreover, transgenic mice carrying the entire human β-globin locus show proper developmental switch of the human genes (7). The human ε and γ genes behave like the embryonic β^{h1} and ε mouse genes, whereas the human β gene behaves like the adult β^{maj} mouse gene. Thus, these data strongly suggest that the regulatory mechanisms underlying the control of time- and tissue-dependent expression of globin genes are highly conserved in these two mammals.

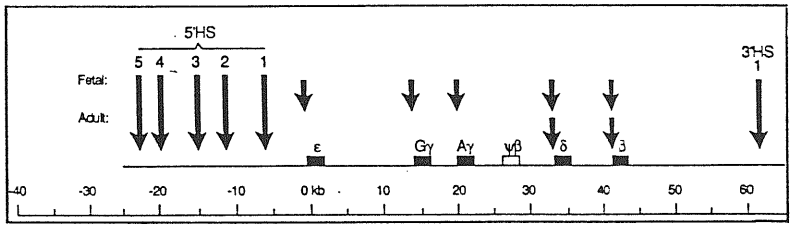

Figure 1. Location of nuclease-hypersensitive sites within the β-globin gene cluster. These sites in nuclear chromatin mark the position of nucleosome-free major sequences. Developmentally stable hypersensitive sites flank the cluster; HS-1 to -4 are erythroid specific and contain the locus control region (LCR) activity. Each promoter has a hypersensitive site in fetal stage erythroid cells, but only the δ-and β-genes are nuclease sensitive in adult stage erythroid cells (3).

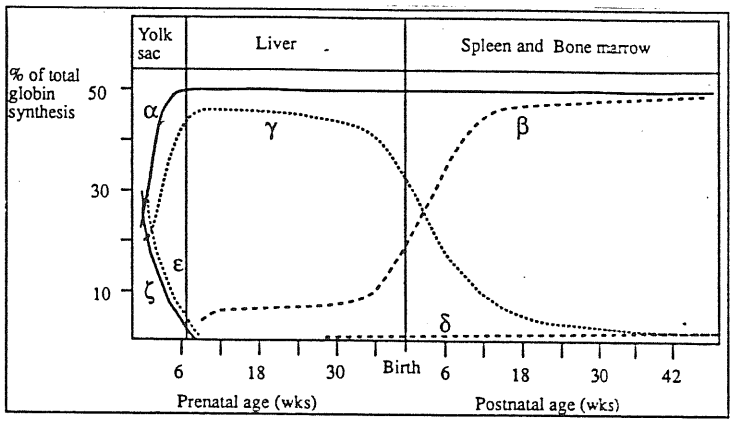

Figure 2. The human β locus and its expression during human development. (4).

1.1.1 β-Globin Locus Control Region (LCR)

The proper sequential expression of β-globin genes relies upon the integrity of *cis*-regulatory sequences located both proximally to each gene and distally (LCR). These sequences consist of multiple *cis*-acting elements that are targets for ubiquitous and tissue- and stage-specific transcription factors (8, 9) (Fig. 3).

The human β-globin LCR was initially identified as a cluster of distinct 200- to 500-bp erythroid-specific DNase I hypersensitive sites (HS-1 to HS-5) (1, 10, 11). These have been termed "major" or "super" HSs to distinguish them from the less sensitive gene-linked HSs (12). HS-1-4 (or HS-I-IV -13-) are located 6, 11, 14.5, 17.5 kb upstream the ε-globin gene, respectively, and are detected in early erythroid cell precursors - when gene-linked HSs are still inactive -, while HS-5 is constitutively and ubiquitously active. A less characterized sixth HS (HS-6, or 3' HS-1) is located 20 kb downstream the β-globin gene.

The first strong evidence that the LCR is an important *cis*-regulatory element consisted in the observation that in naturally occuring HS-2-4 deletions - Hispanic γδβ-thalassemia - the entire β-globin locus became late replicating and DNase I resistant and no β-globin mRNA was detected (13). From then, the LCR has been the subject of extensive analyses.

LCR activity appears to be dominant as it confers high-level and erythroid-specific expression on a number of *cis*-linked heterologous genes both in transgenic mice and in transfected cells (2, 14-16).

The main principle derived from the data collected so far is that the LCR contributes to the organization of the globin locus as a single large "chromatin-open" domain. This goal is achieved by an exquisite additive effect of the single HSs (17, 18, 49). To this regard, the three central tissue-specific HS sites (HS-2, HS-3, and HS-4) play a major role in elaborating the multiple activities attributed to the LCR, while the roles of HS-1 and HS-5 are poorly defined. Anyway, LCR must be complete to overcome heterochromatin silencing and it does so by ensuring that the locus is active all of the time in all of the (red) cells (254).

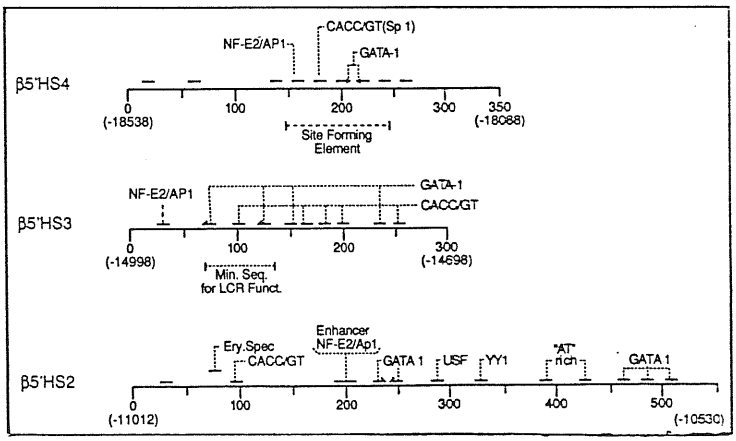

Figure 3. Disposition of binding motifs for various proteins within individual hypersensitive sites. (3).

1.1.1.1 Hypersensitive Site 2 (HS-2)

HS-2 consists of about 500 bp containing target sequences for several *trans*-acting factors such as NF-E2, GATA-1, USF and YY1 (19). Since its initial characterization, the HS-2 site has been considered the most effective regulatory element of the β-globin LCR: it has enhancer activity (20-24) and, in transgenic mice, can act as LCR when present in at least two copies (25). However, the insertion by homologous recombination of a hygromycine phosphotransferase gene between HS-2 and HS-1 of the β-globin locus, abolishes expression of the human β-globin gene, even though no deletion of LCR sequences occurs (26). The expression of the β-globin gene can be restored by recombinase-mediated deletion of the selectable marker, suggesting that when *neo* gene is present within the LCR it disrupts regulation, possibly by out-competing globin gene promoters (27). A similar result has been obtained in mice where HS-2 deletion, though decreasing to 70% the expression of the β-globin gene, has virtually no effect on ε- and γ-globin genes expression (28).

The conclusive assumption is that though HS-2 may have unique gene-specific functions in the intact LCR, its deletion does not abolish a sufficient number of binding sites to significantly alter the topology of the entire locus. Likely other HS sites might perform HS-2 functions whenever it is deleted.

1.1.1.2 Hypersensitive Site 3 and 4 (HS-3 and HS-4)

HS-3 and HS-4 confer position independent (18, 24, 29) and developmental stage-specific expression of linked γ- and β-globin genes, though they have weak enhancer activities. HS-3 is the most active site in embryonic yolk sac and fetal liver (30). It is capable of directing β-globin gene expression from several independent integration sites in mice containing single-copy transgene (31), due to a likely chromatin-opening activity. HS-3 core fragment contains a triple repeat of a combination of GATA-1 binding sites and G-rich sequence that are spaced 30 bp apart (3). Deletion analyses have revealed that the combination of a G-rich sequence flanked on each side by one binding site for GATA-1 is essential to obtain position-independent expression of a linked β-globin gene in erythroid cells (33). On the contrary, binding sites for NF-E2 and GATA-1 are required for the formation of HS-4 chromatin structure following stable transfection into murine erythroleukemic cells (MEL) (34).

1.2 Heterokaryons

Analysis of globin gene expression in heterokaryons between human fibroblasts and MEL cells and between MEL and K562 cells, demonstrated that diffusible cellular factors are involved in the specific activation of the globin locus

(11). Papayannopoulou et al. (35), by transferring normal human lymphoid cell chromosomes into MEL cells, obtained hybrids which expressed only the human adult β gene. Moreover, minicell hybrids carrying human chromosomes containing mutations responsible for the Hereditary Persistence of Fetal Hemoglobin (HPFH) either deletional - δβ-globin gene deletion - or non-deletional - -117 $^{a}\gamma$ G->A - expressed the human γ-globin gene. More recently, Stanworth et al. (36) have elegantly shown that hybrids between MEL cells and erythroblasts derived from mice carrying human γ- and β-globin transgenes express these depending on the pattern of expression in the erythroid cells at the time of fusion, sustaining that epigenetic modifications acquired during development and differentiation may also play a role in determining expression of globin genes in hybrid cells.

1.3 Globin Genes Regulation

Developmental and tissue restricted expression of the γ- and β-globin genes in transgenic mice can be obtained also in the absence of LCR (37-40). However, the level of expression is low and sensitive to the site of chromosomal integration and copy number of the transgene (41). Only the ε-globin gene expression seems to be totally LCR-dependent (42-44). These data confirm that the genetic informations leading to the developmental regulated expression of β-globin genes reside within the genes themself. The LCR sequences, somehow, seem to perform two major activities: the optimization of the erythroid expression of *cis*-linked genes and the commitment to a temporal-scheduled expression of the linked genes by first activating the closest one. To this respect, it is worthy to note that a transgene consisting of human LCR and a linked β-globin gene is prematurely expressed in the yolk sac (45, 46). When another gene is placed in *cis* between the LCR and the β-globin, or the whole cluster is used (47), the correct temporal expression of the β-globin gene is restored.

Several studies of this sort lead to the model of the competitive interaction, that is the expression of the β-globin genes is influenced by the polarity of the genes with respect to the LCR and possibly by the formation of a single complex between LCR and proximal promoter sequences at any given developmental stage (7, 45-48). In addition, Peterson et al. (50) proposed that in transgenic mice carrying tandemly arranged human γ- or β-genes and γβ or βγ constructs in μLCR, the primary determinant of appropriate gene expression during development is the *trans*-acting environment. They shown that gene order, or distance from the LCR, influences the degree of expression - especially when two genes are structurally identical - but promoter competition plays the major role in controlling developmental switching.

All together, these data confirm that crucial DNA sequences, necessary to mediate developmental stage-specific expression of globin genes, are located within or near the genes themselves and are discriminatory partners of interaction with LCR sequences.

1.3.1 The Enhancer Competition Model

The enhancer competition model was proposed by Choi and Engel (51) to account for the embryonic ε- to the adult β^A-globin gene switching in chicken (52). The model hypothesizes competition between different promoters and a unique shared enhancer. In the embryonic environment, physical interaction between the ε-promoter and the enhancer should be favoured, while in the adult environment the stage-specific *trans*-acting factor NF-E4, which binds to a purine-rich sequence (the Stage Selector Element -SSE-) of the β^A promoter (53), should favour the association of this promoter with the enhancer, leading to the activation of the β^A gene and to the silencing of the ε-globin transcription. This model is supported by the observation that NF-E4 is present in mature adult but not in primitive erythrocytes.

1.3.2 Polar Competition Model

Hanscombe et al. (48) have shown that the β-globin genes order plays an important role in the developmental control of transcription. They showed that all globin genes are in polar competition for the LCR activating function at any developmental stage, being the more proximal genes advantaged for interacting with LCR sequences over the distal ones and being the β-globin gene suppressed in the embryo only when there is transcriptional competition from linked γ-globin genes (46, 48, 54).

Polarity is also sustained by the analysis of individuals with non-deletional Hereditary Persistence of Fetal Hemoglobin mutations (HPFH). The persistent expression of the fetal γ-globin gene, associated with the presence of point mutations in its promoter, results in a down-regulation of the *cis*-linked β-globin gene. Also the precise γ/β and $^G\gamma/^A\gamma$ ratios observed during development might reflect their polar distribution,.

1.3.3 Autonomous Gene Regulation

This model accounts for the relevant role played during development by specific elements in directing the interaction of the single ε-, γ-, and β-globin promoters with LCR sequences. These elements are targets for stage-specific factors which mediate the functional interaction between LCR and promoter regions. The

obvious consequence is that there have to be positively and negatively acting factors. In agreement to this model, it is worthy to note that functional silencer elements have been mapped in the promoters of the ε- and γ-globin genes (43, 55).

How do LCR and promoter complexes interact? What is the nature of this communication network and how is it carried out?

A sort of dynamic binary association of LCR and gene promoter sequences must be achieved and a remarkable number of models and mechanisms have been proposed to explain how cross-talk between distant elements might be accomplished.

Looping-out Mechanism. This model implies that the functional interaction between LCR and promoter sequences occurred via protein-protein interactions between factors bound to specific targets in the regulatory regions. This physical network can be simple and direct or can be mediated by additional partners (e.g. coactivators). Whenever this sort of zip is formed, the chromatin spanning between the regulatory regions will be looped-out (56).

Tracking Mechanism. According to this model, the transcription factors and/or helicases can bind to the LCR and migrate to a downstream promoter by progressive and developmental-specific sliding along the DNA (23, 57).

Wave of Supercoiling. This model is related to the previous and accounts for a determinant role played by chromatin structure in directing the hierarchy of globin genes expression. Supercoiled chromatin regions should progressively involve the promoters of the genes to be silenced. At the same time, this stress will favour chromatin changes in the downstream genes making their promoter elements available for the interaction with activating factors.

Whatever mechanism, dynamic chromatin changes are considered to play a pivotal role in the regulation of β-globin genes expression. In agreement to this principle, Wijgerde et al. (58) have shown *in vivo* that the β-globin LCR activates only one gene at a time by a monogenic interaction and that coexpression of multiple genes from a single locus involves alternate transcription. Furthermore, they have shown that these interactions are not static but dynamic, with a LCR holocomplex stably involved with one gene at a time and switching back and forth between genes in a flip-flop mechanism. The LCR appears to interact in a stochastic fashion with gene promoters. The ratio of γ- to β-gene expression would depend on the specific affinities between their promoters and the LCR, reflecting the milieu of nuclear regulatory proteins within the cell (Fig. 4).

1.4 Stage Specific Factors

The existence of stage-specific activators or repressors able to control the developmental profile of globin gene expression has been debated for years. What is becoming clear is that globin switching reflects the cumulative interactions of *trans*-acting factors within the LCR and promoter sequences throughout the locus, rather than being the effect of a single stage-specific regulatory factor.

1.4.1 Stage Selector Elements (SSE)

A putative Stage Selector Element (SSE), mapped in the chicken β-globin promoter, mediates temporal regulation through the binding of the developmentally specific protein NF-E4 (51, 53). A similar element has been identified in the human γ-globin gene promoter. *In vitro* and *in vivo* - in K562 cells - such an element allows preferential interaction of the γ-promoter with HS-2 when in competition with a linked β-promoter (59). The γ-globin SSE is bound by three proteins: Sp1 (59, 226), CTF/NF-1 (60) and a complex (SSP, Stage Selector Protein) consisting of CP2 and an unknown partner protein of 40-45 KDa (61, 62). Binding of SSP correlates with SSE activity, a fact emphasized by evolutionary phylogenetic footprinting studies that demonstrate loss of the protein binding-sites in species without fetal expression of the γ-gene (63).

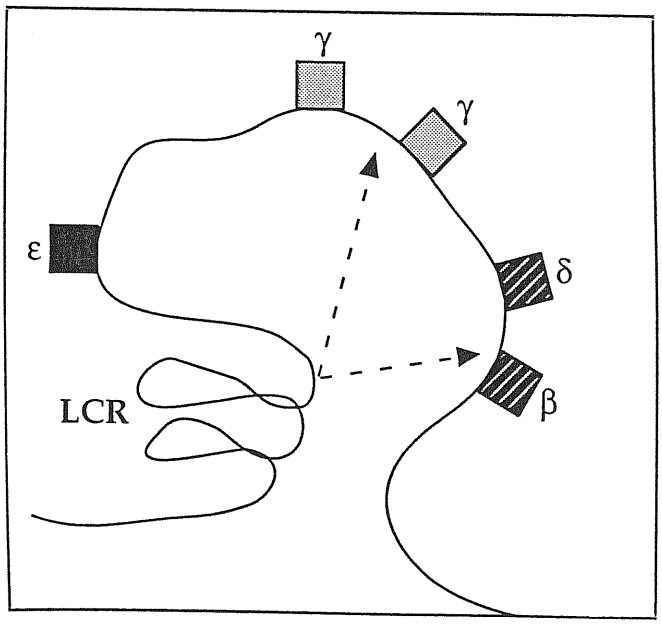

Fig. 4. A model of the interactions of the LCR with the different globin genes through a stochastic looping mechanism. The LCR is indicated as a squiggly line to indicate that the different regions that are hypersensitive to DNase I of the LCR could act together or even form a holocomplex to establish an interaction of the LCR with one of the genes (58).

1.5 *Trans*-Regulators

The majority of models proposed to explain the details of globin gene switching, considers protein-protein interactions as the main mechanism to build the functional framework between LCR and gene promoters (reviewed in 8, 9, 64, 65).

1.5.1 GATA-1

GATA-1 [or Eryf1 (66), or GF-1 (67), or NF-E1 (68)] is a 47-KDa protein that binds to the consensus sequence T/AGATAA/G. The DNA-binding domain consists of two related zinc-fingers (carboxy- and amino-terminal finger) of the type C-X-N-C-(X_{17})-C-N-X-C. GATA-1 is the founder member of a wide family of proteins that show a high degree of homology restricted to their finger domains (38, 75, 76, 87-90), among mammals (66, 69, 85, 86), avians and amphibians.

The C-terminal finger mediates both the binding to DNA (70) and the physical interaction with other zinc finger-containing proteins, such as Sp1, EKLF (71), other GATA-binding factors, and GATA-1 itself (72, 73). The GATA-1 N-terminal finger is dispensable for binding to DNA, though its presence stabilizes the binding to low affinity GATA motives (74).

GATA-1 is expressed in most hematopoietic cell lineages (75-78) and in Sertoli cells (79). Targeted disruption of the mouse GATA-1 locus is associated with lack of erythroid differentiation in chimeric mice (80), while the development of mast cells and megakaryocytes appears largely unaffected. *In vitro* differentiation of GATA-1 null ES cells shows that primitive erythroid cells differentiate normally up to the stage of pro-erythroblast and then undergo to apoptosis (81). This observation emphasizes the role of GATA-1 as a survival (or anti-apoptotic) factor (82) in differentiating erythroid cells.

Recently, a new 40-KDa isoform of GATA-1 has been identified. This isoform results from an alternative translation initiation site usage, seems differently expressed in developing mouse and shows a decreased transactivation potential with respect to the full-length protein (83).

A related protein, GATA-2 (75, 76, 89), is expressed in selected hematopoietic cell types and mice lacking the GATA-2 gene die prematurely with severe anemia (84). The effect of GATA-2 in the regulation of β-globin genes expression seems to be indirect. Likely, GATA-2 is one of the targets of the hematopoietic growth factor signals that sustain erythroid cell differentiation.

1.5.2 NF-E2

Nuclear Factor Erythroid 2 (NF-E2) is a leucin-zipper heterodimer consisting of a large subunit (p45), expressed in hematopoietic tissues and a small

subunit (p18), widely expressed. The DNA-binding domain of p45 is related to the *Drosophila* protein cnc (cap 'n collar), involved in head segmentation and to the *C. elegans* skn-1, a protein critical for blastomere determination. p18 is related to the human retinal protein NRL and to the chicken oncoprotein v-Maf (91-93).

In erythroid cells, NF-E2 is the major AP-1-like DNA binding protein. NF-E2 recognizes and binds a site containing an intact AP-1 binding motif, TGA(C/G)TCA, preceded by a G residue two base pairs upstream. Functional analysis of tandem AP-1-like motifs present in the HS-2 (21, 22) and the HS-4 (94) of the β-globin LCR, indicates that NF-E2 partially mediates activity of these elements *in vivo*. Target sites for NF-E2 have been shown in other gene promoters such as those coding for the porphobilinogen deaminase and the ferrochelatase , as well as in the α-globin LCRs.

Mice lacking the gene for the p45 subunit of NF-E2 develop normally (95). These findings lead to the hypothesis that, though NF-E2 might be an important regulator of globin gene expression and iron metabolism, it is dispensable for hematopoiesis in mice. On the other hand, given the type of binding site, it might be that other AP-1-like proteins can compensate for the absence of NF-E2.

1.5.3 EKLF

The erythroid *Kruppel*-like factor (EKLF) is a Sp1-related transcriptional activator containing three zinc fingers. It binds specifically the GC or GT/CACC motif (CACCC box) (96), found in several erythroid specific gene promoters and within the LCR hypersensitive sites (97). The expression of EKLF is restricted to few hematopoietic cell lineages, being abundant in both embryonic and definitive erythroid cells and low in mast cells (98). The binding of EKLF to the β-globin CACCC-boxes is abolished by mutations responsible for some β-thalassemias (96).

The role of EKLF has been defined by gene disruption experiment. EKLF null mice die of anemia. Interestingly, either Wijgerde et al. (99), analysing transgenic mice homozygous for a single copy of the entire human β-globin locus in an $EKLF^{+/-}$ or $EKLF^{-/-}$ background, or Perkins et al. (100), interbreeding $EKLF^{+/-}$ mice with mice harboring a human β-globin transgene, observed that in absence of EKLF the γ-globin gene expression is normal or increased in fetal liver, while the human β-globin gene expression is absent (99) or dramatically reduced (100). Thus, despite the presence of putative consensus sites in other globin genes, EKLF lack leads to a highly selective deficiency in β-globin gene expression, mimicking the thalassemia data. This defect has to be ascribed only to the lack of adult β-globin expression since the embryonic globin genes are correctly expressed (98, 101).

These results, consistent with the competitive switching model, indicate that EKLF plays an important role in the LCR/β-promoter interaction and in the γ-/β-gene competition.

1.5.4 Tal-1/SCL/TCL-5

The protein SCL (Stem Cell Leukemia), known also as Tal-1 (T-cell acute lymphoblastic leukemia), or TCL-5, is a bHLH transcription factor that binds to the consensus sequence CANNTG (CAGATG) and is a critical regulator of hematopoiesis. It is normally expressed in erythroid cells, in mast cells, in megakaryocytes, in early myeloid cells and in committed progenitor cells ($CD34^+/CD38^+$) (89, 102-104). The SCL gene encodes a helix-loop-helix (HLH) transcription factor which is aberrantly expressed in the majority of cases of pediatric T-cell acute lymphoblastic leukemia (105). The constitutive expression of SCL in T-cell leukemic cells might represent a critical step in the leukemogenic process as SCL is not expressed in normal mature T-cells. The N-terminal transactivation domain is dispensable for oncogenesis, supporting the hypothesis that the scl gene product exerts its oncogenic action through a dominant-negative mechanism (256).

Several SCL transcripts are observed *in vivo* resulting either from alternative transcription start sites or by alternative splicing. Only one splice variant is predicted to affect the protein product and produce an amino-truncated polypeptide (22-26 KDa) which lacks a putative transactivation domain (106). The full-length protein is phosphorylated on serine *in vitro* and, possibly, *in vivo*, by the ERK1 protein kinase (107).

SCL null mice die around day 9 of gestation due to failure of hematopoiesis (108, 109). In agreement with previous data, $scl^{-/-}$ ES cells make a substantial contribution to all non-hematopoietic tissues but do not contribute to any hematopoietic lineage (110).

It has been suggested that GATA-1 and GATA-2 are involved in the expression of the scl gene since two functional GATA-motifs have been mapped in the erythroid specific scl gene promoter Ia (111, 112).

1.5.5 Sp1

Sp1 is an ubiquitous *trans*-activator that binds to the consensus sequence GGGCGG (GC box) (113, 114). It contains three contiguous zinc-fingers and is involved in transcriptional activation of a variety of cellular genes (115, 116). Cooperation between GATA-1 and Sp1 factors has been clearly proved in promoters, enhancers, and LCRs of erythroid specific genes (33, 71).

1.5.6 LCR-F1 (Nrf1)

NF-E1 Related Factor 1 (Nrf1) (117) or Locus Control Region Factor 1 (LCR-F1) (118) and NF-E2 related factor 2 (NRF2) (119) are widely expressed proteins, closely related to the p45 subunit of NF-E2 (65). LCR-F1 is a basic leucine-zipper protein that binds to the duplicated AP1-like sites within HS-2. Although LCR-F1 expression is not restricted solely to erythroid cells, it contains an acidic domain that exerts a strong erythroid-specific transactivation activity (118).

Gene inactivation by homologous recombination in embryonic stem cells demonstrates that LCR-F1 is essential for mesoderm formation, while it is dispensable for globin gene expression (255).

1.5.7 USF/MLTF

USF (Upstream Regulatory Region Factor) or MLTF (Major Late Transcription Factor), is an ubiquitous basic helix-loop-helix zipper protein that binds an E-like box within the HS-2 of the human β-globin LCR (25).

1.5.8 HS2NF5

HS2NF5 (HS-2 Nuclear Factor 5) is a 35-KDa protein that binds the sequence TGTTCTCA within the HS-2 of the β-globin LCR. This site overlaps the binding site for the erythroid-specific transcription factor SCL. Mutations that abolish its binding reduce the enhancer activity of HS-2 in transient and stable transfection assays (120).

1.5.9 NF-E6

NF-E6 (Nuclear Factor Erythroid 6) is an erythroid-specific factor that binds to the γ- and β-globin CCAAT boxes. NF-E6 binding does not correlate with changes in transcription of the wild type γ-globin gene and is not affected by the -117 HPFH mutation (121).

1.5.10 CTF/NF-1

CTF/NF-1 (CCAAT binding Transcription Factor; Nuclear Factor 1) consists of a family of polypeptides of 52-66 KDa, that bind with high affinity to the symmetrical sequence TTGGCT-(N3)-AGCCAA, but that can also strongly bind to the half AGCCAA sequence (122).

CTF binds to two regions in the γ-promoter: from -53 to -35 (SSE) (60) and from -140 to -118, overlapping the transcriptionally active CACCC element. CTF interferes with the synergistic transcriptional activation of the γ-globin promoter mediated by

GATA-1 and Sp1 via competitive displacement of Sp1 from the CACCC element (123).

The very high affinity CTF-binding sites located in the distal upstream region of the human β-globin gene promoter (-125 to -300) do not contribute to the expression of the gene in transient expression experiments, whereas the much lower affinity binding site located at position -80 is fully functional. Thus, neither the mere presence of of a CCAAT sequence in the vicinity of a promoter nor the binding of CTF to fortuitous CCAAT element is sufficient per se to create a promoter responsive to CTF, but rather, its ability to activate transcription is determined by the context of the binding site in relation to the location of other transcription factors (122).

1.5.11 CP1/NF-Y

CP1/NF-Y is a heterodimer made of a 42-KDa (YA) and a 32-KDa (YB) protein (124, 243), homologous to the yeast transcriptional activator HAP2 and HAP3 respectively (126). CP-1/NF-Y is an ubiquitous activator that recognizes CCAAT boxes in a large number of gene promoters, including those of the α- and β-like globin genes. It binds very strongly to the CCAAT box of the human β-globin gene promoter (127); its *in vitro* binding to the distal CCAAT box of the γ-gene promoter (see later) does not show any modification during development (121).

NF-YA has histidines in a $H(X)_3H$ disposition sequence, while NF-YB has cysteine residues, $C(X)_3C$, in the same disposition, together reminescent of the zinc finger motif. It has been hypothesized that these two "half-fingers" may participate in a tetrahedrical metal coordination complex (128).

1.5.12 CDP

CDP (CCAAT Displacement Protein) binds with high affinity to sequences overlapping the CCAAT element thus preventing the binding of CCAAT-binding factors. In the histone H2B gene promoter of sea urchin embryos, the binding of CDP abolishes the expression of the H2B gene in sperm (129). In vertebrates, the CDP homologue is an ubiquitously expressed protein that exhibits binding activities very similar to that of CTF/NF-1, but is unable to bind the single CCAAT box from β-, ε-, and α1-globin promoters (130).

1.5.13 YY1

The zinc finger protein YY1 (Yin Yang 1 -131-), also known as NF-E1 (Nuclear Factor μE1 -132-), or δ (133), or UCRBP (Upstream Conserved Region Binding Protein -134-), or NMP-1 (Nuclear Matrix Protein 1 -135-), is a DNA-binding protein belonging to the GLI-*Kruppel* family. It is widely expressed and highly

conserved between human and mouse. It is a 414 amino acids protein, with a calculated molecular weight of 44 KDa that migrates in SDS-PAGE as a 68-KDa polypeptide. YY1 is heat-stable (136) and contains four zinc fingers of the Cys-Cys-His-His type, while a very unusual N-terminal domain includes stretches of 11 consecutive negatively charged amino acids and 12 consecutive histidines.

Binding sites for YY1 have been reported in a variety of promoters.
Binding site selection analysis (136, 137) established that the CCAT sequence represents the core of YY1 high-affinity binding target. A lower affinity consensus contains the ACAT motif. At present, there is no evidence that ACAT and CCAT sites are functionally different, but the ACAT consensus has been observed only in adeno-associated virus P5 (AAV-P5) and in the human γ-globin gene promoters (131, 61).
Another core sequence, TCAT, found in the human papillomavirus type 18 promoter (139), in the E6/E7 promoter of HPV16 (140) and in the human ε-globin upstream region (141) is bound by YY1 *in vitro* 5 to 10-fold less avidly than the CCAT/ACAT motives (136).

The nucleotides located on the 5' and 3' boundaries of the core are relatively flexible. In principle, this variability in binding sites should allow YY1 to bind and influence transcription within a wide variety of promoters. YY1 is considered a multifunctional nuclear factor since it can activate, repress, or initiate transcription depending on the context of *cis*-acting genetic elements.

The suppressive effect of YY1 can be obtained through: i) competition with other transcriptional activators for overlapping binding sites; ii) DNA-bending that turns bound activators away from the initiation complex; iii) sequestration of an activator protein through protein-protein interactions.

YY1 represses transcription of a variety of cellular promoters, including those from skeletal α-actin (143, 144), c-fos (145), ε-globin (146, 147), γ-globin (61), α-globin (148), creatine kinase M, α1 acid glycoprotein and β-casein genes. In addition, YY1 represses the immunoglobulin (Ig) κ 3' enhancer (132) and it inhibits transactivation of the human γ-IFN (γ-Interferon -149-) and GM-CSF (Granulocyte-Macrophage Colony Stimulating Factor -150-) promoters mediated by AP-1 or Sp1/AP-1 proteins, respectively. YY1 mediates transcriptional repression also in viruses: Moloney murine leukemia virus (134), human papillomaviruses (139), Epstein Barr virus (134), human cytomegalovirus, human immunodeficiency virus type I (151), parvovirus (131), adenovirus, and AAV. In particular, it binds to a motif centered at -60 relative to the initiation site of AAV P5 promoter that mediates transactivation by the E1A protein (152). Recently, it has been shown that E1A protein normally relieves the YY1-mediated transcriptional repression by complexing with it (153). Also c-myc

(154) and B23 proteins (155) bind to YY1 and abrogate its ability to repress transcription.

As a positive regulator, YY1 activates c-myc (156), ribosomal protein (133), and GM-CSF (157) promoters. Activation and, moreover, transcription initiation of AAV-P5 promoter by YY1 involves protein-protein interaction between YY1 and Sp1, in particular between one and half zinc fingers, as well as the portion immediately preceding the zinc finger region of YY1 and the C-terminal region, encompassing the zinc finger DNA-binding domain and the transcriptional activation domain D of Sp1. The interactions with Sp1 protein occurs both *in vivo*, and *in vitro*, and in solution also without DNA (143, 158, 159).

YY1 recognizes and binds *in vitro* to the c-fos SRE (Serum Response Element) and the skeletal actin MRE (Muscle Regulatory Element). Functional antagonism between YY1 and the SRF (Serum Response Factor) at the level of both c-fos SRE and α-actin MRE was initially demonstrated (145). Then, Natesan et al. (160) shown that YY1 facilitates the association of SRF with c-fos SRE. They interpreted this behaviour as a competition between the SRF-DNA complex and free DNA for YY1; increased SRF concentration directs YY1 toward the ternary complex (SRE-SRF-YY1) and away from the simple binary complex (SRF-SRE) with DNA.

YY1 plays a role also during development: in primary myoblasts, it simultaneously inhibits and activates expression of the skeletal α-actin and c-myc genes, respectively, modulating in such a way myoblasts growing and differentiation; the transrepression activity depends on its C-terminal zinc finger region, while its transactivation activity requires an additional N-terminal domain (144).

Under the appropriate *in vitro* conditions YY1 can direct and initiate transcription at nucleotide position +1 in the AAV-P5 promoter (131, 162), being able to interact directly with TFIIB and RNA Polymerase II, in the absence of other known auxiliary factors, including TFIID (163). Besides, YY1 can also specifically interact with $TAF_{II}55$, a human TFIID TAF (164).

It is very likely that YY1 may mediate the critical switch between gene activation and repression; in all cases studied this phenomenon requires the interaction of YY1 with a second protein, highlighting its role in establishing protein-protein interactions within gene promoters.

1.5.14 Ku/DNA-dependent Protein Kinase (DNA-PK)

Chromosomal double strand breaks (DSB_s) may occur spontaneously, or during DNA recombination (e.g. V(D)J recombination), or may be induced by DNA damage (ionizing radiations -IR-, alkylating agents, oxidative stresses).

Nonhomologous DNA end-joining recombination is the primary mechanism of DSB_s repair in eukaryotes.

Intriguingly, it was shown that a DNA-dependent protein kinase [DNA-PK (165)] binds to, and is activated by, DNA double stranded (ds) ends. It was also observed that Ku protein (166) interacts with free ends of ds DNA without almost any sequence-specificity, serving as the DNA binding component of DNA-PK. All together, these findings led to the suggestion that both factors might be involved in DNA repair/recombination process, and studies on DSB_s repair-defective mutants isolated from bacteria, yeast and mammalian cells recently confirmed it.

1.5.14.1 Ku

Ku is an abundant nuclear protein originally identified as an autoantigen recognized by sera from patients with autoimmune disorders ('Ku' being derived from the first patient's name -166-). It is a heterodimer composed of two polypeptides of 70 KDa (p70) and 86 KDa (p80) (167), that behaves as a 300 KDa-polypeptide when analysed by gel-filtration chromatography (166, 168). cDNAs of both subunits have been cloned (169-171) and shown to contain no significant homologies to other proteins (169, 171).

Ku binds indifferently to DNA ends whether 5' protruding, blunt, or 3' recessed, phosphorylated or not, and binds - with lower affinity - to linear or circular single stranded (ss) DNA (167, 172). It binds also to DNA nicks and to DNA termini ending in hairpin structures, perhaps by recognizing transition forms from dsDNA to ssDNA without specificity for any particular nucleotide sequence or termini configuration. Ligation of DNA ends after binding of Ku can trap the protein on the DNA. By this trik it has been shown that Ku might have some sequence-specificity (173-175). It seems that also RNA/DNA duplex may provide Ku targets, as they contain ssRNA ends (176).

Ku binds and selectively protects from DNase I digestion approximately 25 nucleotides at both 5' and 3' termini, and 30-35 bp are required per Ku molecule (173). Once attached to DNA termini, it translocates along the DNA in an ATP-independent manner (167). The sliding is processive; once loaded onto a fragment of DNA, there is very little exchange of Ku protein to other fragment, unless DNA ends are present (175). These peculiarities can explain why in mobility shift assay ladders of shifted bands proportional to the length of DNA are constantly seen (173).

Initial data indicated p70 as the active DNA-end binding member of the dimer (177, 178), being able to carry out this activity even in the absence of p80 (179-182). However, more recent and convincing works have indicated that p70 is not enough, and p80 is dispensable for DNA binding (168, 183).

Ku is the DNA-binding component of the DNA-PK holocomplex in eukaryotic cells (184-186); it is phosphorylated *in vitro* by DNA-dependent protein kinase catalytic subunit (DNA-PK_{cs}) and possesses also ATPase activity (125). Ku first binds to the DNA, then it recruits p350 DNA-PK to the complex. Its ability to slide along the DNA may have a physiological significance in terms of delivering the kinase catalytic component to substrates bound at fixed positions on the template.

Besides its characteristic ability for non-specific interactions with DNA, the participation of ku antigen in sequence-specific DNA binding has been described in some cases. This observation is of considerable importance because it might create a linkage between the excision repair mechanism and the transcriptional apparatus. Footprints of Ku proteins have been observed on the promoters of U1 RNA (by PSE1, Proximal Sequence Element-binding protein -187, 188-), transferrin receptor (by TREF-1 and -2, phorbol 12-tetradecanoate 13-acetate-Responsive Element Factor -189-190-), human collagen IV (by CTCBF, CTC box Binding Factor -252-), heat shock response element (by CHBF, Constitutive hse Binding Factor -194, 195-), and ribosomal DNA genes (by E1BF, Enhancer 1 Binding Factor -191, 192-), U5 region of the human T-cell leukemia virus type 1 (HLTV-1) LTR (193) and T-cell receptor β enhancer (178). In addition Ku gives footprints on the octamer motif (196). It has also been shown that Ku/DNA-PK can directly and specifically bind to NRE1 (Negative Regulatory Element 1) of mouse mammary tumor virus (MMTV), repressing glucocorticoid-induced MMTV transcription (197); other data suggest that Ku protein may repress Polymerase I transcription and that UBF (Upstream Binding Factor) may relieve this inhibition (198).

Ku seems to bind, together with UBF, in a human DNA region (B48) previously recognized to contain a DNA replication origin (199) and, linked to this observation, Ku has also been identified as human helicase (211).

However, some of these reports are a bit controversial and the Ku binding-specificity is still not universally accepted; more than one authors have isolated Ku as contaminant during DNA-affinity chromatography (201).

Ku exists within the nucleolus and nuclear cortex (202) and the nucleolar localization varies throughout the cell cycle; Ku and DNA-PK_{cs} are present in the nucleoli of G2-phase cells, but neither are detectable in the nucleoli of S-phase cells, suggesting that gene expression for these proteins is associated with the proliferative state of the cells (203, 204).

Ku homologs have been found in monkey, mouse and *Drosophila* (205); the recent identification of a Ku-like activity also in *S. cerevisiae* suggests that Ku/DNA-PK is ubiquitous in the eukaryotic kingdom.

1.5.14.2 DNA-PK

DNA-PK (DNA-dependent protein kinase) is a serine/threonine kinase that is absolutely dependent for activity on binding to ds DNA containing broken ends, nicks, and ss gaps (165, 200). DNA-PK consists of a catalytic subunit (DNA-PK_{cs}, a large polypeptide of 350 KDa) and a DNA binding component (Ku). DNA-PK becomes activated only when recruited on DNA by Ku. Once activated, DNA-PK is capable of phosphorylating a number of substrates *in vitro*, including several transcription factors (Sp1, Oct1, Oct2, SRF, Fos, Jun, Myc) (113, 176), proteins implicated in the cellular response to DNA damage (p53, Ku, hsp 90, and the 34 KDa subunit of replication protein A) (165, 200), topoisomerase II, SV40 large T-antigen, and the C-terminal domain (CTD) of the largest subunit of RNA Polymerase II (207, 208). These substrates are only phoshporylated effectively when they are bound to the same DNA molecule as DNA-PK, suggesting that DNA-PK activity *in vivo* might be restricted to proteins relatively close to the site of its specific activation (176).

DNA-PK is capable of autophosphorylation, a process that leads to its inactivation (165) and it has been implicated in several nuclear processes, besides DSB_S repair and V(D)J recombination, including transcription (RNA Polymerase I transcriptional repression -204-, collagen IV genes transcriptional regulation -252-), and DNA replication (185, 186, 209, 210).

The three subunits of DNA-PK (DNA-PK_{cs}, p70- and p80-Ku DNA-binding subunit) are all needed for both IR repair and V(D)J recombination: Ku plays a direct role in the signal join formation (end joining), while DNA-PK_{cs} preferentially acts during the processing of coding joining formation (coding ends).

p80 Ku defects are responsible for the DSB_S repair deficient mutational group 5 (XRCC5) and for the absence of any DNA-PK activity in mutant cell lines like sxi-1 and xrs-6 (161, 184-186, 210). V3 and scid group mutant cells are defective in DNA-PK_{cs} gene (210).

1.6 γ-Globin Genes Regulation

The evolutionary duplicated human $^{g}\gamma$- and $^{a}\gamma$-globin genes contain identical proximal promoter elements, which are separated by about 5 kb of DNA and are transcriptionally active in human fetal liver and in some human erythroleukemic cell lines (e.g. HEL, K562). These genes encode proteins which differ in only a single amino acid at position 136: glycine in $^{g}\gamma$ and alanine in $^{a}\gamma$ (138).

Despite major advances in the understanding of the structure of human hemoglobin genes, the mechanism controlling the γ- to β-switching has remained

elusive and both autonomous and competitive models have been proposed to explain the turning off of γ-globin genes in adulthood.

a) The autonomous control is thought to be achieved through the action of *cis*-acting silencer elements within the γ-globin gene promoters. In transgenic mice, negative elements located between -382 and -730 and between -141 and the transcription initiation site, are required for γ-globin gene silencing (206). The latter *cis* element includes the distal CCAAT box (-115 -111 region) and the SSE (-53 to -35 region).

b) The competitive model assumes that the preferential interaction between the β-globin gene and the LCR in the adult stage of development determines the turning off of γ-globin genes.

Data accumulated so far, strongly support the autonomous silencing rather than gene competition as the primary mechanism for turning off γ-globin expression in adulthood.

Pharmacological induction of fetal hemoglobin (HbF, α2γ2) in humans has been achieved by several agents. These include cytotoxic drugs (5-azacytidine, hydroxiurea) that induce HbF by altering the kinetic of erythropoiesis, hemopoietins (erythropoietin) that stimulate HbF production by inducing rapid erythroid regeneration, and butyric acid that acts by preventing the silencing of γ genes rather than by reactivating γ genes that have already been silenced. Butyrate (212) and its analogues (short-chain fatty acids with up to nine carbons -213- and α-amino butyric acid, α-ABA -214-) inhibit the fetal-to-adult globin gene switching both *in vitro* and *in vivo*. It has been proposed that the mechanism responsible for this effect might involve the increase of hyperacetylated histones due to inhibition of the histone deacetylase (215). Up to date, two Butyrate Response Elements (BRE) have been mapped between -382/-730 and -730/-1350 of the ^{a}γ-gene promoter, but their relevance *in vivo* is not precisely established (216).

1.7 γ-Globin Gene Promoters

The γ-globin promoters contain several elements whose sequences and localization are highly conserved; *cis* elements for transcription factors such as GATA-1, CTF/NF-1, Sp1, OTF-1, NFE-3, CDP and CP-1/NF-Y have been well characterized (Fig. 5).

GATA-1 binds to the -195 to -170 region and contacts two TATCT (AGATA) motifs (-189/-185; -175/-171) (67). It also binds to the consensus CAAG found at -120/-117 (-120 region) (74) and to a canonical consensus sequence at -104/-98 (88).

CTF/NF-1 binds to two regions in the γ-promoter: from -53 to -35 (SSE -60-) and from -140 to -118, overlapping the transcriptionally active CACCC element (123). It has been shown that CTF/NF-1 interferes with the synergistic transcriptional

activation of the γ-globin promoter mediated by GATA-1 and Sp1, throughout the competitive displacement of Sp1 from its target at -140 (123, 63).

The ubiquitous OTF-1 protein binds to the octamer element at position -182/-175 (217).

A peculiar characteristic of γ-globin promoters is a duplication of a 12-bp sequence containing the CCAAT box, a *cis* acting element of many eukaryotic gene promoters. A number of nuclear factors binds to overlapping sequences within the duplicated CCAAT boxes and the binding of CP1/NF-Y, CDP and NFE-3 have been better characterized.

CP1/NF-Y binds more tightly to the proximal (more similar to the β-globin CCAAT box) than to the distal γ-CCAAT (130).

CDP binds flanking sequences of both γ-CCAAT boxes containing the GCCTTGAC motif. It has been reported that CDP binding displaces CP1 from its target and, as a consequence, acts as a putative repressor (127, 130).

NFE-3 binds to the distal but not to the proximal CCAAT box (234, 246) (see later).

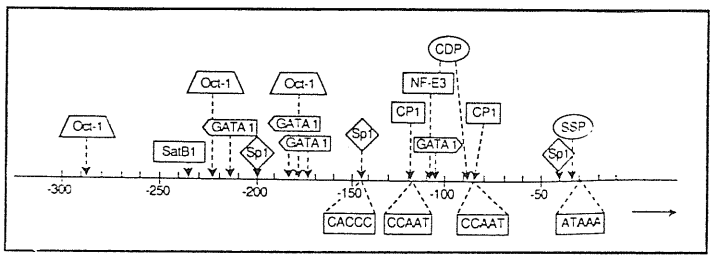

Figure 5. The γ-globin gene promoter. General disposition of the conserved "boxes" and position of binding motifs for various proteins are shown (3).

1.8 Hereditary Persistence of Fetal Hemoglobin (HPFH)

The Hereditary Persistence of Fetal Hemoglobin (HPFH) is a genetically inherited condition characterized by the constitutive expression of the fetal γ-globin genes in adult life. The adult expression of γ-globin has no adverse effects and, when a HPFH mutation is coinherited with a β-thalassemia allele, the presence of γ chains can partially compensate for the absence of β chains, giving rise to a much milder disease. Individuals heterozygous for one of the several types of HPFH express one or both the nonallelic ^{G}γ- and ^{A}γ-globin genes at levels exceeding by 15-200-fold the physiological one. The HbF may represent more than 30% of total hemoglobin in contrast to the normal adult level of less than 1% HbF.

HPFH alleles can be of two types.

Deletional HPFHs are caused by large deletions within the 3' end of the β-like globin cluster; they are almost invariably associated with increased expression of both γ genes (or the only $^{G}\gamma$-, if the $^{A}\gamma$-globin gene is eliminated by the deletions) in *cis* to the deletion.

Mutational HPFHs (or non-deletional HPFHs) are associated with a number of point mutations or small deletion within the promoter sequences of the $^{G}\gamma$- or $^{A}\gamma$-globin genes (Tab. I and Fig. 6). The casual relationship between the point mutations and the HPFH phenotype is strongly supported by genetic evidences. First, the single point mutations are constantly associated with HPFH phenotype. Second, the same mutations have been observed in HPFH cases of different ethnic origins.

Although the molecular mechanisms responsible for γ-globin protracted expression in mutational HPFH are not perfectly known, alterations in the recognition of *cis*-acting elements by regulatory factors have been invoked. The mutations might increase γ promoter strength, allowing the mutant promoter to interact with adult stage-specific *trans*-activating factors, or they might eliminate or inhibit binding of repressors (142, 240, 241).

TYPE AND RACIAL GROUP	MUTATION IN GLOBIN GENE	% Hb F in HETEROZYGOTES
$^{G}\gamma$ HPFH		
Japanese	$^{G}\gamma$–114 C to T	11–14
Black	$^{G}\gamma$–175 T to C	20–30
Sardinian	$^{G}\gamma$–175 T to C	17–21
Black	$^{G}\gamma$–202 C to G	15–25
$^{A}\gamma$ HPFH		
Georgia	$^{A}\gamma$–114 C to T	3–6.5
Black	$^{A}\gamma$–114 to –102 deleted	30–32
Greek	$^{A}\gamma$–117 G to A	10–20
Sardinian	$^{A}\gamma$–117 G to A	12–16
Black	$^{A}\gamma$–117 G to A	11–16
Black	$^{A}\gamma$–175 T to C	36–41
Brazilian	$^{A}\gamma$–195 C to G	4.5–7
Southern Italian	$^{A}\gamma$–196 C to T	12–16
Chinese	$^{A}\gamma$–196 C to T	14–21
British	$^{A}\gamma$–198 T to C	3.5–10
Black	$^{A}\gamma$–202 C to T	1.6–3.9

Table I. Non-deletional forms of Hereditary Persistence of Fetal Hemoglobin (HPFH). (3).

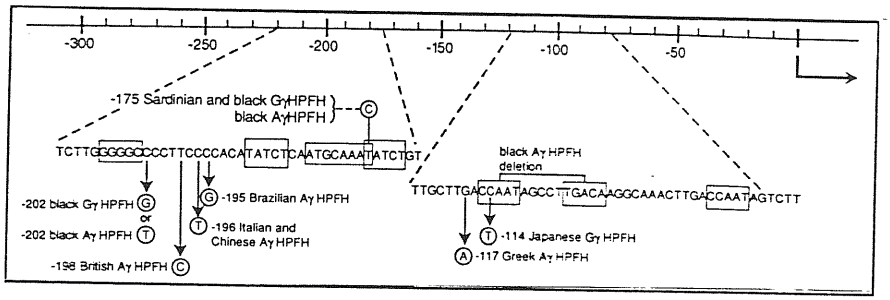

Figure 6. Point mutations in the γ-globin gene promoter that cause Hereditary Persistence of Fetal Hemoglobin (HPFH). (3).

1.9 Mutational HPFH

1.9.1 -200 region HPFHs

The region between positions -141 and -202 of the $^{a}\gamma$- and $^{g}\gamma$-promoters is essential for their expression in transgenic mice (218), and it is the site of several HPFH mutations: a -175 T->C in Blacks $^{g}\gamma$ (219), in Blacks $^{a}\gamma$ (220), and in Sardinian $^{g}\gamma$ (221) HPFHs; a -195 C->G in Brazilian $^{a}\gamma$ HPFH (222); a -196 C->T in Italian (223) and Chinese (224) $^{a}\gamma$ HPFH; a -198 T->C in British $^{a}\gamma$ HPFH (225); a -202 C->T and C->G in African $^{a}\gamma$ and $^{g}\gamma$ HPFH, respectively (227-229).

As already mentioned, a variety of transcription factors binds to this region. Among them, binding of GATA-1, Sp1 and OTF-1 has been shown to be critically affected.

1.9.1.1 -198 $^{a}\gamma$ HPFH

-198 $^{a}\gamma$ is the only HPFH mutation that creates a significantly stronger Sp1 binding site and a 4-5-fold increased activity in transient transfection experiments (63, 230, 231), in sharp contrast to -202 and -196 HPFH mutations.

In the -198 HPFH promoter, OTF-1 canonical binding is apparently displaced by Sp1 and also GATA-1 shows an unusual binding, protecting from position -172 to -211, instead of the normal -195 to -172 region (63) (Fig. 7). However, the apparent erythroid-specificity of -198 HPFH mutation in transfected cells suggests that the molecular mechanisms acting *in vivo* are probably more complex than the simple augmentation of the binding of ubiquitous transcription factors observed *in vitro*.

1.9.1.2

Much more contrasting data concern the **-202 C->G $^{g}\gamma$ HPFH** (228, 229) and the **-202 C->T $^{a}\gamma$ HPFH** (227) mutations. Some authors have reported an increased Sp1 binding to these sites (230, 232), while others have not (63, 231).

1.9.1.3 -175 HPFH

It has been proposed that the -175 T->C HPFH mutation, that induces a 4-fold increased activity of the γ-globin promoter in transfection assays in K562 cells (233), "increases" (234) or "alters" (67) the binding of GATA-1 to its distal site in the -195 to -170 region (Fig. 7). It is unclear whether this change is relevant to the *in vivo* HPFH phenotype, but the transfection data do suggest that GATA-1 binding to the distal site has a positive, physiological effect on γ-globin promoter activity.

The same mutation abolishes OTF-1 binding *in vitro* (234, 235) but inactivation of its binding site does not cause overexpression of the γ-globin gene in transfection

experiments (233), weakening the possibility that it might be really involved in the -175 HPFH phenotype.

1.9.2 dCCAAT box HPFHs

1.9.2.1 -114 HPFH

The -114 C->T mutation has been associated with increased adult $^{a}\gamma$ expression (3%-5%) in a Georgian family (236) and with increased $^{g}\gamma$ expression in a Japanese HPFH (237). The -114 transition occurs within the distal CCAAT box of the γ-gene promoters and abolishes binding of CP1 and NFE-3, without affecting CDP and GATA-1 binding affinity (237, 238) (Fig. 7). In transgenic mice the -114 HPFH mutation results both in a significant increase in γ-globin expression as well as in a stronger transcriptional competition of γ- toward β-globin gene (239).

1.9.2.2 -117 HPFH

The G->A substitution at position -117, two bases upstream of the distal CCAAT box of the $^{a}\gamma$ gene, is known as the Greek $^{a}\gamma$ form of HPFH (240, 241). Individuals heterozygous for this mutation produce from 10% to 20% $^{a}\gamma$-HbF in adulthood and Greek HPFH mutation is sufficient to produce HPFH phenotype in transgenic mice, resulting in persistence of γ-globin expression at high level and a concomitant decrease in β-globin expression in fetuses and adults (121). GATA-1, CP1/NF-Y, CDP and NFE-3, all contact the G residue mutated in the Greek HPFH (130). Methylation interference assay using affinity-purified proteins shows that the G at -117 is crucial for NFE-3 as well as for GATA-1 binding and the substitution for this G in A results in a significantly decreased of their specific binding (130, 238, 242).

The -117 substitution occurs in a TGA trinucleotide within a sequence that is repeated three times at intervals of 13 and 14 bp in the duplicated CCAAT box regions (Fig. 7). This arrangement resembles the interaction site for the sea urchin CCAAT displacement protein (CDP) in the sperm H2B promoter (129) and the G->A susbstitution in the first 5' TGA triplet leads to a two-fold increase of CDP binding affinity (130).

The mutated sequence (GACCAAT to AACCAAT) exhibits also a stronger similarity to the canonical high-affinity CP1/NF-Y binding site (AACCAATT) and this is probably the reason why also CP1/NF-Y binds the -117 HPFH distal CCAAT box with higher affinity than the wild type box (130, 234, 238, 243, 244).

The -117 HPFH phenotypic changes correlate with the loss of binding of the transcription factor GATA-1 to the -120 region of γ-globin promoter, suggesting that it may also act as a negative regulator of the γ-globin gene in adults (130). Moreover,

-117 G->A point mutation is clearly associated with almost complete disappearance of typical NFE-3 *in vitro* binding to the distal γ-CCAAT box (130, 234).
The ability to completely override the adult suppression of the γ-globin gene could result from the loss of bound suppressor/s (GATA-1, NFE-3), the gain of bound activator/s (CP1/NF-Y), or modification in CCAAT boxes reciprocal interaction, so that inactivation of the distal box leads to more effective use of the proximal one.

The behaviour of a 13-bp deletional HPFH, that lacks the distal γ-CCAAT box, strongly favours the first possibility and highlights the critical role played by NFE-3 in regulating γ-globin gene expression (see later).

1.9.2.3 Δ13 deletion

Δ13 indicates a thirteen-base pairs deletion in the $^{a}\gamma$ gene distal CCAAT region described in two black adolescents with $^{a}\gamma$-β+/HPFH and sickle cell trait. In these individuals the expression of the fetal gene exceeds 100-fold the normal levels. It is usually referred as -115 to -103 deletion, but the 5' end of this deletion is ambiguous: it could be anywhere, between -122 and -114 (245).

The deletion removes: one of the TGA repeats thought to be essential for CDP binding, the third GCCTT repeat at -109/-105 and the distal CCAAT sequence (-115 to -111) (Fig. 7). This deletion reduces the *in vitro* binding of GATA-1 by about 50% and abolishes binding of CP1, CDP and NFE-3 (245, 246).

In the -117 HPFH γ-promoter GATA-1 and NFE-3 binding is dramatically affected, while CDP and CP1/NF-Y binding is slightly increased. CDP, CP1/NF-Y and NFE-3 binding is also abolished by the Δ13 HPFH deletion, while GATA-1 binding is weakened. Though the two mutations are located in the same region, it is possible that the molecular mechanisms underlying the HPFH phenotype are heterogeneous. On the other hand, if similar mechanisms are responsible for the two types of HPFH, the loss of NFE-3 binding is the only common alteration in both the -117 and Δ13 HPFH, suggesting a negatively acting role for this factor.

It has been demonstrated that the abrogation of GATA-1 binding site by a G->A point mutation at -104 results in a substantial decrease of γ-globin promoter activity. Furthermore, this mutation abolishes the increased activity of the -175 HPFH mutant, showing that the proximal GATA-1 site is the major GATA-like determinant of γ-globin gene erythroid-specific expression (242). This result has important implications for understanding the -117 and Δ13 HPFH. It weakens the possibility that GATA-1 might behave as the only and crucial negatively acting factor when bound to the mutated sequences, and indirectly gives relevance to the role of NFE-3.

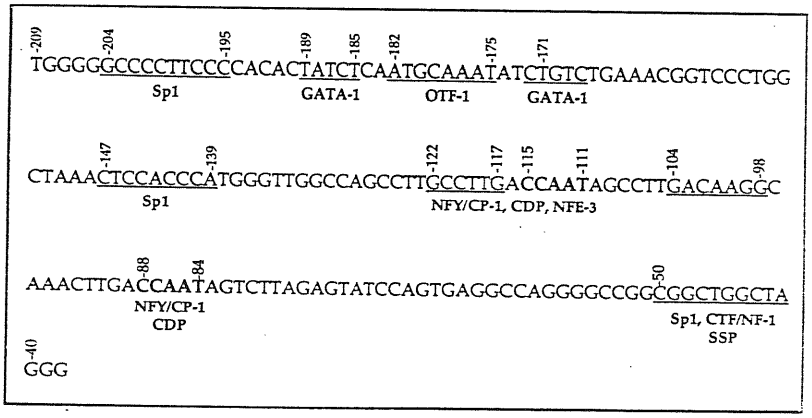

Figure 7. The γ-globin gene promoter. Nucleotide sequence of the human γ-globin promoter (from -209 to -38). Sequences corresponding to the proximal and distal CCAAT box are indicated in bold. Consensus binding sites for *trans*-acting factors are underlined.

CHAPTER 2

MATERIALS AND METHODS

2.1 Cell culture

Human erythroleukemic K562 cells (kindly provided by Dr. M.E. Lopez) were grown at 37°C in suspension to a density of $5x10^5$ cells/ml in RPMI 1640 medium supplemented with 2 mM L-glutamine, 10% fetal calf serum and 5 μg/ml gentamycin.

2.2 Whole cell extract preparation and protein purification

The protein purification was carried out by consecutive extractions of approximately 20 g of frozen cell pellets. K562 whole cell extracts (WCEs) were prepared accordingly to the method of Manley et al. (247) with sligth modifications. The salt precipitated proteins were dissolved in TM buffer (50 mM Tris-HCl pH 7.9, 12.5 mM MgCl2, 10 μM ZnCl2, 1 mM EDTA, 10% glycerol) or TM* buffer (as TM, plus 0.009% Tween20), dialyzed at 4°C against TM 0.1 M NaCl, 2mM DTT (TM 0.1M), or TM 0.1 M NaCl, 2 mM DTT, 0.009% Tween20 (TM* 0.1M) for a total of 12 hr, with one buffer change, and loaded onto a Heparin-Sepharose column, equilibrated with TM 0.1 M or TM* 0.1 M. Bound proteins were step-eluted with TM or TM* buffer with 0.4, 0.6 or 1 M NaCl. The elution of NFE-3 from the Heparin-Sepharose and subsequent columns was checked by electrophoretic mobility shift assay (EMSA). The fractions eluted at 0.6 M salt were pooled, and the proteins were precipitated with ammonium sulfate to 60% of saturation. Precipitated proteins were dissolved in TM or TM*, dialyzed at 4°C against TM 0.5M or TM* 0.5M for a total of 12 hr, with one buffer change and applied to a Sdx-200 column equilibrated in TM 0.5M or TM* 0.5M. Active fractions eluted from gel filtration column were pooled, precipitated and dialyzed against TM 0.2M or TM* 0.2M. Sample was mixed with various amount of polydIdC and calf thymus DNA (see results), cleared by centrifugation and/or filtration through 0.2 μm filters, and the supernatant was loaded onto a NHS-activated Superose column coupled to γdCCAAT, εCCAAT or γ42-36 concatamerized ONs. All the ONs used are listed in Tab. II. The affinity chromatography was performed on a SMART System, and the bound proteins were eluted either with a linear 0.1-1 M NaCl gradient or with a single step at 1 M NaCl.

Tab. II: Sequences of oligonucleotides utilized in electrophoretic mobility shift assays. Nucleotide positions are relative to the CAP site. Bases within the CCAAT boxes are indicated in boldface; bases added at the extremities are indicated in italics.

γdCCAAT	-127	*GATC*GCCTTGCCTTGA**CCAAT**AGCCTTGACA	-101
γ42-36	-132	*CC*GGCCAGCCTTGCCTTGA**CCAAT**AGCCTTGACAAGGCAA	-95
γ42	-132	GGCCAGCCTTGCCTTGA**CCAAT**AGCCTTGACAAGGCAAACTT	-91
εCCAAT	-96	*GATC*GTCAGCCTTGA**CCAAT**GACTTTTA	-73
-117 HPFHo	-127	GCCTTGCCTTAA**CCAAT**AGCCTT	-105
-117 HPFHn	-127	*GATC*GCCTTGCCTTAA**CCAAT**AGCCTTGACA	-101
-114 HPFH	-132	GGCCAGCCTTGCCTTGA**CTAAT**AGCCTTGACAAGGCAAACTT	-91
-110 HPFH	-132	GGCCAGCCTTGCCTTGA**CCAAT**CGCCTTGACAAGGCAAACTT	-91
Δ13	-132	GGCCAGCCTTGCCTTG----------ACAAGGCAAACTT	-91
biotynil. εCCAAT	-96	*GTCAGCCTTGA**CCAAT**GACTTTTA	-73
γpCCAAT	-99	GGCAAACTTGA**CCAAT**AGTCTTAGA	-75
MHCII (5'->3' AS) (CP1)	-44	*GATC*CTTTTAA**CCAAT**CAGAAAA	-63
Ad2ori	-22	TTTTGGATTGAAG**CCAAT**ATGATC	-45
NF-Y-box alb. (NF-Y)		*TGAAAC*GGGTAGGAA**CCAAT**GAAATGAAATGAAAGCA	
YY1 P5	-60	GTTTTGCGACATTTTGCGACAC	-39
γ-50 region	-59	AGGGGCCGGCGGCTGGCTAGGATGA	-34
γ-198 Sp1	-211	CTTGGGGGCCCCTCCCCCACACTATCTCAATG	-180
Sp1 (SV40)		GATCCCCCGCCCC	
mαglobin (5'->3' AS)	-167	*GATC*CGGGCAACTGATAAGGATTCCCA	-189
octa H2B histone	-58	CTTCACCTTATTTGCATAAGC	-38

Mono Q anionic exchange chromatography (performed on a SMART System) was used to concentrate protein samples at any stage of purification, as a valid alternative to the ammonium sulfate precipitation. Usually the column was equilibrated in TM 0.1M NaCl or TM* 0.1M NaCl and elution was carried out with a linear 0.2-1 M NaCl gradient. All chromatographies were run at 4°C.

2.3 Preparation of DNA-affinity columns

γdCCAAT, εCCAAT and γ42-36 ONs were annealed, phosphorylated and ligated essentially as described by Kadonaga et al. (248). The ligation product was coupled to NHS-activated Superose column (Pharmacia) accordingly to the manufacturer's instructions.

2.4 YY1 purification/deprivation

Partially purified NFE-3 preparations in which YY1 activity was also detected were pooled and further subjected to Ni^{2+}-NTA chromatography, performed on a SMART System at 4°C. The proteins were dialyzed at 4°C against buffer A (20 mM Hepes pH 7.9, 12% glycerol, 100 mM KCl, 0.1% Tween20, 30 mM imidazole) for a total of 12 hr and loaded onto a Ni^{2+}-charged Hi Trap column (Pharmacia), equilibrated with the same buffer. Bound proteins were eluted with buffer A plus 250 mM imidazole; binding activity of eluted fractions was checked by EMSA.

2.5 Magnetic beads DNA-affinity purification

Twenty milligrams of Streptavidin magnetic particles were transferred to a 15 ml Falcon tube and washed extensively by resuspension and magnetic separation: six times with TE buffer (10 mM Tris-HCl pH 7.9, 1 mM EDTA), once with 1 M NaCl, once with 2 M NaCl and six times with TE buffer. After removal of the last wash solution, the beads were resuspended in 1 M NaCl containing 200 μg of double stranded (ds) 5'-biotynilated εCCAAT ON. The suspension was gently rocked in a roller for 12 hr at room temperature and then washed three times with TE buffer. The efficiency of coupling was estimated by reading the OD_{260} of the solution before and after coupling. Protein samples were pooled, diluted to 100-150 mM NaCl with TM buffer and added to the DNA-coupled magnetic particles suspension. The reaction tube was placed on ice for 15 min, mixed by gentle flicking every 5 min to prevent sedimentation and the supernatant recovered after magnetic separation. Beads were washed three times with TM buffer; bound proteins were step eluted with TM buffer at a final concentration of 0.5 M, 1 M and 2 M NaCl, assayed in EMSA and visualized by silver staining (249).

2.6 Electrophoretic mobility shift assay (EMSA)

Fifty nanograms of sense strand ONs were 5' end labeled with γ^{32}P-ATP and T4 polynucleotide kinase at 37°C for 45 min in 50 mM Tris-HCl pH 7.5, 10 mM MgCl2, 5 mM DTT, 1 mM spermidine, 1 mM EDTA. The mixture was then heated at 95°C for 5 min; 55 ng of complementary ONs were added, and the sample heated at 95°C for 5 min, slow cooling to room temperature and separated from unincorporated γ^{32}P-ATP by gel filtration through Sephadex G-50 column. EMSA standard binding reaction (10 µl) contained 0.1-0.3 ng of labeled ds ON (approximately 50.000 cpm), 2 µg of bovine serum albumin (BSA), 4 mM spermidine, 150-200 mM NaCl, 1 mM EDTA, 10 mM Tris-HCl pH 7.9, 2 mM DTT and 0.009% Tween20, when indicated. For YY1 *in vitro* binding, NaCl concentration was lowered to 50 mM. When crude extract proteins were used (2-5 µg), 1 µg of polydIdC/10 µl reaction was added and competition performed with 100-fold molar excess of cold ds ONs. When affinity-purified fractions were used, polydIdC was added at a concentration of 10-500 ng/10 µl reaction or omitted at all, and unlabeled competitors were usually added in 50-fold molar excess. Apart from few experiments, competition and supershift assays were performed by mixing competitor DNA or specific antibodies (or phosphate-buffered saline - PBS -) with all the other components both before and after the addition of the protein to be tested. EMSA mixtures were kept on ice 30 min; 2 µl of loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF, 10% ficoll) were added and samples were loaded onto a 5% 29:1 acrylamide/bisacrylamide non denaturing gel; electrophoresis was carried out in 50 mM Tris borate pH 8.2 at 200 V for 2.5 hr at 4°C. Gels were vaccum dried and exposed to X-ray films.

2.7 NFE-3 on-off rate analysis

EMSA mixture was scaled up to 100 µl and incubated on ice for 30 min with labeled εCCAAT ON and without polydIdC; 100-fold excess of unlabeled competitor ON was then added. After 0 time, 2, 5, 15 and 30 min, 10 µl aliquots were loaded onto a continuously running non-denaturing gel; electrophoresis was carried out exactly as reported in 2.6.

2.8 NFE-3 affinity study

NFE-3 affinity for wild type and mutated CCAAT consensus sequences was estimated in EMSA using the ON for which the affinity had to be determined (usually εCCAAT and γdCCAAT) as labeled probe, and varying the amounts of competitor over at least a 100-fold range with 1.5- to 2-fold serial dilutions. Binding condition and experimental procedure were exactly as reported in 2.6.

2.9 UV-crosslinking

2.9.1 UV-crosslinking to end-labeled probe: affinity-purified fractions were incubated with ^{32}P-end labeled γd- and ε-CCAAT ONs under gel shift conditions, for 30 min on ice. After binding, one fifth of the reactions was loaded onto a non denaturing gel and the rest of the mixtures was transferred to a 60-well microtiter and irradiated for 30 min at 4°C using a 254 nm UV lamp at a distance of 3 cm. Samples were then mixed with SDS sample buffer (5% β-mercaptoethanol, 60 mM Tris-HCl pH 6.8, 1% SDS, 10% glycerol, 4 mM EDTA), heated for 5 min at 90 °C and electrophoresed on 10% or 12.5% SDS-PAGE. Gels were fixed in 50% methanol, 7.5% acetic acid for 15 min, dried and exposed to X-ray films.

2.9.2 UV-crosslinking to uniformly labeled probe: partially purified fractions were incubated with ^{32}P-uniformly labeled γ42 ON probe. γ42 ON was body labeled by annealing a 18 bp ON (3'-GGAACTGTTCCGTTTGAA-5') and filling in with Klenow fragment and α^{32}P-dCTP/-dTTP in 10 mM Tris-HCl pH 7.5, 5 mM MgCl2, 7.5 mM DTT. Labeled probe was purified by gel filtration through Sephadex G-50 column and crosslinking reactions performed as above. After UV irradiation each sample was transferred to eppendorf tube. One volume of solution 1 (0.4% sarcosyl, 2 mM PMSF) was added and samples mixed and briefly cenrifuged. One volume of solution 2 (1% NP-40, 10 mM MgCl2, 5 mM CaCl2, 20 mM Tris-HCl pH 8.0) was further added and samples digested with 9.4 U of DNase I and 3 U of micrococcal nuclease for 30 min at 37°C. Samples were precipitated with 20% Trichloroacetic acid (TCA), resuspended in SDS sample buffer and analysed as above.

2.10 Southwestern blot

Partially purified samples were fractionated on 12.5% SDS-PAGE and transferred to nitrocellulose membrane in 20 mM Na2HPO4/NaH2PO4 pH 7.0 buffer, for 12 hr, at 8 V at 4°C. Membrane was renatured and blocked for 8 hr with two changes (3 ml/each) of buffer A (40 mM Tris-HCl pH 7.6, 50 mM NaCl, 10% glycerol, 2.5% NP-40, 5% BSA, 0.002% NaN3, 10 μg/ml herring sperm DNA, 0.02 mM ZnCl2, 1 mM EDTA pH 8.0, 2 mM PMSF) by gentle rocking at 4°C. After renaturation, filter was washed in 30 ml of buffer B (as buffer A, but without glycerol and NP-40 and with 0.25% BSA), rocking at 4°C for 20 min. Filter was then probed with about 10^6 cpm/ml of ^{32}P-end labeled and concatamerized εCCAAT ON, rocking at 4°C for 12 hr. Nitrocellulose membrane was washed two times, 30 min/each with buffer B and exposed to X-ray films. Competition was performed by addition of 100-fold molar excess of competitor DNA to the labeled probe.

2.11 DNA-binding protein renaturation

Affinity-purified samples were precipitated in 20% TCA and resuspended in SDS sample buffer, heated for 15 min at 45 °C and fractionated on a preparative 12.5% SDS-PAGE. Lane corresponding to NFE-3 sample was excised and cut out in four different areas (using standard protein molecular weights as reference); each area was transferred into a dialysis tube and electroeluted in SDS-PAGE running buffer for 3 hr at 50 V. After precipitation with 20% TCA, pellet was washed 4-5 times with cold acetone. The resulting pellet was resuspended in 50 µl 6 M guanidinium hydrochloride (GdHCl) in TM buffer (without glycerol and DTT) and incubated on ice for at least 30 min. GdHCl was removed by gel filtration chromatography (Fast desalting on SMART System) and samples were left 16 hr at 4°C to allow their slow refolding. Column equilibration and elution was done in TM0.1 M NaCl, 2mM DTT. Each fraction was tested in EMSA.

2.12 Sample preparation for mass-mass spectrometry analysis

Affinity-purified material was concentrated by a Mono Q chromatographic step. Active fractions were precipitated with 20% TCA or directly mixed with SDS sample buffer, heated 5 min at 90 °C and loaded onto a 10-12.5% SDS-PAGE (prepared with ultra-pure reagents). At the end of electrophoresis, gel was stained with Coomassie blue (0.1% Coomassie blue R-250, 45% methanol, 10% acetic acid) for 5 min and then destained (10% methanol, 10% acetic acid) for 15-20 min. The specific bands were excised out and each gel slice was transferred into an 1.5 ml tube containing 1 ml of MilliQ water, frozen in dry ice-ethanol bath and sent to M. Mann - Protein Peptide group - at the EMBL.

2.13 Western immunoblot analysis

Proteins were separated by electrophoresis on 10-12.5% SDS-PAGE and electroblotted to nitrocellulose for 48 min at 20 V in 25 mM Tris base, 192 mM glycine, 20% methanol, 0.1% SDS. When Phast System transblot apparatus (Pharmacia) was used, manufacturer's instructions were followed. Membranes were stained with 0.1% Ponceau red in 1% acetic acid, washed in PBS, blocked in 5% milk-TBST (5% w/v non-fat dried milk in TBST buffer: 10 mM Tris-HCl pH 7.9, 150 mM NaCl, 0.05% Tween20) for 1 hr at room temperature and incubated for 60-90 min at room temperature (or overnight at 4°C) with antisera or polyclonal antibodies in 5% milk-TBST. Antibodies anti-Ku were kindly provided by the A. Falaschi's and S. Jackson's labs; antibodies anti-CP1/NF-Y were a generous gift of S. Ottolenghi's lab. Antibodies anti-CTF and anti-murine GATA-1 were made in our laboratory, while all the others (anti-YY1, anti-human GATA-1) were commercially supplied (Santa Cruz).

Filters were incubated with anti-rabbit or anti-mouse alkaline-phosphatase-conjugated antibodies for 45 min at room temperature. Bound antibodies were revealed with AP buffer (100 mM Tris-HCl pH 9, 100 mM NaCl, 5 mM $MgCl_2$) containing BCIP (5-bromo-4-chloro-3-indolyl phosphate) and NTB (Nitro blue tatrazolium) reagents. When UV-crosslinked proteins were subjected to immunoblot analysis, filters were also dried and exposed to X-ray films.

CHAPTER 3

RESULTS: Part I

The first purification procedure is detailed in Fig. 31.
Typical NFE-3 *in vitro* binding activities were observed in crude K562 WCEs (3.1, 3.2) and in partially purified preparations that means samples eluted both from Heparin-Sepharose (3.3.1, 3.6) and Sdx-200 (3.3.2, 3.4, 3.5) columns. These activities were indicated as "band 1, band 2 and band 3", beginning from the slower migrating to the faster migrating one.

Earliest approaches to perform DNA-affinity purification led, instead, to a variety of problems. In particular: i) low recovery of NFE-3 activity (3.3.3) ii) presence of NFE-3 activity in flowthrough sample (3.3.3).
This behaviour was considered the direct consequence of: i) fractionation of NFE-3 components (complementation in 3.7) ii) high off ratio of NFE-3 binding (3.8) iii) partial inhibition of NFE-3 binding to the affinity column by YY1 (3.9); this protein in fact copurified with NFE-3.

In the attempt to circumvent these problems: i) YY1 protein was removed from the NFE-3 active samples before subjected them to affinity purification (3.9) ii) a low amount of polydIdC was used during the affinity step (3.3.3, 3.3.5) iii) DNA-affinity chromatography was replaced by DNA-coated magnetic beads purification (3.3.5).

The latter two modifications led to a better recovery of DNA-binding activities, but provoked both a variation in mobility shift and the loss of typical NFE-3 binding specificities.

For these reasons the *in vitro* binding activity obtained after DNA-coated magnetic beads purification was indicated as "affinity-purified activity" and no more as NFE-3 protein/s or NFE-3 activity (3.3.5, 3.10, 3.11, 3.12).

The activity enriched during the DNA-coated magnetic beads purification turned out to be due to contaminant Ku heterodimer (3.13, 3.14).

3.1 NFE-3 Binds to the distal CCAAT Box of the Human γ-globin Promoter

To better characterize the nature of the binding activity known as NFE-3, whole cell extracts (WCEs) from the human erythroleukemic cell line K562 were tested in electrophoretic mobility shift assay (EMSA) with a labeled oligonucleotide (ON) spanning the distal CCAAT box region of the human γ-globin promoter (γdCCAAT) and 1 μg of polydIdC as aspecific competitor.

A reproducible pattern of shifted complexes was observed as shown in Fig. 8 (lane 1, complexes 1-6). To assess their specificity, the binding was challenged by competition with 100-fold molar excess of several ONs (all the ONs used are listed in Tab. II).

The retarded band 1 is known to be due to CP1/NF-Y, the major factor interacting with the human γ-globin distal CCAAT box (130, 234, and Fig. 8 and 9). In terms of binding specificity, CP1/NF-Y can be distinguished from NFE-3 by its ability to bind to both the HPFH -117 mutant and the wild type sequence of the γ-globin distal CCAAT box, whereas NFE-3 binds only to the latter (234, 246).

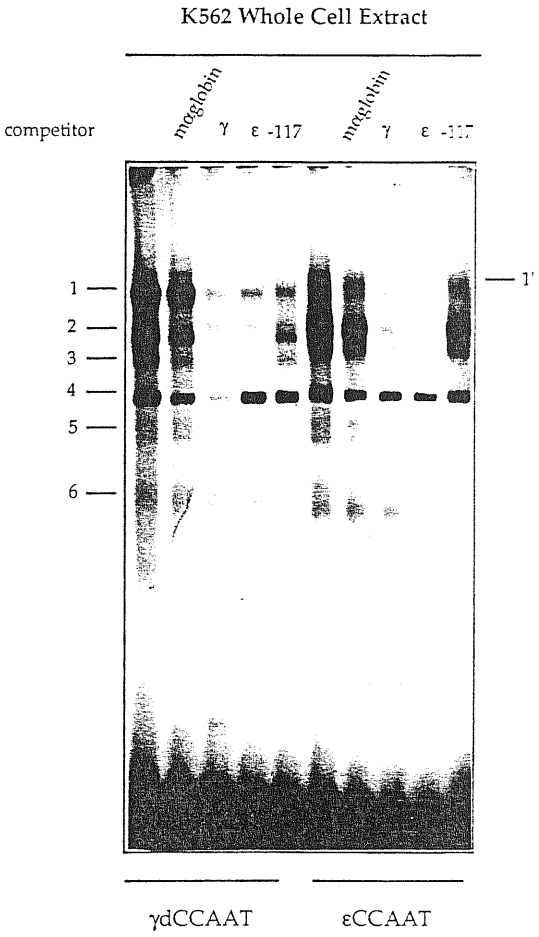

Fig. 8. Binding specificity of crude K562 extracts to ε- and γ-dCCAAT boxes. K562 whole cell extract was incubated with either γdCCAAT or εCCAAT labeled ONs, in the presence of 1 μg of polydIdC. Binding reactions were carried out in the absence (lane 1 and 6) or presence of 100-fold molar excess of competitors, as reported on the top of the figure. The position of complexes 1-6 is indicated on the left of the figure.

In Fig. 9 a heparin fraction enriched for CP1/NF-Y is analysed by EMSA with labeled γdCCAAT probe. CP1/NF-Y binds the γdCCAAT (lane 1) and is competed out by unlabeled γdCCAAT (γ, lane 2) or by -117 HPFH (-117, lane 4) ONs. In contrast, the unlabeled human ε-globin promoter CCAAT box (εCCAAT ON, ε, lane 3) is not bound by CP1/NF-Y (166).

CP1/NF-Y in crude K562 WCEs was also competed out by both γdCCAAT (Fig. 8, γ, lane 3) and -117HPFH ONs (Fig. 8, -117, lane 5).

Anyway, competition with -117HPFH revealed another band comigrating with, and weaker than, CP1/NF-Y. As detailed later on, it corresponds to a newly identified NFE-3 binding activity.

In Fig. 8, bands 2 and 3, that for electrophoretic migration resemble the NFE-3 complexes, were indeed specifically competed by either γdCCAAT (γ, lane 3) and εCCAAT (ε, lane 4), while no competition or a clearly weaker competition was observed when mαglobin (GATA-1 consensus binding site from the murine α1 globin gene promoter) (lane 2) and -117 HPFH (-117, lane 5) ONs, respectively, were used. These results are in agreement with the criteria originally adopted to identify NFE-3 activity in crude extract (234, 246).

Band 4 was only competed by the wild type γdCCAAT (γ, lane 3), while band 5 was competed by all the ONs tested (γ, ε, -117, lanes 3-5), with the exception of mαglobin (lane 2), suggesting that this complex might be due to a general CCAAT-binding protein. Band 6 appears to be non-specific, since it was neither competed by any of the tested ONs (mαglobin, γ, ε, -117, lanes 2-5).

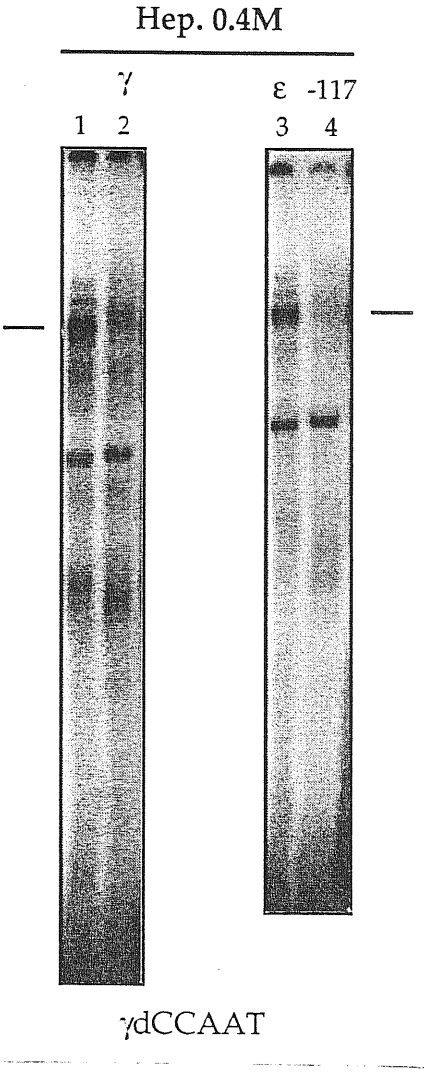

Fig. 9. DNA binding specificity of CP1/NF-Y. Partially purified CP1/NF-Y eluted from Heparin-Sepharose at 0.4 M salt was assayed in mobility shift with labeled γdCCAAT (lane 1). The specificity of binding was tested by competition experiment with 100-fold molar excess of γdCCAAT (γ, lane 2), εCCAAT (ε, lane 3) and -117 HPFH (-117, lane 4) ONs. CP1/NF-Y retarded complexes are indicated by bars on both sides of the figure.

3.2 NFE-3 Binds to the Single CCAAT box of the Human ε-globin Promoter

On the basis of binding specificity, NFE-3 belongs to the heterogeneous class of CCAAT box-binding proteins. Among the tested human globin gene CCAAT boxes, only the single ε-globin CCAAT box is significantly bound by NFE-3 and NFE-3 represents the major ε-globin CCAAT box-binding protein (238). A reproducible pattern of retarded bands was observed when crude K562 WCE was incubated with εCCAAT labeled probe, as shown in Fig. 8. The slower migrating band (named 1') comigrates with the CP1/NF-Y-γdCCAAT complex; the other two strong and broad bands partially overlap NFE-3-γdCCAAT complexes (2 and 3). In competition assay, all three retarded bands behave exactly as expected for NFE-3:

they were competed out by both γdCCAAT (γ, lane 8) and εCCAAT (ε, lane 9), while neither mαglobin (lane 7) nor -117 HPFH (-117, lane 10) ONs were able to abolish the binding. These data strongly suggest that: i) NFE-3 can efficiently bind both ε- and γd-CCAAT boxes; ii) γdCCAAT -117 mutation associated with HPFH abolishes NFE-3 binding; iii) the retarded band 1 (with labeled γdCCAAT) or 1' (with labeled εCCAAT) can not be exclusively due to CP1/NF-Y, but very likely also to a comigrating NFE-3 complex. On the basis of these discriminatory behaviours, the purification procedure was scaled up to accumulate enough NFE-3 protein suitable for peptide microsequencing or mass spectrometry analysis.

3.3 Purification of NFE-3 from Human Erythroid K562 Cells

3.3.1 Heparin-Sepharose Chromatography

NFE-3 activity is detectable only in mouse (MEL) and human (K562) erythroleukemia cell WCEs (data not shown). K562 cells were chosen as starting material in consideration of their human origin and their stronger NFE-3 activity in mobility shift assay. K562 cells were first processed by conventional chromatography to fractionate NFE-3 from CP1/NF-Y. This was achieved by salt step elution of proteins bound to Heparin-Sepharose. While NFE-3 activity eluted in the 0.6 M fractions, CP1/NF-Y eluted exclusively in the 0.4 M fractions, as demonstrated by both gel retardation and western blotting assays. A typical EMSA of this purification step is shown in Fig. 10. Once again, the specificity of binding was assessed by competition experiments. Binding of NFE-3 to both the γd- and ε-CCAAT probes (lane 1 and 6, bands 1', 2/2', 3/3') was specifically competed by 100-fold molar excess of either γd- (γ, lane 3 and 8) and ε-CCAAT (ε, lane 4 and 9) unlabeled DNA, while both mαglobin (lane 2 and 7) and -117 HPFH (-117, lane 5 and 10) ONs did not.

The novel NFE-3 binding activity, named 1 or 1', as well as the previously reported ones (234, 246), here indicated as 2 and 3 (when assayed with labeled γdCCAAT) or 2' and 3' (when assayed when εCCAAT probe), were all enriched by this chromatographic step. They showed stronger binding affinity for εCCAAT than for the wild type γdCCAAT box, suggesting that the εCCAAT box is a better site for NFE-3. NFE-3 complexes bound to εCCAAT migrate slightly slower than when bound to the γdCCAAT ON (compare lane 1 and 6).

Western blot analyses confirmed the DNA-binding results (Fig. 11): both subunits of CP1/NF-Y (NF-YA of 42 KDa and NF-YB of 32 KDa) (128) were eluted from Heparin-Sepharose column at 0.4 M NaCl (Panel on the left: 0.4M, lane 2; Panel on the right: 0.4M, lane 3), while at 0.6 M (Panel on the left: 0.6M, lane 1; Panel on the

right: 0.6M, lane 4) neither of them were detected, even though a larger amount of protein was tested.

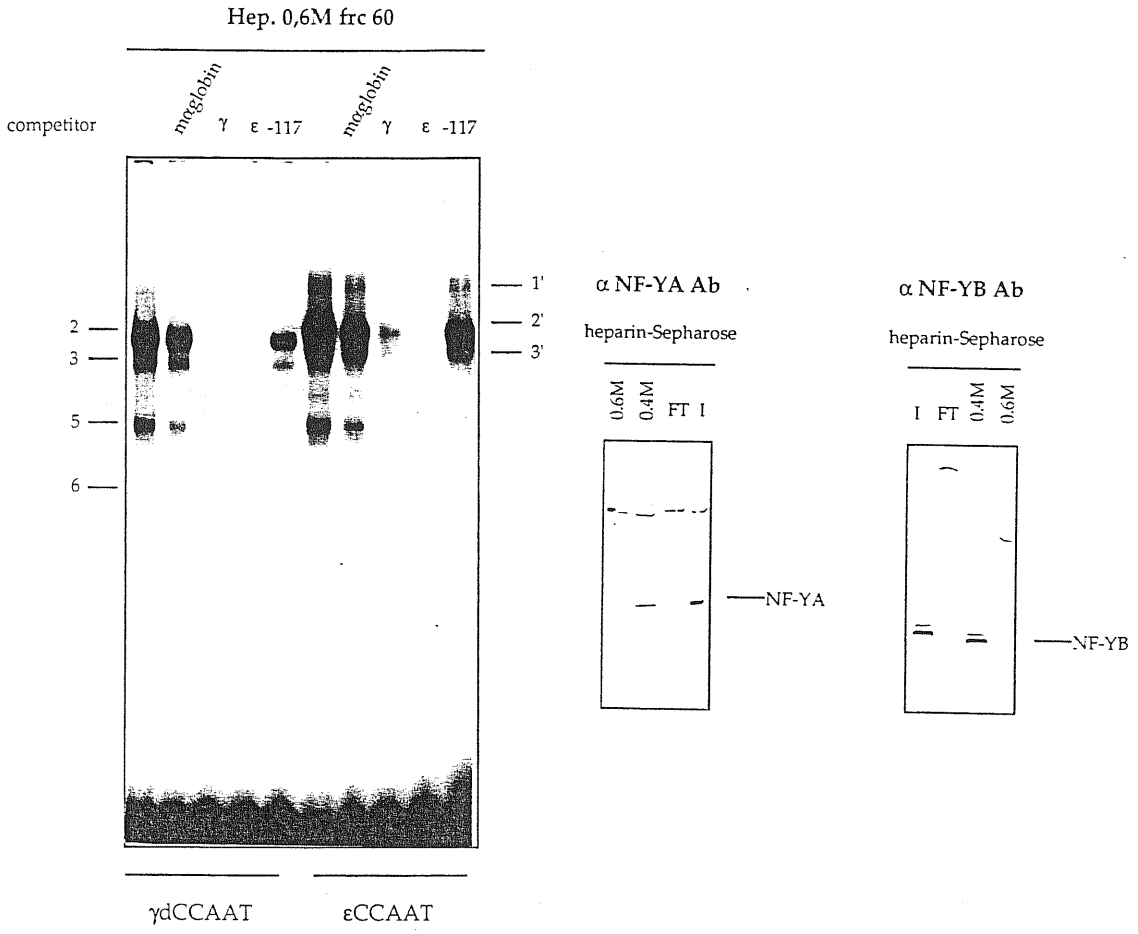

Fig. 10. EMSA: Fractionation of K562 whole cell extract by Heparin-Sepharose chromatography. Gel retardation assay with fractions eluted at 0.6 M NaCl from Heparin-Sepharose column; samples were incubated with either γdCCAAT or εCCAAT labeled ONs and 1 μg of polydIdC. Binding reactions were carried out in the absence (lane 1 and 6) or presence of 100-fold molar excess of competitors, as reported on the top of the figure. The position of complexes 2-6 and 1'-3' is indicated on the left and right side of the figure, respectively.

Fig. 11. Western Blot: CP1/NF-Y subunits are eluted at 0.4 M salt from Heparin-Sepharose column. Immunoblot analysis of 0.4 M salt heparin-eluted fractions with antibodies specific for NF-YA (panel on the left) and NF-YB (panel on the right) proteins. I: input; FT: flowthrough; 0.4M and 0.6M: 0.4 M and 0.6 M NaCl eluted fractions, respectively. The migration of the 42-KDa NF-YA and 32-KDa NF-YB proteins is shown.

3.3.2 Gel Filtration Chromatography

The Heparin-Sepharose 0.6 M fractions were further processed by gel filtration chromatography on a Superdex-200 column. As shown in Fig. 12, size exclusion chromatography resolved band 1' from bands 2' and 3' (see, as an example, fraction #67, #82 and #91).

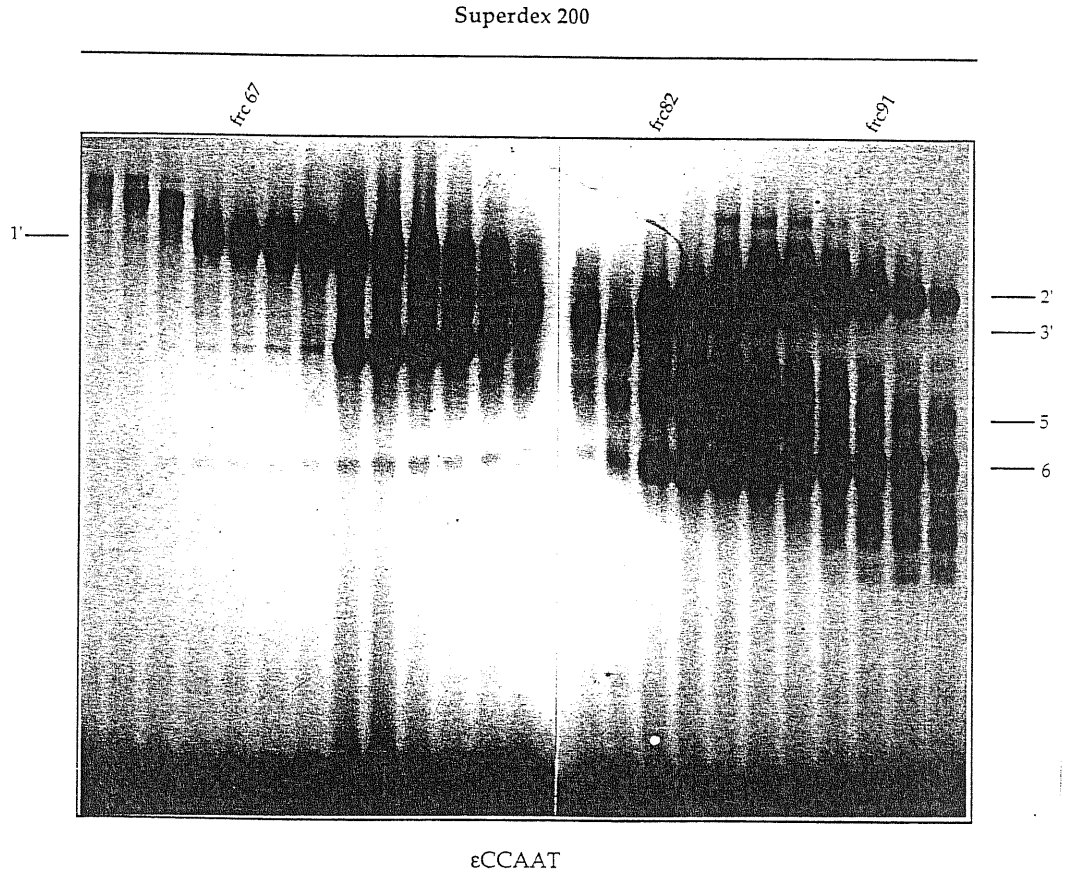

Fig. 12. Segregation of NFE-3 binding activities. Heparin-Sepharose active fractions were pooled and applied to a FPLC Sdx-200 column; two µl aliquot of the eluted fractions were incubated with εCCAAT ON in the presence of 1 µg of polydIdC and samples resolved in a non denaturing gel. The position of major retarded complexes are indicated.

To demonstrate that NFE-3/γdCCAAT (lane 1) and NFE-3/εCCAAT (lanes 2-14) activities were exactly the same observed in both K562 crude WCE and in heparin samples, Sdx fractions #67, #91, and #82 containing the activities 1', 2' and 3' respectively, were each challenged with 100-fold molar excess of the unlabeled ONs reported on the top of Fig. 13. The specific retarded bands were all competed out by γdCCAAT (γ, lanes 4, 8, 12) and εCCAAT (ε, lanes 5, 9, 13), but not by -117 HPFH (-117, lanes 3, 7, 11) and CP1/NF-Y consensus binding site (NFY, lane 2) ONs.

Protein standards were also subjected to Sdx-200 chromatography, under the same conditions. Comparison of their elution profile allowed to make an evaluation of NFE-3 molecular mass of about 400 KDa for the heaviest complex and 60 KDa for the lightest one.

Since the 1' complex represents an unreported NFE-3-like DNA binding activity, further studies were done to deeply analyse its biochemical properties and to definitively exclude any relationship with the other predominant and ubiquitously expressed CCAAT-binding protein CP1/NF-Y.

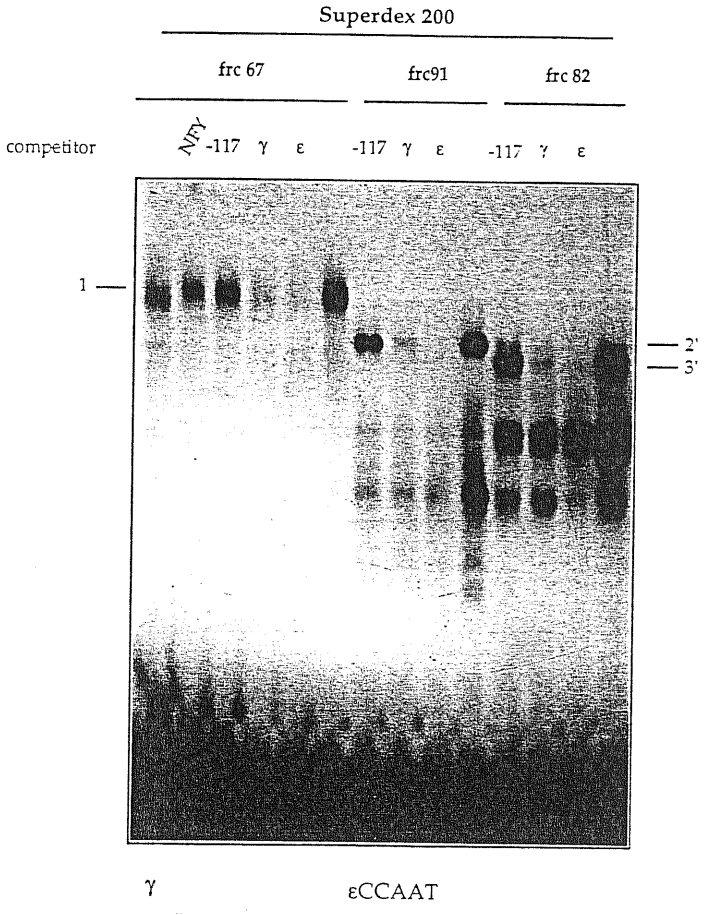

Fig. 13. Competition assay of samples purified by gel-filtration chromatography. Fractions #67, #91, and #82 eluted from Sdx-200 column were incubated with εCCAAT labeled ON (or γdCCAAT, lane 1) and 1 μg of polydIdC. Competition experiments were performed using 100-fold molar excess of competitors, as reported on the top of the figure. The position of complexes 1, 2' and 3' is indicated on both sides of the figure.

3.4 NFE-3 Binding Affinity Study

To directly estimate the relative binding affinity of NFE-3 for HPFH mutants, it was assessed the ability of NFE-3 to bind γdCCAAT, -117 HPFH or -114 HPFH ONs in competition experiments. A series of parallel binding reactions containing the NFE-3 complex 1 and the γdCCAAT labeled ON were incubated with increasing concentrations of unlabeled competitors (10- to 100-fold molar excess). Fig. 14 shows that wild type γdCCAAT (γWT, lanes 2-4) fragment is a much higher effective competitor than either -117 HPFH (-117, lanes 5-7) or -114 HPFH (-114, lanes 8-10) mutants confirming the *in vitro* binding identity between canonical NFE-3 activities (bands 2, 3 - data not shown -) and this newly identified one.

3.5 Further Characterization of NFE-3 Binding Specificity

It has been shown that NFE-3 binds to the distal CCAAT box of the γ-globin gene promoter but not to ONs carrying either -117 (G->A), -114 (C->T) HPFH point mutations (41, 42, 50) or 13-base pairs HPFH deletion (43). In the attempt to

evaluate the behaviour of the Sdx fractions containing NFE-3 activity (activities 2/3, not shown; activity 1, in Fig. 15), direct binding to wild type and mutant probes were performed. In Fig. 15 is shown that NFE-3 failed to bind to the proximal CCAAT box of the human γ-globin promoter (PγH, lane 1) in spite of the high degree of homology with the distal one; no binding was also observed when -117 (-117, lane 3), -114 (-114, lane 5) and 13-bp deletion (Δ13, lane 4) ONs were tested.

These findings are in agreement with reported DMS interference data (238).

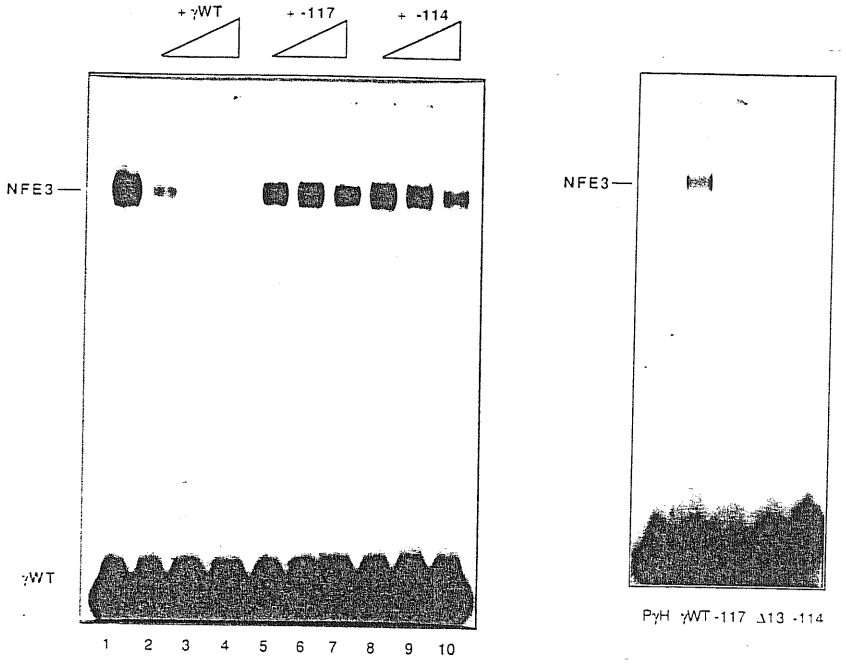

Fig. 14. Titration assay (left). A fixed amount of Sdx-purified NFE-3 protein and labeled γdCCAAT probe were mixed with increasing amounts of cold competitors and of 1 μg of polydIdC. Lanes 2, 3, and 4: 10-, 50-, and 100-fold molar excess of γdCCAAT ON (γWT); lanes 5, 6, and 7: 10-, 50-, and 100-fold molar excess of -117 HPFH ON (-117); lanes 8, 9, and 10: 10-, 50-, and 100-fold molar excess of -114 HPFH ON (-114). Retarded complex, corresponding to band 1, is indicated on the left.

Fig. 15. Direct binding of Sdx-purified fractions to γdCCAAT and mutant probes (right). NFE-3 active samples eluted from Sdx-200 column were tested with different labeled ONs, in the presence of 1 μg of polydIdC. Lane 1: proximal CCAAT box of the human γ-globin promoter (PγH); lane 2: wild type γdCCAAT (γWT); lane 3: -117 HPFH (-117); lane 4: Δ13 HPFH (Δ13); lane 5: -114 HPFH (-114). NFE-3 retarded band is indicated on the left.

3.6 NFE-3 is not an Alternative Form of CP1/NF-Y

In the active form CP1/NF-Y consists of a heterocomplex composed of at least two subunits, A and B (128 and Fig. 11). It has been shown that alternative forms of CP1/NF-Y can originate from alternatively spliced NF-YA mRNA (250). Before proceeding to further purification, the possibility that NFE-3 could be one alternative form (band 1) or a degradation product (bands 2 and 3) of CP1/NF-Y was tested. Antibody-mediated supershift assay was performed on Heparin-Sepharose

Antibody-mediated supershift assay was performed on Heparin-Sepharose fractions containing either CP1/NF-Y or NFE-3. As shown in Fig. 16, an antisera raised against an invariant peptide of the NF-YA subunit (251) did not alter NFE-3-γdCCAAT complexes (Hep 0.6M + αNF-YA, lane 5) even at concentration capable of supershifting CP1/NF-Y-γdCCAAT complex (Hep. 0.4M + NF-YA, lane 2), excluding any immunological correlation between these two CCAAT-binding factors.

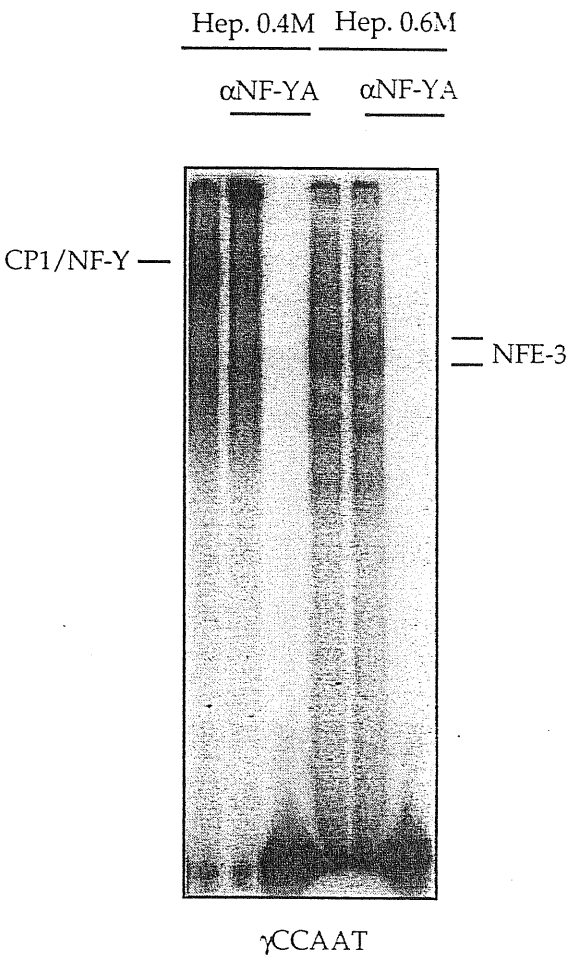

Fig. 16. Supershift of CP1/NF-Y by anti-NF-YA antibodies. EMSA reactions were performed with 0.4 M (Hep. 0.4M) and 0.6 M (Hep. 0.6M) salt eluted fractions from Heparin-Sepharose column in the absence or presence of anti-NF-YA antibodies (αNF-YA). Proteins were preincubated with antibodies first, and then with labeled γdCCAAT ON, before being subjected to gel mobility shift analysis. Specific CP1/NF-Y and NFE-3 retarded bands are indicated on the left and right of the figure, respectively.

3.3.3 DNA-Affinity Purification

The two major NFE-3 binding activities individuated and deeply characterized after Sdx-200 purification step (band 1 and 2/3), were mixed with polydIdC (corresponding to about 10 ng/10 µl EMSA reaction) and singly purified by affinity chromatography on resin containing concatamerized γdCCAAT or εCCAAT ONs. Unfortunately, first attempts to purify NFE-3 by affinity approach were frustrating by the extreme low recovery of the protein. Fig. 17 shows a paradigmatic example: pooled Sdx-200 active fractions, containing the NFE-3 haviest complex, were loaded onto an εCCAAT-affinity column and eluted samples tested in mobility shift with the same labeled probe. The recovery was extremely low (sample #13 and #14) with a flowthrough containing most of the activity (F.T., lane 2). Flowthrough fractions were generally rechromatographied, but the expected recovery of additional NFE-3 was never observed (data not shown).

Among the possible explanations for such a result, the following were considered more plausible: i) inappropriate preparation of the affinity column; ii) very high εCCAAT-NFE-3 dissociation constant (Kd) that allows loaded proteins to bind to the immobilised DNA, but that also causes their immediate release from the consensus binding site (high off ratio); iii) fractionation of NFE-3 components (as suggested by gel filtration data) critically involved in DNA binding. To address the quality of the affinity resin, γdCCAAT-NHS affinity column was used to purify CP1/NF-Y. It succeeded as expected (data not shown) proving that affinity-column was correctly prepared.

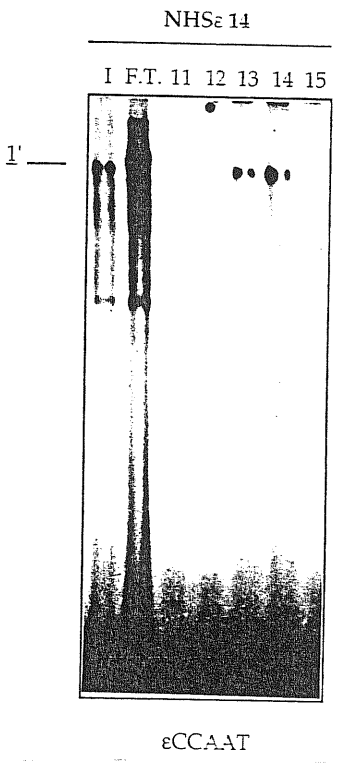

Fig. 17. EMSA of affinity-purified NFE-3. Sdx-eluted fractions containing NFE-3 1' binding activity were pooled, mixed with polydIdC and applied to εCCAAT-coupled NHS-column. Bound proteins were eluted with a linear 0.1-1 M NaCl gradient and tested in mobility shift with labeled εCCAAT ON and 100 ng of polydIdC. Lane 1: Input (I), lane 2: flowthrough (F.T.) and lanes labeled as 11 to 15: fractions eluted at 0.5 M NaCl. Specific retarded band is indicated on the left.

3.7 Complementation Assay

Mobility shift analyses of samples eluted from Sdx-200 column, led to hypothesize that the specific slower migrating NFE-3 complex might be the result of protein-protein interaction occuring in parallel with DNA binding. The possibility that band 1 might be due to overlapping components, partially resolved during affinity-chromatography, was tested by complementation assay.

Usually mobility shifts performed with DNA-affinity purified samples show NFE-3 binding activity in both the input material and the flowthrough sample and very weak or no activity in the other eluted fractions. The experiment shown in Fig. 18 is an example; fractions #15 and #18 (unable to bind by itself to the εCCAAT or γdCCAAT ONs) could be complemented by a very high dilution of flowthrough sample (FT 1/10, lane 2) that did not give rise to any specific activity in EMSA (compare lane 2 with lanes 8 and 11). An equivalent result was obtained when a pool of all the eluted DNA-affinity fractions, by itself unable to bind to the γdCCAAT (pool 12-18$^-$, lane 12) was complemented with diluted flowthrough (pool 12-18+, lane 13). The reconstituted activity migrated in EMSA faster than the canonical one.

Conceivably, the reason for this finding could be either degradation of the protein/s or inability to reconstitute the fully active NFE-3 binding activity. These data reinforced the hypothesis that NFE-3 binding activity might be separated in few essential components (one segregating in fraction #15, the other in fraction #18 and the third one lost in the flowthrough) and that their combination might result in a striking, but partial, restoration of binding activity.

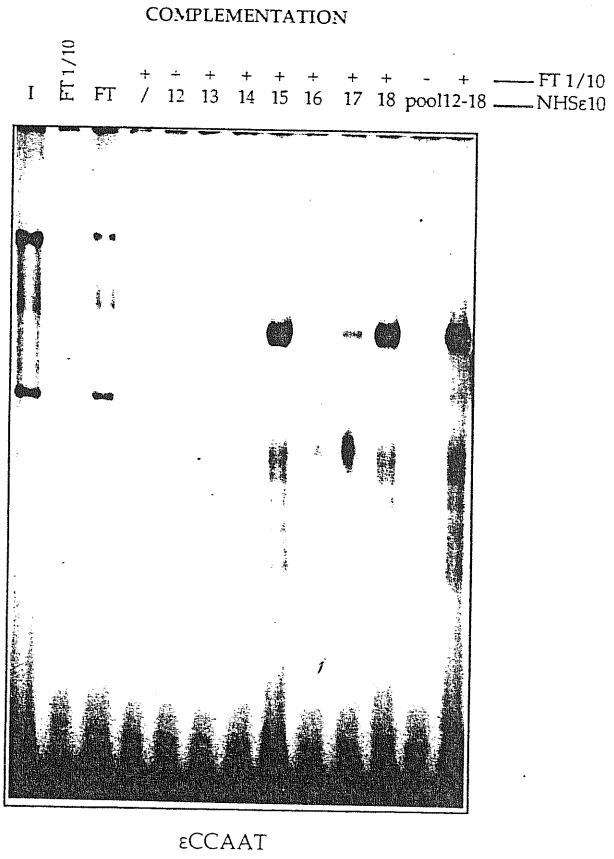

Fig. 18. Reconstitution of NFE-3 binding activity in complementation assay. Affinity-purified samples were tested in EMSA with labeled εCCAAT ON and 100 ng of polydIdC. Lane 1: input (I); lane 3: flowthrough (FT); lane 2: 1/10 dilution of flowthrough sample. From lane 5 to 11: eluted fractions complemented with 1 µl of 1/10 FT dilution; lane12: an aliquot of pooled fractions from #12 to #18 (pool 12-18 -); lane 13: #12-#18 pooled sample mixed with the same amount of 1/10 diluted FT tested in lane 2 (pool 12-18 +).

3.8 On-Off Rate Study

Dissociation rate of pooled flowthrough samples from DNA-affinity column was analysed. An aliquot was preincubated with labeled DNA and 100 ng of polydIdC, as a standard EMSA reaction; then a 100-fold molar excess of cold εCCAAT ON was added, and the disappearance of protein-DNA complexes was monitored over time by loading aliquots onto a continuously running gel. As shown in Fig. 19, NFE-3 binding resulted very unstable. No retarded bands were discernible after few seconds of chase (0 time, lane 6) also after long-exposures of the autoradiogram. This simple dynamic study suggested that εCCAAT-NFE-3 binding instability might be responsible for the low protein recovery occuring during DNA-affinity chromatography.

In the attempt to circumvent such a problem, a tentative purification with DNA-coated magnetic beads was tried to shorten the time between binding to and elution from immobilised DNA.

3.3.4 Anion Exchange Chromatography

In order to increase the protein molar concentration of samples accumulated during various affinity-purification steps, an anion exchange (Mono Q) chromatography was performed. On elution with a linear NaCl gradient, the majority of NFE-3 binding activity was detected at a NaCl concentration of 0.45-0.5 M. As hown in Fig. 20, Mono Q flowthrough (FT, lane 2 and 11) was totally devoided of any binding activity, but the eluted active fractions shown a partially new pattern of retarded bands. In particular, band 1 was no more detectable and two other activities, faster migrating, were observed.

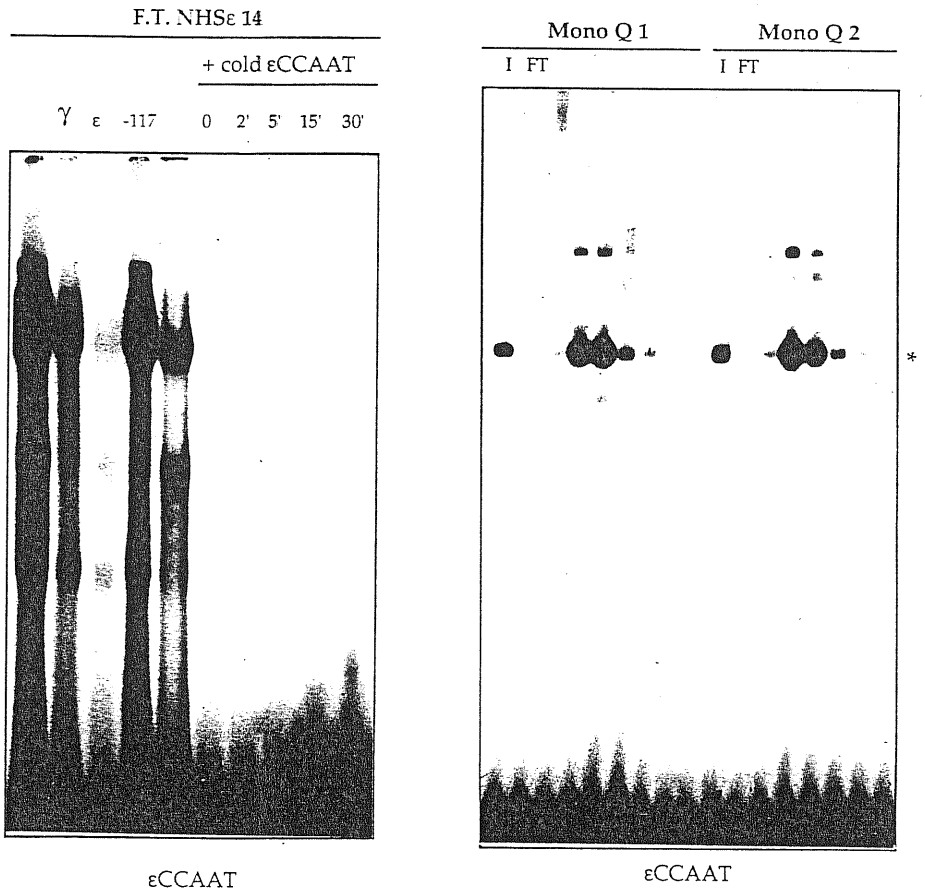

Fig. 19. On-off rate study (left). Active flowthrough sample recovered from εCCAAT DNA-NHS column was preincubated on ice with labeled εCCAAT ON and 100 ng of polydIdC. After 30 min, 100-fold molar excess of cold εCCAAT ON was added and 10 µl aliquots were loaded onto a continuously running gel at different times, as indicated above each lane in minutes. In lane 5 is the control; in the first four lanes, a typical competition experiment performed with γdCCAAT (γ, lane 2), εCCAAT (ε, lane 3) and -117 HPFH (-117, lane 4) ONs is shown.

Fig. 20. Anion exchange chromatography (right). NFE-3 active fractions obtained by DNA-affinity chromatography were pooled and applied to a Mono Q column. Active samples eluted at 0.5 M NaCl were assayed in EMSA with labeled εCCAAT ON. Two identically performed chromatographies are shown (Mono Q1 and Mono Q2). Major retarded band is indicated with an asterisk on the right of the figure. I: input; FT: flowthrough.

3.3.5 Protein Purification with DNA-Coated Magnetic Beads

To purify the binding activity from Mono Q fractions, biotinylated εCCAAT or γdCCAAT ONs were immobilised on streptavidin-coated paramagnetic particles.

As shown in Fig. 21 the proteins responsible for the binding observed in the Mono Q fractions were efficiently captured. No εCCAAT-binding activity was discernible both in flowthrough (FT, lane 3) and after washes (W, lane 4), while fractions eluted at 0.5 M NaCl (0.5M, lane 5 and 6) gave rise to a strong retarded band when assayed in EMSA with labeled εCCAAT ON and without polydIdC.

Unfortunately, such activity was efficiently competed out by -117 HPFH ON (data not shown), in contrast to the well established NFE-3 binding peculiarities.

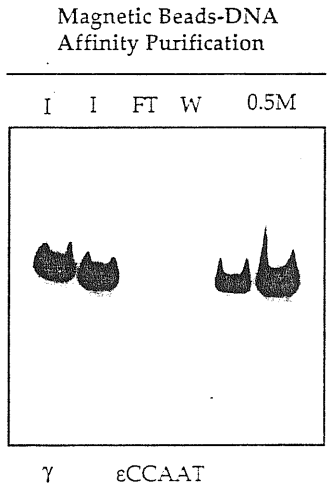

Fig. 21. Protein purification by εCCAAT-coated magnetic particles. NFE-3 active fractions from Mono Q chromatography were pooled, mixed with polydIdC and purified with εCCAAT-coated magnetic particles. Active samples eluted at 0.5 M NaCl were tested in EMSA using labeled γdCCAAT (lane 1) and εCCAAT (lanes 2-6) ONs, without polydIdC. Lane 1 and 2: Input (I); lane 3: flowthrough (FT); lane 4: first wash (W); lanes 5 and 6: 0.5 M salt eluted fraction (0.5M).

This finding could be due either to the fractionation of non specific DNA-binding proteins or to the purification of a DNA-binding subunit that lacks NFE-3 specificity because missing the discriminatory components. The latter hypothesis was preferred on the basis of several indications: i) the nature of the input material - proteins undergone to several steps of fractionation -; ii) the complementation data shown above; iii) the possibility that the lack of the NFE-3 specificity in these fractions reflects the loss of discriminatory components that confer specificity to a DNA-binding subunit; iv) the possibility that protein degradation could influence either the binding specificity or the on-off rate.

The complexity in protein constituents of active fractions was evaluated by silver staining. As shown in Fig. 22 (Panel A), DNA-binding activity segregated with three major polypeptides with an estimated molecular weight of 66-, 68-, and 82-KDa.

Among the transcription factors known to play a role in the modulation of globin gene expression, an involvement of the ubiquitous transcription factor YY1 has been clearly demonstrated (61, 132, 141, 147). YY1 migrates in SDS-PAGE as a 68-

KDa protein (131) and therefore the possibility to have YY1 within the three isolated polypeptides was assessed by cross-competition, supershift, and immunoblot assays.

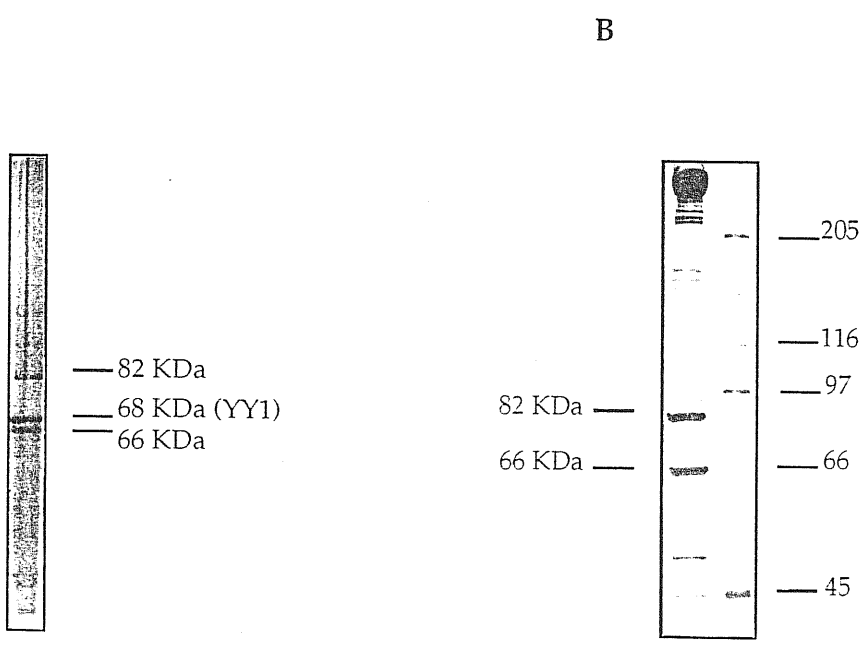

Fig. 22. Silver staining of samples purified with DNA-coated beads, before and after YY1-deprivation. Panel A: an aliquot of sample purified with εCCAAT-coated magnetic particles was separated on 10% SDS-PAGE and visualized by silver staining. Three major polypeptides of 82- 68- and 66-KDa are marked on the right of the figure. Panel B: silver staining of the same affinity-purified sample after YY1-deprivation. Two predominant bands of 66- and 82-KDa are indicated on the left, while standard molecular weights are indicated on the right of the figure.

YY1 binds to the CCAT or ACAT consensus sequences (136, 137), while its ability to recognize and bind the CCAAT box has not been reported. Using YY1 consensus binding site ON (-70 to -50 region of the adenovirus associated P5 promoter -P5 ON-) a tentative cross-competition was performed. Affinity-purified fractions were either incubated with labeled P5 ON and competed out with 100-fold molar excess of εCCAAT ON, or, vice versa, incubated with labeled εCCAAT ON and competed out with 100-fold molar excess of P5 ON. Unfortunately, no clear and definitive results were obtained, presumably due to the different experimental conditions required for optimal εCCAAT affinity-purified proteins and YY1 *in vitro* binding (see Materials and Methods). The same rather confusing conclusions were achieved when supershift experiments with anti-YY1 antibodies were performed (data not shown). More informative data came from Western blot analysis with anti-YY1 antibodies. Lane 3 of Fig. 23 (MonoQ7) shows that active fractions eluted from Mono Q column (and used as starting material for DNA-magnetic beads purification) contained a high amount of YY1 protein, strengthening the assumption that the 68

KDa polypeptide seen in the silver stained gel of Fig. 22 might be really ascribed to YY1.

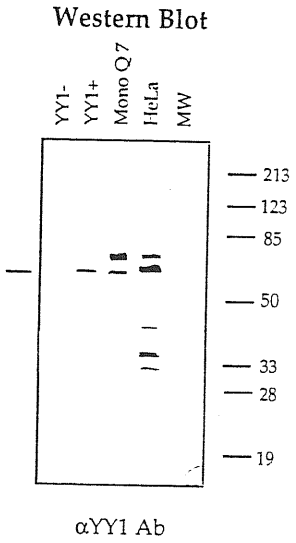

Fig. 23. Western blot: Immunodetection of YY1 protein in NFE-3 active samples. Nuclear extract from HeLa cells (HeLa, lane 4), Mono Q eluted fractions, before (Mono Q7, lane 3) and after (YY1$^-$, lane 1) YY1 deprivation, were fractionated by 10% SDS-PAGE, transferred to nitrocellulose and probed with anti-YY1 antibodies. In lane 2 (YY1+) is an aliquot of partially purified YY1 eluted from Ni^{2+}-NTA column. Molecular weights are indicated on the right and YY1 protein on the left of the figure.

3.9 YY1 Deprivation

A valid procedure to demonstrate whether YY1 participates in NFE-3 binding activity, was to eliminate YY1 protein from the eluted Mono Q samples and to investigate whether any change in their *in vitro* DNA-binding did occur.

YY1 was efficiently sequestered by Ni^{2+}-NTA chromatography, as demonstrated by mobility shift (using as labeled probe its consensus binding site -P5 ON) and immunoblotting assays (with rabbit anti-YY1 polyclonal antibodies).

EMSA in Fig. 24 shows that in Mono Q eluted sample (lane 5), YY1 activity identical to that observed in HeLa cells (lane 1), disappeared after only one round of Ni^{2+}-NTA chromatography (lane 3), while no change in the pattern of retarded bands obtained with εCCAAT probe could be demonstrated, being an increased binding affinity of either two reported NFE-3 complexes, the only striking difference (compare lane 2 and 4), to suggest a no better characterized YY1-mediated inhibition of NFE-3 binding.

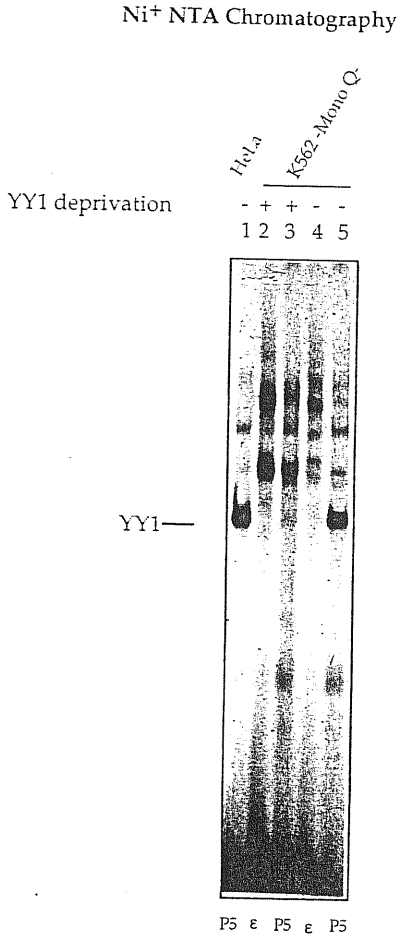

Fig. 24. EMSA: YY1 binding activity in NFE-3 active samples. Nuclear extract from HeLa cells (HeLa, lane 1) and Mono Q eluted fractions (K562-Mono Q-, lanes 2-5), before (lanes 4 and 5) and after (lanes 2 and 3) YY1 deprivation, were subjected to mobility shift analysis using both P5 (P5, lanes 1, 3, 5) and εCCAAT (ε, lanes 2 and 4) labeled ONs. YY1 retarded band is indicated on the left of the figure.

A precise correlation between EMSA and immunoblot analysis was also demonstrated. Western blot of Fig. 23 shows in fact that YY1 protein was effectively absent in Ni^{2+}-NTA flowthrough sample ($YY1^-$, lane 1), while still detectable in eluted fractions ($YY1^+$, lane 2), Mono Q fractions (MonoQ7, lane 3) and in HeLa cell nuclear extract (HeLa, lane 4).

3.10 Large-Scale Protein Purification

In light of the results reported above, large-scale protein purification was performed. Pooled Mono Q NFE-3 active fractions were subjected to Ni^{2+}-NTA chromatography and the flowthrough sample (deprived of YY1 protein/activity) further subjected to εCCAAT DNA-coated magnetic particles purification. The DNA-binding activity was efficiently and quantitatively purified.

Silver staining of purified proteins is shown in Panel B of Fig. 22. YY1 was no more detected (68 KDa polypeptide seen in Panel A of Fig. 22) and only two proteins of about 66- and 82-KDa, in apparently equimolar amount, were observed.

3.11 UV-Crosslinking

To establish whether both p66 and p82 polypeptides directly interact with γdCCAAT ON *in vitro*, EMSA mixture (in the presence of 10 ng of polydIdC and the same amount of protein tested in retardation shift) was irradiated with UV light and the cross-linked products analysed by SDS-PAGE. A predominant labeled complex of about 85 KDa was clearly seen (Fig. 25, lane 1). Correction of the molecular weight for the bound DNA led to protein of about 70 KDa, in good agreement with the apparent size of one of the two purified polypeptides. However, competition experiments revealed that unlabeled γdCCAAT (γ, lane 2), as well as -117 HPFH mutant (-117, lane 3) or CP1/NF-Y unrelated consensus sequence, from Y box of MHCII gene (CP1/NF-Y, lane 4) ONs could equally well abolish binding.

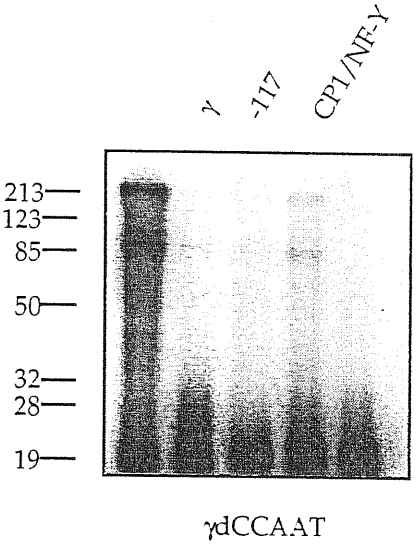

Fig. 25. UV-crosslinking of "affinity-purified activity". UV-crosslinking was performed with active fractions eluted from εCCAAT-coated magnetic particles, labeled γdCCAAT ON and 10 ng of polydIdC. After UV irradiation and electrophoresis onto 10% SDS-PAGE, labeled proteins were visualized by autoradiography. Competition was performed with 50-fold molar excess of γdCCAAT (γ, lane 2), -117 HPFH (-117, lane 3) and CP1/NF-Y consensus binding site (CP1/NF-Y, lane 4) ONs. In lane 5 no protein was added to the mixture reaction (negative control). Standard molecular weights are indicated on the left of the figure.

This finding was in agreement with gel shift competition data (see 3.3.5), and indicated that p66 subunit has, by itself, a low DNA binding specificity not completely restricted to the γdCCAAT box region.

The band visible at the top of the picture (migrating close to the 213-KDa molecular weight) corrsesponded to the slot of the gel.

3.12 Renaturation of "Affinity-Purified Activity" from SDS-PAGE

An aliquot of pooled active fractions was precipitated and separated by preparative SDS-PAGE. The region of the gel containing both p66 and p82 was excised and proteins electroeluted from the gel. After precipitation of the eluted proteins SDS was removed by acetone washing and the precipitated proteins resuspended in 6 M GdHCl. Subsequent removal of GdHCl by gel filtration chromatography allowed the slowly refolding of the polypeptide chains and furthermore, the recovery of DNA binding activity, as determined by EMSA (data not shown). The renatured polypeptides gave rise to the same pattern of retarded bands observed when samples eluted from εCCAAT-coated beads were tested (see, 3.3.5).

3.13 Mass-Mass Spectrometry (I)

Active samples eluted from DNA-coated magnetic beads were concentrated by TCA precipitation, separated by 10% SDS-PAGE and quickly stained with Coomassie; the two major bands of 66 KDa and 82 KDa were excised out and subjected to mass spectrometry analysis (M. Mann's lab - EMBL -). Proteins were partially sequenced by nanoelectrospray mass spectrometry and their coding genes identified by database search. Such analysis revealed, with no doubt, that p66 and p82 KDa were identical to the p70 and p80 subunits of the human Ku autoantigen.

3.14 p66 and p82 are Identical to Ku Autoantigen Subunits

To discriminate between a real participation of Ku subunits in NFE-3 binding activity and the possiblity of their unspecific enrichment during DNA-affinity purification, a variety of experiments was performed.

K562 WCE samples and active fractions eluted from Heparin-Sepharose column and DNA-coated beads were subjected to SDS-PAGE, blotted and reacted with antisera raised against human Ku (200). Ku subunits were abundantly detectable in almost all fractions tested and not only after affinity-chromatography where a sufficient enrichment of NFE-3 should have been achieved. As shown in Fig. 26, Panel A, p70- and p80-Ku were present in K562 WCE (lane 1) as well as in Heparin-Sepharose input (I, lane 6), flowthrough (F.T., lane 5), 0.4 M salt (H. 0.4M,

lane 3) and 0.6 M salt (H. 0.6M, lane 4) fractions. Ku coeluted also with γd- and ε-CCAAT binding activity in Sdx-200 (Fig. 26, Panel B, lanes 3-8) and, of course, in fractions purified DNA-coated magnetic beads purified fractions (Fig. 26, Panel A, Aff. purif., lane 2).

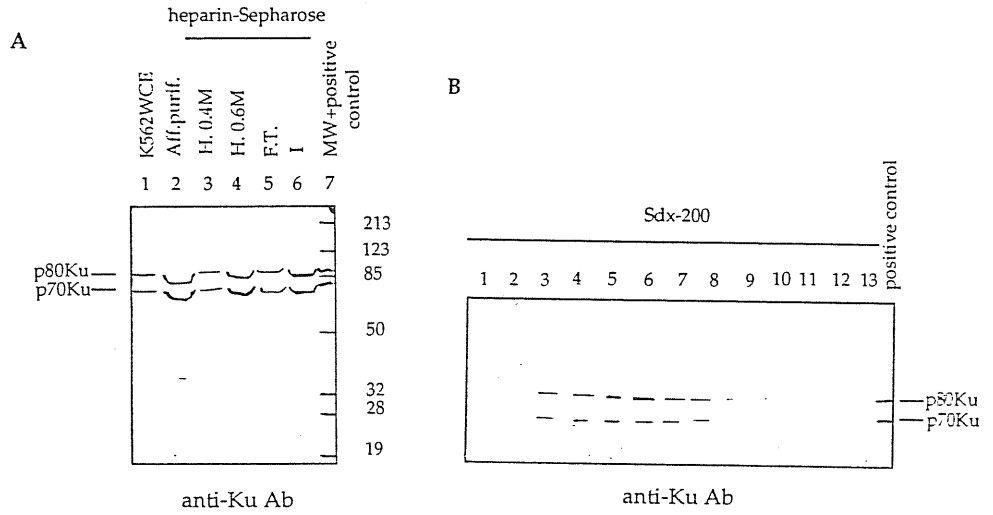

Fig. 26. Immunodetection of Ku autoantigen in NFE-3 active samples. Western blot: proteins were separated onto 10% SDS-PAGE, transferred to nitrocellulose and probed with anti-Ku antibodies. Panel A. Protein samples: K562 WCE (lane 1), Heparin-Sepharose input (I, lane 6), flowthrough (F.T., lane 5), proteins eluted from Heparin column at 0.4 M (H. 0.4M, lane 3) and 0.6 M (H. 0.6M, lane 4), and proteins eluted from DNA-coated magnetic beads (Aff.purif., lane 2). Positive control was run together with standard molecular weights (lane 7). p70 and p80 Ku polypeptides are indicated on the left and protein standards are indicated on the right of the figure. Panel B: fractions eluted from Sdx-200 column. Positive control was loaded in lane 14; p70 and p80 Ku are indicated on the right of the figure.

With the purpose to demonstrate the functional involvement of Ku in NFE-3 *in vitro* binding, polyclonal antisera to Ku were added to EMSA reactions containing: K562 WCE samples or active NFE-3 fractions derived from several chromatographic purification steps, 10 ng of polydIdC and labeled εCCAAT ON. Each sample was preincubated with increasing amount of anti-p70, anti-p80, or anti-heterodimeric Ku antisera, before and after the addition of labeled probe and analysed in gel retardation assay. Fig. 27 shows that anti-Ku antibodies specifically ablated (anti-p70, lane 2 and 3; anti-p80, lane 5 and 6) or supershifted (anti-p70, lane 4; anti-p80, lane 5) -according to their final concentration in the reaction- the "affinity-purified activity" complex. On the contrary, the same antibodies, in the same reaction conditions (but with the addition of 500 ng of polydIdC) were not able to affect binding between εCCAAT probe and partially purified NFE-3. As shown in Fig. 28, the binding of complexes 1', 2' or 3', present in Sdx-200 fractions, were not affected by the addition of anti-p70 Ku (lane 2 and 5), anti-p80 Ku (lane 3 and 6) or anti-p70/80 Ku (lane 7) antibodies, indicating that Ku did not take part in NFE-3 activity and

reinforcing the possibility of its non specific enrichment during DNA-affinity purification procedure.

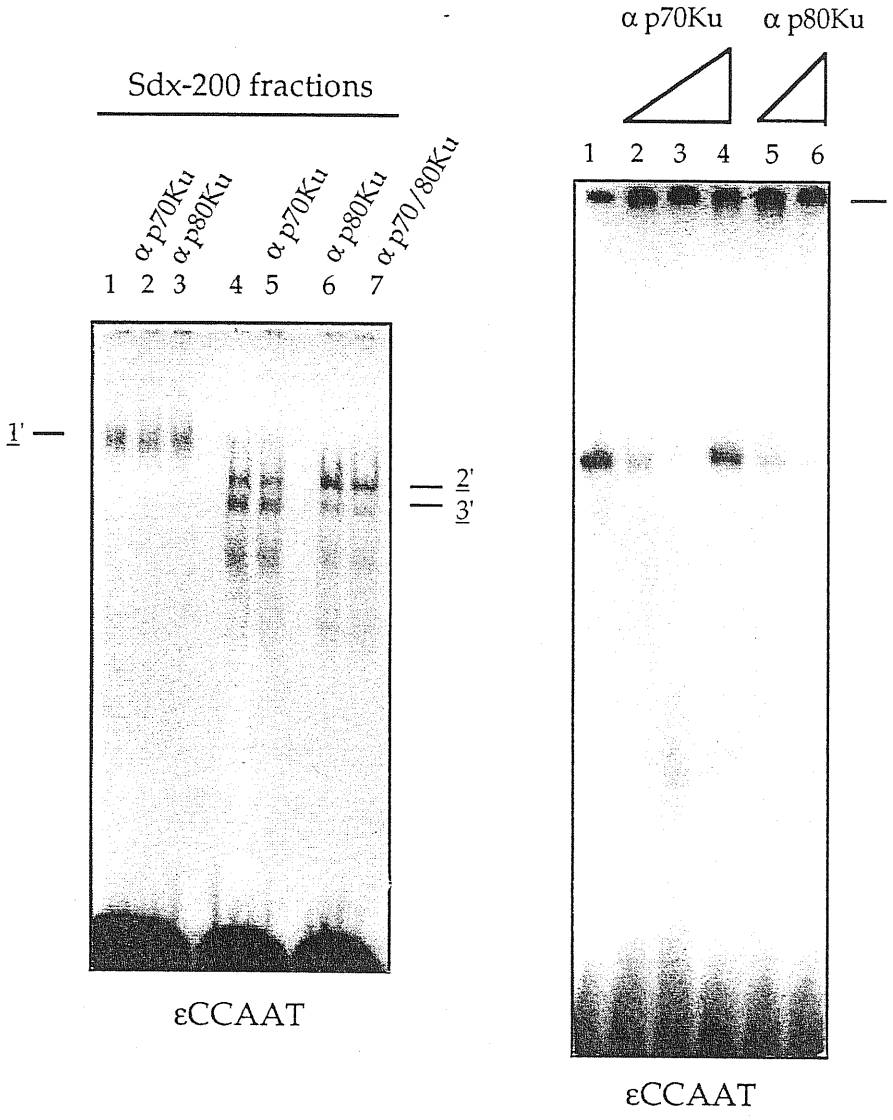

Fig. 27. Supershift assay with anti-Ku antibodies. "Affinity-purified activity" obtained after εCCAAT-coated magnetic particles step was preincubated with antibodies directed against either p70 or p80 Ku subunits and tested in mobility shift assay in the presence of 10 ng of polydIdC and labeled εCCAAT ON. Lane 1: control without antibodies; lanes 2 to 4: increasing amounts of anti-p70 Ku antibodies; lanes 5 and 6: increasing amounts of anti-p80 Ku antibodies. Shifted bands are indicated with a bar on the right of the figure.

Fig. 28. Supershift assay with anti-Ku antibodies. Partially purified NFE-3 samples eluted from Sdx-200 column were tested in mobility shift with anti-p70, -p80 and -p70/80 Ku antibodies (as reported on the top of the figure), in the presence of 500 ng of polydIdC and labeled εCCAAT probe. Lane 1 and 4: controls without antibodies. 1', 2' and 3' specific complexes are indicated on both sides of the figure.

The formation of ladder of retarded bands concomitant to increased protein concentration is a well established *in vitro* binding characteristic of Ku (173).

Fig. 29 (10X, lane 10) shows that increasing the protein to DNA ratio led to the appearance of more slowly migrating complexes also when affinity-purified NFE-3 is tested in EMSA.

The loss of NFE-3 binding specificity observed immediately after DNA-affinity purification was further investigated. In Fig. 29 is shown that binding of "affinity-purified activity" to γdCCAAT probe is displaced not only by addition of 50- fold molar excess of specific γdCCAAT (γ, lane 7), -117 HPFH, (-117, lane 8) and unrelated Sp1 consensus (Sp1, lane 9) ONs but also by an excess of synthetic DNA fragments, such as 150 and 500 ng of polydIdC (dIdC, lane 2 and 3) and, to a lower degree, 250 ng of calf thymus DNA (CTD, lane 4) and 250 ng of polydAdT (dAdT, lane 5); on the other hand, 250 ng of yeast tRNA (tRNA, lane 6) were unable to compete the binding.

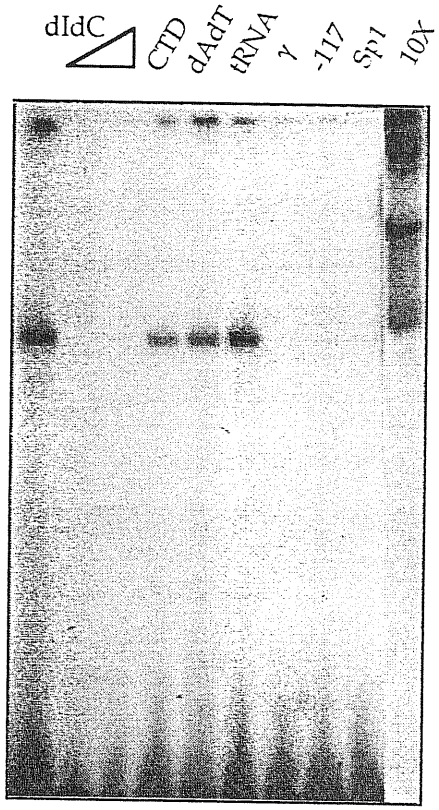

Fig. 29. Binding specificity of "affinity-purified activity". Active samples eluted from εCCAAT-coated magnetic beads were tested in EMSA with a variety of specific and aspecific DNA competitors added either before or after labeled γdCCAAT. Lane 1: control; lanes 2 and 3: 150 ng and 500 ng of polydIdC (dIdC) respectively; lane 4: 250 ng of calf thymus DNA (CTD); lane 5: 250 ng of polydAdT (dAdT); lane 6: 250 ng of yeast tRNA (tRNA); lanes 7, 8 and 9: 50-fold molar excess of γdCCAAT (γ), -117 HPFH (-117) and Sp1 consensus (Sp1) ONs, respectively; lane 10: the same amount of labeled probe tested with 10 times more affinity-purified protein (10X).

All together, these results established that Ku subunits were enriched during the DNA-affinity chromatographies, in a non-specific manner with no relationship with the ONs used for NFE-3 purification, even though this is in apparent contradiction with the finding that the protein was strongly bound to the DNA-affinity matrix, that was eluted from the column only at high salt concentration and, above all, it was not sequestered by others NHS-affinity columns used in series during NFE-3 purification (data not shown). Anyway, the most plausible conclusion is that Ku is not a component of NFE-3 complex, but it simply coeluts with such activity. Likely, Ku enrichment was caused by the low amount of aspecific competitor DNA used during affinity purifications. As mentioned above, polydIdC was reduced to favour NFE-3 binding. But unfortunately, being coeluted with NFE-3 and having a much higher affinity for free DNA ends, Ku was probably able to efficiently displaced NFE-3 from its specific consensus binding site ONs.

RESULTS: Part II

3.15 New Course in NFE-3 Purification

On the basis of accumulated data, it was really essential to confirm the strict binding specificity of partially purified NFE-3 samples before subjecting them to DNA-affinity chromatography.

Competition experiments were carried out with a large number of ONs added to the reaction mixture either before or after the εCCAAT and the γdCCAAT probes. In each reaction the binding was challenged with 100-fold molar excess of unlabeled competitor and at least 1 μg of polydIdC.

Fig. 30 shows an EMSA performed with Sdx fractions containing either the 1' (lanes 1-13) or the 2'/3' activities (lanes 14-26) and labeled εCCAAT ON. The binding was challenged with the ONs reported on the top of the figure. Both εCCAAT and γdCCAAT competed, while all the others competed less efficiently or did not compete at all. In particular, -117 HPFH (-117, lane 5 and 18), -114 HPFH (-114, lane 7 and 20), Δ13 HPFH deletion (Δ13, lane 6 and 19), γ proximal CCAAT (γp CCAAT, lane 4 and 17) ONs, and CP1/NF-Y (NF-Y, lane 12 and 25) and YY1 (YY1, lane 13 and 26) consensus binding site ONs were unable to displace NFE-3 complexes.

In competition assay, two different -117 HPFH ONs were tested: the ON already used in earlier experiments with blunted extremities (old -117 ON, -117o) and a new one (new -117 ON, -117n) with 5' protruding ends identical to those of εCCAAT and γdCCAAT ONs.

This was done in order to: i) reveal different binding affinities exclusively due to the specific HPFH point mutation and not to difference occuring in other regions of the DNA fragments; ii) avoid masking non specific Ku binding in mobility shift competition assays since it has indeed been proposed that once loaded onto a fragment of DNA, there is very little exchange of Ku protein to other fragments, unless identical DNA ends are present (175).

To conclude investigations on NFE-3 binding affinities, a titration curve with polydIdC was performed. Increasing amounts of polydIdC were added to EMSA reactions containing Sdx-purified NFE-3, bound to labeled γdCCAAT. NFE-3 retarded bands were not disturbed by polydIdC at a final concentration up to 3 μg/10 μl binding reaction (data not shown), in net contrast with the easy displacement of Ku by 100-150 ng/10 μl reaction of polydIdC (Fig. 29, dIdC, lane 2).

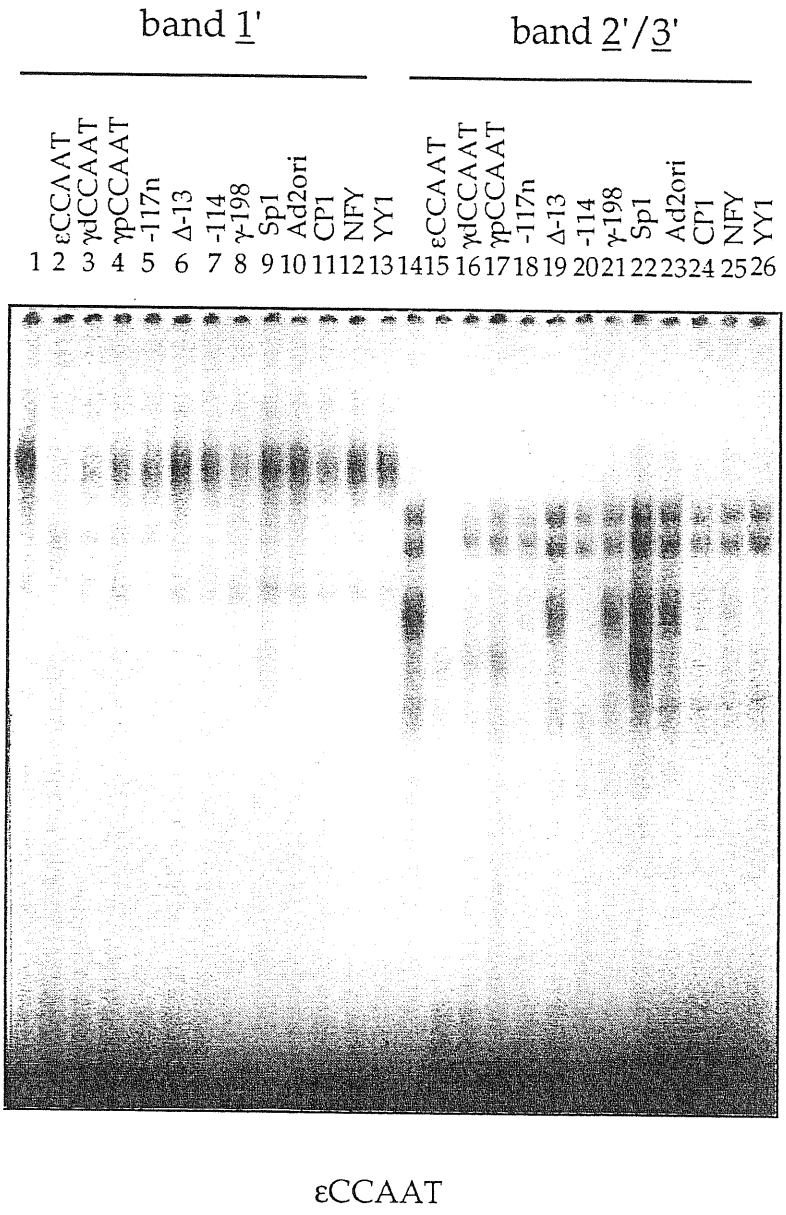

Fig. 30. ***In vitro*** **binding specificity of 1', 2', and 3' NFE-3 complexes.** NFE-3 active fractions (band 1', lanes 1-13; bands 2'/3', lanes 14-26), eluted from Sdx-200 column were tested in EMSA with labeled εCCAAT ON and 1 μg of polydIdC. Binding was challenged with 100-fold molar excess of CCAAT-related and unrelated double-stranded ONs, as reported on the top of the figure and listed in Tab. II.

3.16 New Purification Scheme

3.16.1 Non-Specific Competitor DNA

K562 WCEs were loaded onto a Heparin-Sepharose column as already described. Heparin fractions eluted at 0.6 M NaCl, containing NFE-3 activities 1 and 2/3 together were pooled, mixed with an appropriate amount of non-specific competitor DNA (equivalent to about 400 ng of polydIdC and 11 ng of calf thymus DNA/10 μl EMSA reaction) and applied onto a newly-made DNA-affinity column. A very large amount of polydIdC (about 700 μg) and calf thymus DNA (about 20 μg)

per ml of input material was used (approximately 40 times more than in previous DNA-affinity chromatographies) in order to avoid aspecific binding.

3.16.2 Buffer Composition

Buffer composition was modified by the addition of low levels of non-ionic detergent (0.009% Tween20) in all chromatographic and analytical procedures to further reduce non-specific DNA-protein interactions.

3.16.3 Oligonucleotides

A new synthetic ON (named γ42-36) was designed. γ42-36 is longer than γdCCAAT ON, it covers position from -132 to -95 relative to the cap site of the human γ-globin promoter and contains only two Cs 5' protruding, instead of the GATC sequence present in γdCCAAT ON. These partial modifications were done to avoid handling of 5' protruding ends already shown to be efficiently recognised by Ku.

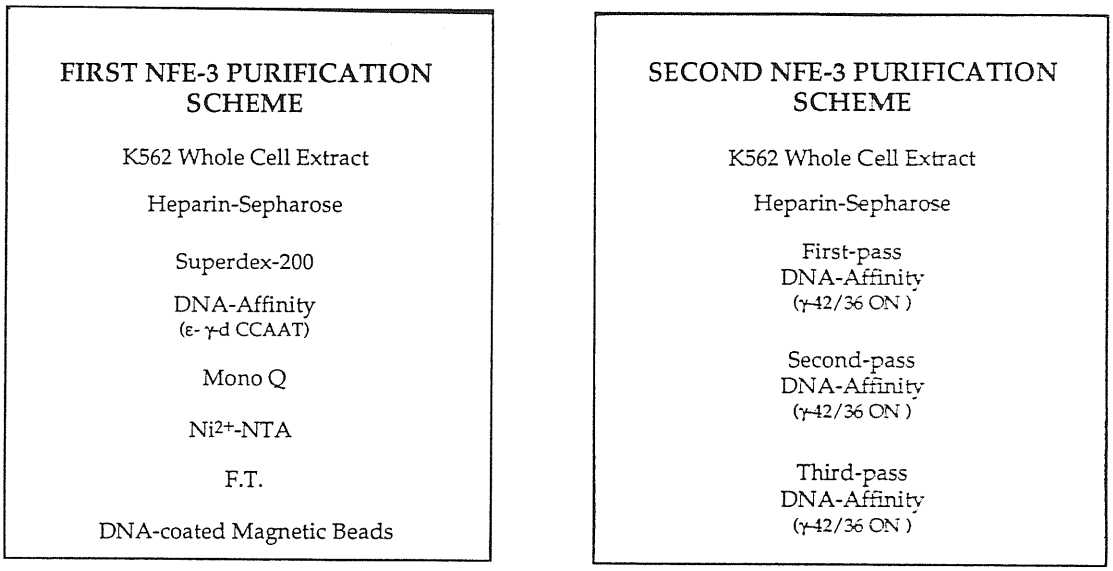

Fig. 31. Purification of NFE-3 from K562 cells. Purification schemes relative to the first and second purification approaches. The steps in purification of NFE-3 are indicated in simplified forms as a flowchart. For details, see Materials and Methods.

3.16.4 Affinity Chromatography

In the course of the first purification it turned out that NFE-3 binding activity might be attributed to a heterodimeric protein complex. To avoid potential and undesired segregation of the components, elution from the DNA-affinity column was carried out in a single step, with 1 M NaCl. The high competitor DNA/protein ratio, the presence in all buffers of 0.009% Tween20, the new γ42-36 ON and the

single one-step elution were all improvements that together allowed to purify NFE-3 proteins near to homogeneity after three rounds of affinity-chromatography (Fig. 31).

Gel retardation of samples recovered after the second round of DNA-affinity is shown in Fig. 32. EMSA reactions were performed with labeled γdCCAAT ON and 150 ng of polydIdC. Flowthrough (FT, lane 2) is devoided of any binding activity, while active fractions (lanes 5-8) gave rise to two classes of complexes (indicated on the right of the figure). The first one corresponds to the already characterized specific NFE-3 binding activities (band 1 and 2/3), the second one corresponds to three faster migrating complexes never identified during earlier purification and/or other gel retardation experiments (*a, *b, *c).

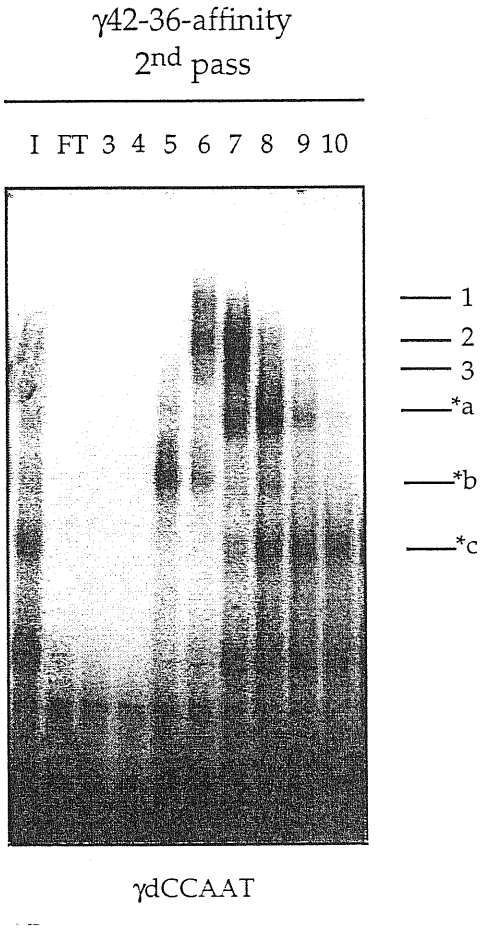

Fig. 32. Mobility shift of γ42-36 DNA-affinity purified NFE-3. Active samples eluted at 0.6 M NaCl from Heparin-Sepharose column were pooled and fractionated by two rounds of DNA-affinity chromatography. Fractions eluted at 1 M NaCl (from #3 to #10) were tested in EMSA with labeled γdCCAAT ON and 150 ng of polydIdC. I: Input; FT: flowthrough. Predominant retarded bands are indicated on the right of the figure.

3.17 Competition Experiments

Binding specificities of all the complexes were estimated by competition experiments with labeled γdCCAAT and 150 ng of polydIdC, as shown in Fig. 33. They all were competed out by 50-fold molar excess of cold γdCCAAT ON (γ, lane 3, 7, 11, 15). Δ13 HPFH ON (Δ13, lane 4, 8, 12, 16) did not compete, while -117 HPFH ON (-117, lane 5, 9, 13, 17) competed a little less efficiently than γdCCAAT ON, suggesting that even with this newly devised purification procedure, a segregation of some cofactor/s essential for specific NFE-3 binding might occur.

Interestingly, one complex (*b) exhibited typical NFE-3 binding specificities: it was competed out by wild type γdCCAAT (γ, lane 3), but it was not competed by

both Δ13 HPFH (Δ13, lane 4) and -117 HPFH (-117, lane 5) ONs. The relative difference in binding affinity between wild type and -117 HPFH distal CCAAT region was even much stronger than previously observed with partially purified NFE3 sample. Furthermore, *b activity seemed to be characterized by a lower dissociation rate and/or by a higher binding affinity for the wild type probe relatively to the other retarded bands. When competition experiments with cold γdCCAAT ON were performed on affinity-purified fractions containing band *b together with band 1 and 2 (γ, lane 7), band 2 and 3 (γ, lane 11) and band *a and *c (γ, lane 15), complex *b was less efficiently displaced. Presumably, limiting amount of cold competitor allowed the displacement of labeled γdCCAAT probe from complexes characterized by a low binding affinity and/or a high dissociation rate, to the expense of the high-affinity *b binding species.

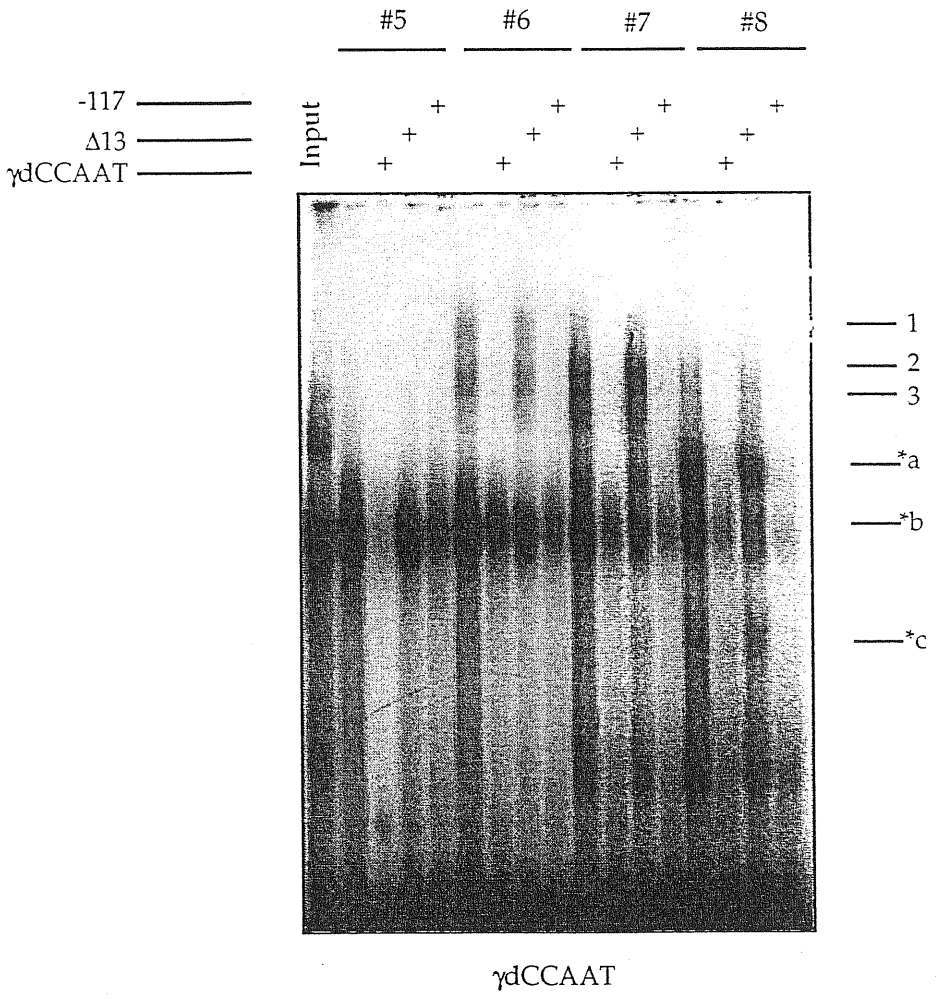

Fig. 33. Mobility shift competition assay of γ42-36 DNA-affinity purified NFE-3. Exactly the same fractions analysed in EMSA of Fig. 32 were incubated with labeled γdCCAAT ON, 150 ng of polydIdC and competition performed with 50-fold molar excess of cold competitors, as indicated on the top of the figure. -117: -117 HPFH ON; Δ13: Δ13 HPFH ON. Predominant retarded bands are indicated on the right of the figure.

3.18 Supershift Assay

To determine whether any of the NFE-3 retarded bands could be due to known DNA binding factors, some of which previously shown to coelute with NFE-3 activity, supershift experiments were performed. Fig. 34 shows that addition of anti-NF-E2 (αNF-E2, lane 2), anti-YY1 (αYY1, lane 4) and anti-heterodimeric Ku (αp70/p80, lane 6) antibodies, even at a concentration capable of affecting specific protein-DNA complexes (data not shown), did not affect NFE-3 binding. Therefore, YY1 and Ku proteins are no more present in affinity-purified NFE-3 fractions and they did not take a part in NFE-3 *in vitro* binding.

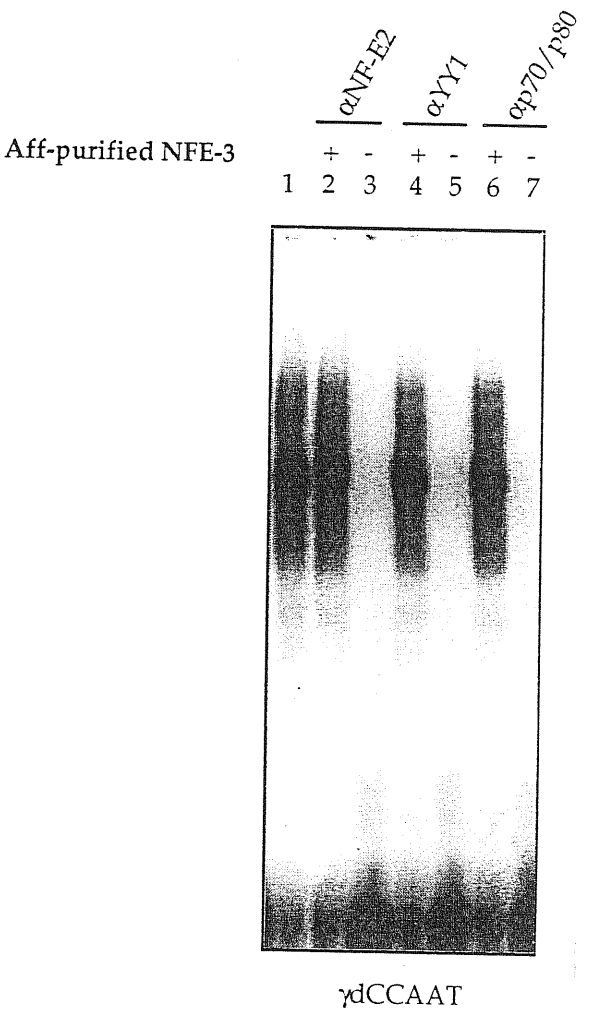

Fig. 34. Supershift assay with γ42-36 DNA-affinity purified NFE-3. Second-pass affinity-purified NFE-3 samples were tested in EMSA with labeled γdCCAAT ON and 150 ng of polydIdC. For supershift assay, anti-NF-E2 (lane 2), anti-YY1 (lane 4) and anti-p70/p80 Ku (lane 6) antibodies were added to the EMSA mixtures. Lane 1 is the positive control. In lanes 3, 5, and 7, anti-NF-E2, anti-YY1, and anti-p70/p80 Ku antibodies were tested in the absence of protein samples.

3.19 Silver Staining

Samples recovered after a third-passage of DNA-affinity chromatography were analysed by SDS-PAGE and stained with silver. Two polypeptides of 120 KDa (p120) and 40 KDa (p40) were observed (indicated with two dots in Fog. 35). Active samples were pooled and concentrated by anion exchange chromatography; a small aliquot of samples exhibiting NFE-3 binding activities in EMSA was loaded onto a 10% SDS-PAGE and silver stained. Fig. 35 represents a typical silver staining performed with NFE-3 samples eluted from different columns, at various purification steps. In particular, in lane 4 is the protein pattern of the γ42-36 DNA-affinity purified NFE-3 (silver staining of the relative input and flowthrough samples

were inconclusive due to smearing caused by high amount of aspecific competitor DNA present - data not shown -). Although it is well established that staining with silver does not accurately reveal the true ratio between different proteins, it seemed that p120 and p40 were not in a perfect equimolar amount, being p120 in slight preponderance.

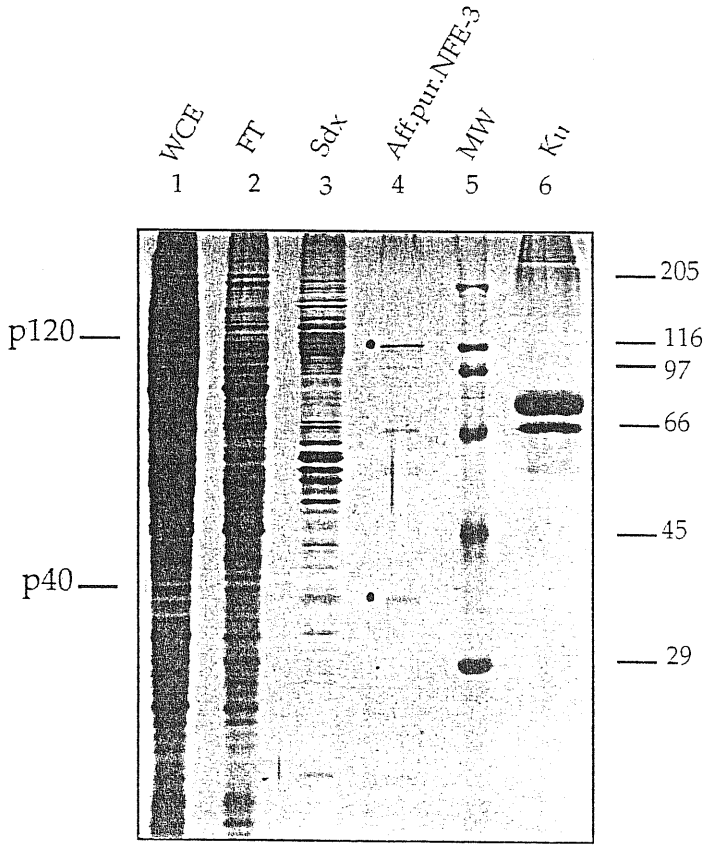

Fig. 35. Analysis of protein profiles after more significant purification steps. Active samples eluted from various chromatographic columns were precipitated with TCA, electrophoresed on 10% SDS-PAGE and silver stained. Lane 1: K562 WCE (WCE), or Heparin-Sepharose input; lane 2: Heparin-Sepharose flowthrough (FT); lane 3: Sdx-200 eluted fraction (Sdx); lane 4: NFE-3 active fraction obtained after three rounds of γ42-36 DNA-affinity purification (Aff.pur.NFE-3); lane 5: about 30 ng of each marker protein (MW); lane 6: DNA affinity-purified Ku autoantigen (Ku). The two predominant bands in lane 4 are indicated with dots and with bars on the left of the picture (p120, p40).

3.20 UV-Crosslinking of Affinity-Purified NFE-3 to "End-Labeled" DNA

To provide conclusive evidence that the two major polypeptide species eluted from γ42-36 DNA-affinity column represent NFE-3, and also to investigate whether both subunits are involved in DNA binding, UV-crosslinking to ^{32}P-end labeled γdCCAAT ON was performed.

Fig. 36 shows two labeled proteins of 120-130-KDa and 60-70-KDa. Since the irradiated sample was not digested by DNases, the observed molecular weight might be consistent with the labelling of the two major components observed in silver stained gel.

Even though the lightest lebeled polypeptide migrated in SDS-PAGE as a doublet of 60-70 KDa, both the UV-crosslinked complexes (120-130 KDa and 60-70 KDa) were

competed out by 50-fold molar excess of unlabeled γdCCAAT ON (γ, lane 3) but not by the same amount of Δ13 HPFH ON (Δ13, lane 4).

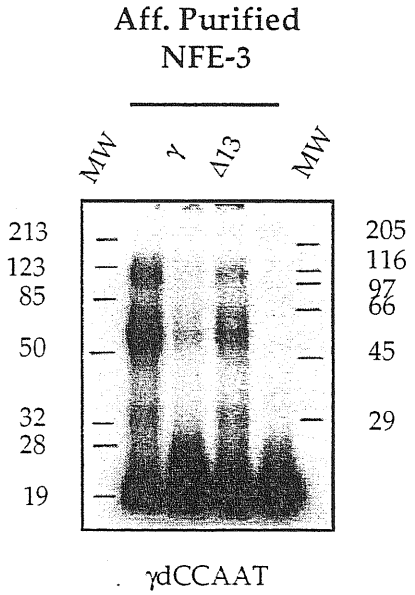

Fig. 36. UV-crosslinking of affinity-purified NFE-3 to end-labeled DNA. ^{32}P-end labeled γdCCAAT ON was crosslinked to NFE-3 active samples obtained after two passages through γ42-36 DNA-affinity column and complexes were separated on 12.5% SDS-PAGE (lane 2). Competition was challenged with 50-fold molar excess of either cold γdCCAAT (γ, lane 3) and Δ13 HPFH deletion (Δ13, lane 4) ONs; all reaction mixtures contained 150 ng of polydIdC. Prestained and high molecular weight protein standards are indicated on the left and on the rigth of the figure, respectively.

Considering the very low recovery of NFE-3, it was not possible to correlate these polypeptides with specific DNA-binding activity, by denaturation-renaturation experiment, as it was done with Ku. Anyway, further DNA binding studies (southwestern blot and UV-crosslinking with uniformly labeled DNA) were approached to test whether there are any significant differences between the specificity of the highly purified NFE-3 relative to crude fractions.

3.21 UV-Crosslinking of Partially-Purified NFE-3 to "Uniformly-Labeled" DNA

NFE-3 active fractions eluted from Heparin-Sepharose and Sdx-200 columns were UV-crosslinked to a γ42 ON (encompassing the dCCAAT box of the human γ-globin promoter, Tab. II) uniformly labeled with both α^{32}P-dCTP and α^{32}P-dTTP.The cross-linked products were digested with nucleases before loading them onto a 12.5% SDS-PAGE, so that the size of the radiolabeled complexes more closely approximated the size of the protein. As shown in Fig. 37, two major labeled complexes, with electrophoretic mobilities of 50-60 KDa and 120 KDa, were clearly detected in Heparin-Sepharose fractions eluted at 0.6 M NaCl (Hep 0.6M, lane 1) and in fractions #91, #67, and #82 from Sdx-200 column (lanes 3, 4, and 5). All these active samples gave rise to specific NFE-3 retarded bands when tested in EMSA (see, for comparison Fig. 10 and Fig. 12). Each crosslinking reaction was performed with 150 ng of polydIdC.

Competition with 50-fold molar excess of cold γdCCAAT and Δ13 HPFH ONs (data not shown) reproduced exactly the competition pattern observed when purified NFE-3 was tested in UV-crosslinking with end-labeled γdCCAAT (see Fig. 36). Binding of p120 and p50-60 to labeled γ42 ON was competed out by cold γdCCAAT ON, but not by the same amount of Δ13 HPFH ON.

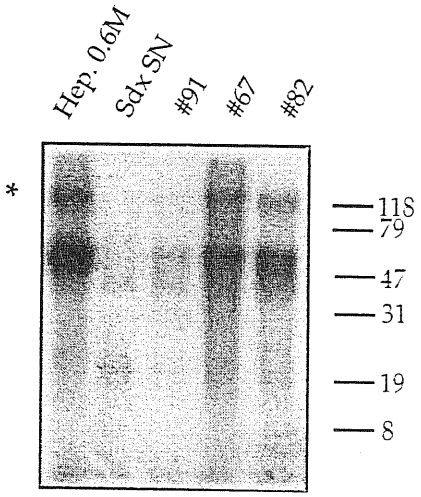

Fig. 37. UV-crosslinking of partially purified NFE-3 to uniformly labeled DNA. Protein samples and 150 ng of polydIdC were incubated with filled in α^{32}P-dCTP/dTTP γ42 ON, irradiated with UV light, digested with nucleases and electrophoresed on 12.5% SDS-PAGE. In lane 1: Heparin-Sepharose active sample eluted at 0.6 M salt (Hep. 0.6M); lane 2: supernatant after precipitation of Hep. 0.6 M fractions (Sdx SN); lane 3 to 5: Sdx-200 eluted active fractions (#91, #67, #82, respectively). Standard molecular weights are marked on the right, while p120 is indicated (with an asterisk) on the left of the figure.

3.22 Southwestern Blot

A southwestern assay was used to test the DNA binding activity of separated putative NFE-3 subunits. Partially purified NFE-3 samples eluted from Heparin-Sepharose at 0.6 M NaCl and from Sdx-200 columns were separated by 12,5% SDS-PAGE, transferred to nitrocellulose and probed with ^{32}P-labeled concatamerized εCCAAT ON.

As shown in Fig. 38, this resulted in the p120, but not the p40 binding to DNA. The labeled 47 KDa polypeptide was considered unrelated to NFE-3, as it was also present in the supernatant recovered after precipitation of fractions eluted at 0.6 M salt from Heparin-Sepharose, a sample totally avoided of any *in vitro* binding activity.

p120 binding was observed in the Heparin-Sepharose sample (Hep 0.6M, lane 5) and in the Sdx-200 fractions (#82, #67, and #91, lanes 1, 2, and 3) characterized by 2'/ 3' and 1' binding activities when tested in EMSA (see Fig. 10 and 12). Labeled proteins in sample #91 and #82 were observed only after longer exposure of the membrane (data not shown). The whole procedure was carried out with a constant amount of sonicated herring sperm DNA (10 μg/ml solution) as competitor DNA to avoid aspecific binding.

Being NFE-3 almost the only nuclear factor in K562 able to bind the εCCAAT box, the above results acquired great interest, also considering that the 120

KDa protein was clearly detected i) in UV-crosslinking (with uniformly labeled γ42- and end labeled γd-CCAAT ONs) ii) in southwestern assay (with end labeled concatamerized εCCAAT ON) and iii) both with partially and highly purified NFE-3 preparations. These data and the strong linkage between p120/p40 elution profile from affinity-column and NFE-3 *in vitro* binding, allowed to conclude that p40 and moreover p120 represent/s NFE-3. On the basis of the estimated molecular weights in SDS-PAGE and supershift experiments on affinity-purified NFE-3 (3.17), the involvement of YY1 and Ku was also definitely excluded.

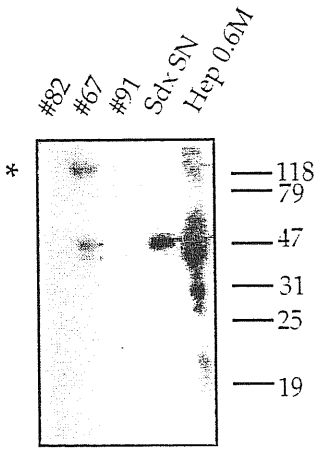

Fig. 38. Southwestern blot analysis of partially purified NFE-3 samples. Protein samples were separated by 12.5% SDS-PAGE, transferred to nitrocellulose and probed with ^{32}P-end labeled concatamerized εCCAAT ON. Lane 1 to 3: Sdx-200 eluted active fractions (#82, #67, #91, respectively); lane 4: supernatant after precipitation of Hep 0.6M fractions (Sdx SN); lane 5: Heparin-Sepharose active sample eluted at 0.6 M salt (Hep 0.6M). Standard molecular weights are marked on the right, while p120 is indicated (with an asterisk) on the left of the figure.

3.23 Mass-Mass Spectrometry (II)

NFE-3 samples recovered after a third passage through the γ42-36 DNA-affinity column were concentrated by an anion exchange chromatographic step, precipitated with TCA, resuspended in SDS sample buffer, resolved onto a preparative 12.5% SDS-PAGE and quickly stained with Coomassie. Both bands were individually excised out, kept in ultra-pure water and recently sent to Heidelberg for mass mass spectrometry analysis (M. Mann's lab).

CHAPTER 4

DISCUSSION

This work deals with the characterization and purification from K562 cells of the protein/s responsible for the NFE-3 binding activity.

NFE-3 was originally described as a factor present in extracts of erythroleukemic cell nuclei (K562, MEL, HEL) and capable to bind exclusively the distal and the unique CCAAT-boxes of the human γ- and ε-globin gene promoters. Genetic evidences point out that these *cis*-regulatory elements play a pivotal role in controlling the expression and the switch of globin genes during development.

Natural non-deletional HPFH mutations within the distal CCAAT-box of the human γ-globin gene promoters affect the binding of several factors *in vitro*. However, NFE-3 is the only factor whose binding is consistently decreased with all different HPFH mutations so far observed: the -117 G->A (130), the -114 C->T (237) and the Δ13 deletion (245) in the $^{A}\gamma$ promoter, and the -110 A->C in the $^{G}\gamma$ promoter (253). These evidences legitimate the efforts put in this work and the accomplishments pursued.

4.1 Characterization of NFE-3 activity

An accurate analysis was carried out to differentiate NFE-3 from the other CCAAT-binding proteins (in particular from the ubiquitously expressed CP1/NF-Y) and to highlight its complex biochemical nature.

Purification of NFE-3 was complicated by three major difficulties:

- first, the source material that is K562 cells
- second, the presence of several copurifying nuclear factors
- third, the limited amount of NFE-3

4.1.1 K562 cells

The choice of the human erythroleukemic cell line K562 as source of NFE-3 was obliged by the lack of large amount of nucleated human adult erythroid cells. K562 cells express the ε- and γ-globin but not the adult β-globin genes (32). These cells virtually represent the embryonic-fetal stage of erythroid development. However, they are transformed and, though inducible, have a narrow and limited

pathway of differentiation. Above all, these cells do not undergo a γ- to β-globin gene swicth, thus hampering the characterization of the molecular mechanisms involved in this phenomenon.

Furthermore, the frozen stage of differentiation of K562 cells reflects a cellular enviroment committed to the fully activation of embryonic-fetal genes and not to the genetic switch occuring during development.

If NFE-3 does control the developmental γ-globin silencing, it comes easy to wonder why is the protein(s) present in K562 cells that abundantly express the γ genes. As observed in other known models, post-translational modifications, productive interactions with other DNA-binding and non-DNA-binding proteins, competition with different proteins for the same cognate site within the promoters might occur and finely modulate/balance the action of any putative repressor, NFE-3 included.

Unfortunately, in the case of NFE-3 we do not know either the state or the relative proportions of the polypeptide(s) in adult erythroid cells (where HPFH phenotype is expressed), relatively to the embryonic/fetal cells .

4.1.2 Copurifying nuclear factors

During the purification procedure, two well characterized transcription factors coeluted with NFE-3 activity: CP1/NF-Y and YY1. Two specific chromatographic steps were devised in order to devoid extracts of both factors.

4.1.2.1 CP1/NF-Y

NFE-3 differs from CP1/NF-Y on the basis of both immunological and functional (*in vitro* DNA binding) criteria. In particular, antibodies directed against CP1/NF-Y do not affect the formation of the NFE-3-DNA complex (Fig. 16), ruling out the possibility that NFE-3 is a heterodimer of either CP1/NF-Y subunit with an unidentified protein.

In addition, we have recently shown, by DMS interference experiments (238), that NFE-3 and CP1/NF-Y both interact with bases -117, -115 and -114 of the γ-globin promoter, but a G at -117 position seems critical only for NFE-3 binding. Furthermore, only NFE-3 extends its contacts to nucleotides -107 and -108. This result is in agreement with the inability of NFE-3 to bind to the -117 and -114 HPFH ONs (Fig. 15).

As shown in Chapter 3, Heparin-Sepharose chromatography was a very efficient tool in fractionating these activities (compare Fig. 9 to Fig. 10).

4.1.2.2 YY1

YY1 is an ubiquitously distributed transcription factor which acts either as an activator or a repressor of viral or eukaryotic promoters. Various mechanisms have been proposed to explain YY1 action.

In particular, in the globin gene context, YY1 might directly interact with gene promoters and repress their expression or it might act via competition with *trans*-activators for overlapping binding sites. The first mechanism seems to occur in the context of the γ-globin gene promoter, since one conserved YY1 element located at -1086 exhibits repressor activity in transient transfection assays (61). Displacement of the transcriptional activator GATA-1 from its consensus binding site at -269 is, instead, the mechanism by which YY1 participates to modulate silencing of the ε-promoter (147).

The procedure flowcharted in Fig. 31 for the purification of NFE-3 included a step dedicated to the exclusion of YY1 (Ni^{+2}-NTA chromatography). This was done to simplify the pattern of protein correlating with NFE-3 activity but furthermore, to investigate whether YY1 could be part of such an activity. The typical NFE-3 retarded bands observed in EMSA are not affected by YY1 deprivation (Fig. 24), and this finding disfavoured any direct YY1 involvement.

On the other hand, the exclusion of this zinc finger protein clearly resulted in an increased *in vitro* binding affinity of NFE-3 to its consensus sequence (Fig. 24). This phenomenon can be interpreted as a consequence of the ability of YY1 to bend DNA. It could be that the lack of YY1/DNA bending activity, might result in an easier recognition and binding of NFE-3.

On the light of these results, as it will be also considered later on for Ku, it is possible to exclude the participation of YY1 in NFE-3 activity, but it cannot be ruled out the possibility that YY1 is included in the formation of a NFE-3 bigger complex.

Of course, cloning of cDNA(s) encoding NFE-3 protein(s) is necessary to solve this issue.

4.1.3 NFE-3 abundance in K562 cell line

In K562 cells, NFE-3 is not the major CCAAT box binding protein. Even though the *in vitro* binding efficiency does not necessarily correlate with the amount of a given DNA-binding factor, it may be assumed that NFE-3 is much lesser represented than CP1/NF-Y, Ku or GATA-1 nuclear proteins in K562 whole cell extracts.

4.2 First purification procedure

To purify NFE-3 from K562 whole cell extracts, it was devised the purification procedure depicted in Fig. 31, consisting of five distinct chromatographic steps with Heparin-Sepharose, Sdx-200, DNA-affinity Superose, Mono Q and Ni^{+2}-NTA columns and a final step of purification with DNA-coated magnetic beads.

Heparin-Sepharose column succeded in fractionating NFE-3 from CP1/NF-Y activity (Fig. 9 and 10). Gel filtration (Sdx-200) chromatography was necessary and sufficient to separate the canonical NFE-3 activities (here named 2/3) from the slower migrating and newly characterized one (named 1) (Fig. 12).

On the contrary, first attempt to purify NFE-3 by DNA-affinity purification resulted in a very low recovery of the activity, whenever concatamerized γdCCAAT or εCCAAT ONs were coupled to the activated Sepharose matrix (Fig. 17). Besides, much of the activity was observed in the flowthrough fractions. Simple experiments were done in order to understand the dynamic of NFE-3 binding to the DNA; the data shown lead to hypothesize that a very high off-rate distinguishes the protein (Fig. 19), explaining in such a way the not easy task in purifying NFE-3 by affinity chromatography.

MonoQ-concentrated affinity fractions contained YY1. It was included in the purification procedure an additional step to get rid of it. YY1 protein in fact has a stretch of 16 histidines in a row, it binds tightly to Hi-trap chelating resin and after only one round of Ni^{+2}-NTA chromatography, both YY1 *in vitro* binding activity and YY1 protein disappeared (Fig. 24).

To keep the active protein highly concentrated and to perform DNA binding and elution as quick as possible, the classical DNA-affinity column was substituted by a DNA-coated magnetic particles purification (Fig. 21). Biotynilated εCCAAT ON was chosen because it was shown that NFE-3 binds the CCAAT box of the ε globin promoter stronger than the distal CCAAT of the γ promoter and also because γdCCAAT is crowded by general CCAAT-binding factors, while the εCCAAT region is almost only bound by NFE-3.

At the end of this laborious procedure two predominant polypeptides of 66- and 82-KDa were observed (Fig. 22).

4.2.1 Identification of Ku autoantigen

Mass-mass spectrometry analisys of both affinity-purified polypeptides revealed their identity to sequences within the two subunits of the human Ku autoantigen.

The identity between Ku and the purified proteins was also inferred by their similar molecular weights, their identical behaviour in mobility shift and UV-

crosslinking assays and by the recognition of p82 and p66 by anti-Ku antibodies both in Western blot and supershift experiments.

It was investigated whether p70 and p80 Ku are integral members of the specific NFE-3 complex. Any attempt to demonstrate this hypothesis failed and finally, reconsidering the collected data, it was concluded that Ku purification and isolation was the direct consequence of its aspecific binding to DNA-coated magnetic beads. This unwanted result is probably due to the fact that: i) Ku and NFE-3 have similar biochemical characteristics - responsible of their coelution during initial purification steps -, ii) DNA-affinity purification and binding assays were often performed in the presence of low amount of unspecific competitor DNA, iii) the purification with magnetic beads was performed with non-concatamerized ONs (and with a very high amount of free DNA ends).

On the other hand, recent data suggest that Ku might play some regulatory role(s) in transcription processes (187, 190, 191) and the partecipation of Ku in sequence-specific DNA binding has also been described (197).

In light of these findings, it is still controversial the observation that Ku binds with high affinity the γdCCAAT region with respect to different consensus sites. In our hands, purification of other factors (e.g. octamer, GATA-1), from affinity resins loaded in series with the γdCCAAT resin, do not appear contaminated by Ku. The same behaviour was observed when batch chromatography was performed.

4.3 Second purification procedure

In order to avoid copurification of both Ku and YY1 and also to increase NFE-3 recovery a novel purification approach was setted up. As shown in Fig. 31, NFE-3 purification was simply accomplished by Heparin-Sepharose and DNA-affinity chromatographies.

K562 whole cell extract preparation and the Heparin-Sepharose purification were performed as in the first purification approach, with only a slight modification in buffer composition, consisting in the addition of 0.009% Tween20. This non-ionic detergent decreases aspecific DNA-protein interactions and increases protein solubilization.

Small aliquots of active samples recovered from the Heparin-Sepharose were loaded onto the affinity column and subjected to three rounds of purifications. This was done to guarantee a net molar excess of specific DNA-binding sites toward the unspecific ones and to permit a faster running procedure that also meant a faster elution of bound proteins.

The recovery of bound proteins during the first two consecutive affinity-purifications was achieved by one-step elution at 1 M NaCl, to avoid potential and

undesired segregation of NFE-3 components (Fig. 32). It was a feasible procedure because the very high concentration of aspecific competitor in the injected material, caused very few proteins to be retained onto the column.

A new designed ON (with different 5' protruding ends) and a very large amount of apolydIdC were used. Interestingly, it was observed that polydIdC elutes almost completely in the flowthrough sample and, as a consequence, when flowthrough had to be rechromatographied, no additional polydIdC was necessary.

4.4 NFE-3 is a "bona fide" hetero(di)meric protein complex

Biochemical analyses shown that, in addition to the well known NFE-3 binding complexes (named 2 and 3), a slower migrating and specific NFE-3 complex (named 1) was also present. This newly identified complex was constantly seen in crude K562 WCEs (Fig. 8), in the Heparin-Sepharose eluted fractions (Fig. 10) and also in the active samples eluted from Sdx-200 column (Fig. 12 and 13) when tested in EMSA with both γd- and ε-CCAAT probes.

The observation that the mobility of the original NFE-3 complexes was significantly faster than that of the NFE-3/band 1 complex is consistent with a scenario in which the molecular weight of the latter protein/s is higher than that of the protein/s involved in complexes 2 and 3. Besides, in gel filtration chromatography, band 1 activity elutes like a 400 KDa native protein complex.

The possibility that band 1 results from a protein-protein interaction and therefore from at least a heterodimeric complex bound to the consensus DNA, came also from the *in vitro* complementation result (Fig. 18).

The heaviest (band 1) and the lightest (band 2 and 3) complexes shown typical NFE-3 binding specificity but *in vitro* band 1 seems to bind less efficiently than the other complexes to the canonical site.

The distal CCAAT box has been considered an essential element involved in the transcriptional regulation of the γ-globin genes, and a region in which crucial interactions between *trans*-acting factors might take place.

The -117 HPFH γ promoter might be considered a different multicomponent structure in which protein-DNA and protein-protein interactions modulating the γ-globin genes silencing are critically affected.
In transgenic mice, the -117 HPFH point mutation leads to a HPFH phenotype, while point mutations within this region that singly affect NFE-3 or GATA-1 binding do not result in such a phenotype (see later).

The finding that NFE-3 can recognize and bind its consensus sequence also as a heterodimeric protein complex, lead to hypothesize that:

i) it might bind the distal CCAAT box in conjuction with other factor/s to modulate the human γ-genes
ii) for such a reason, the -117 HPFH γ-promoter cannot be replaced in transgenic mice studies by a γ-promoter carrying a point mutation that singly abolished only the NFE-3 *in vitro* binding, considering the non-identity between these two modified promoters

Following the second purification scheme (as detailed above), two polypeptides of 40- and 120-KDa were identified.

Both copurified with the NFE-3 activity and were the most abundant proteins observed when active fractions were analysed by silver staining (Fig. 35).
Anyway, UV-crosslinking between affinity-purified NFE-3 samples and either end-labeled or uniformly-labeled consensus site ONs gave rise to a pattern of labeled proteins that perfectly agrees only with the molecular weight of the larger polypeptide. The putative p40-related NFE-3 protein migrated in both experiments slightly above the expected molecular weight. The reason for such a behaviour is still unknown. It could be due to an interference of DNA probe during the separation on SDS-PAGE, but it could also be simply explained with a non real partecipation of p40 in DNA binding.
As shown by southwestern blot experiment, indeed, p120 but not p40 binds by itself to DNA (Fig. 38).

Whether or not p40, even without directly contacting DNA, plays a role in NFE-3 activity cannot be surely assessed until cloning of its cDNA.

The affinity-purified proteins maintain the *in vitro* binding behaviour observed and well characterized in K562 whole cell extracts; the only difference consists in the easier displacement of their binding by one of the HPFH point mutant ON (-117HPFH). As already considered in Chapter 3, this could be due to fractionation of some cofactor(s) dispensable for NFE-3 binding activity, but necessary for binding specificity.

Finally, considering either the different molecular size (in SDS-PAGE) and DNA binding activity (in EMSA) or the lack of any cross-reactivity with specific antibodies, the presence of Ku and YY1 proteins in the affinity-purified NFE-3, can be definitely excluded.

4.5 *In vivo* relevance of NFE-3 protein

4.5.1 Transfection data

We have previously shown (238) that a mutation in the ε-globin CCAAT box that completely abolishes NFE-3 binding, results in the almost complete inactivation of the ε-globin promoter in transfected K562 cells; however, a different mutation that also abolishes NFE-3 binding but allows substantial CP1/NF-Y binding results in essentially normal activity of the promoter. These data suggest that, in the K562 environment, both NFE-3 and CP1/NF-Y may act as positive factors.

The finding that NFE-3 activity in transient transfection experiments does not correlate with the genetic data is intriguing as it could reveal more sophisticated mechanisms underlyining the switching. As already mentioned, NFE-3 - that has been characterized in an embryonic/fetal cellular type - might be a binding activity that only in the adult erythroid environment can interact specifically with other proteins (transcription factors or not) or be subjected to specific post-translational modification; as well as, NFE-3 might be finely modulated in its concentration (e.g. by compartimentalization within the cell) or even its DNA-binding activity might be differentially modulated by defined external stimuli in particular physiologic circumstances. On this light, it is possible that NFE-3 could share homologies with families of other proteins like CREB and Stat.

4.5.2 Transgenic mice

While a single point mutation at -117 in the duplicated CCAAT box region of the $^{a}\gamma$-globin gene promoter associated with Greek HPFH (240, 241) is sufficient to produce a HPFH phenotype in transgenic mice (121), specific disruptions of either GATA-1 or NFE-3 sites within this region failed to give a HPFH phenotype in adult transgenic mice (239). These evidences do not support a direct role for either of these factors as suppressors of γ-globin gene transcription through binding to the CCAAT box region, contrasting the *in vitro* data. However, these findings do not exclude the involvement of GATA-1 and NFE-3 in γ-globin gene regulation. They only reinforce the current model by which γ genes silencing is mediated by an accumulation of weak effects spread across its promoter, with the CCAAT box region being one participating in this effect.

Conceivably, NFE-3 is not the only factor implicated in modulating the γ- to β-globin switch, but being its binding abolished by a variety of HPFH mutations that encompass its canonical binding site, NFE-3 will be an important tool to recognize the nature of the other factors involved in the switch.

Communication between *cis* elements situated many Kb apart, seems to be achieved through physical protein-protein interactions. Erythroid-specific, ubiquitous or even non-DNA binding adaptor molecule(s) might physically link promoter associated*trans* acting factor(s) to LCR-bound protein(s).

Transcription factors which bind DNA as obligate dimers are subjected to an additional levels of regulation. In fact, by contacting or dimerizing with different partners under diverse circumstances, a limited repertoire of proteins may be mixed and matched to assume a greater role in gene regulation. Variation in dimer pairing may differentially affect target gene transcription, titrate complex formation, allow recognition of different DNA-binding sites, promote interaction with different coactivators and co-repressors or may affect the subcellular localization.

In light of this view and considering the importance of protein-protein interaction in establishing and maintaining functional association between *cis* elementes within each promoter and the globin LCR, the possibility that NFE-3 activity might be due to a heterodimeric complex is really intriguingly.
This conclusion is, so far, supported only by biochemical evidences and it will be verified only with the cloning of the corresponding cDNAs.

4.6 Conclusion and perspectives

The characterization of NFE-3 and the molecular cloning of the cDNA encoding its specific component(s) will be a significant step toward an analysis of γ-globin gene regulation. With gene targeting in mouse embryonic stem cells and the methods available for *in vitro* differentiation of hematopoietic cells, it will be possible to initiate a formal genetic analysis of the function on NFE-3 in erythroid cell development.

REFERENCES

1. Grosveld, F., Blom van Assendelft, G., Greaves, D.R., and Kollias, G. (1987) Position-independent, high-level expression of the human beta-globin gene in transgenic mice *Cell* 51, 975-985.
2. Blom van Assendelft, G., Hanscombe, O., Grosveld, F., and Greaves, D. (1989) The beta-globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner *Cell* 56, 969-977.
3. Stammatoyannopoulos, G., and Nienhuis, A. (1994) Hemoglobin switching in *The molecular basis of blood diseases* (Stammatoyannopoulos, G., Nienhuis, A., Majerus, P., and Varmus, P., eds) pp 107-155.
4. Grosveld, F., Antoniou, M., Berry, M., de Boer, E., Dillon, N., Ellis, J., Fraser, P., Hurst, J., Imam, A., Meijer, D., Philipsen, S., Pruzina, S., Strouboulis, J., and Whyatt, D. (1993) Regulation of human globin gene switching *Cold Spring Harbor Symposia on Quantitative Biology* Vol. LVIII pp 7-13.
5. Farace, M.G., Brown, B.A., Raschella, G., Alexander, J., Gambari, R., Fantoni, A., Hardies, S.C., Hutchinson III, C.A., and Edgell, M.H. (1984) The mouse beta-h1 gene codes for the z chain of embryonic hemoglobin *J. Biol. Chem.* 259, 7123-?.
6. Chada, K., Magram, J., and Costantini, F. (1986) An embryonic pattern of expression of a human fetal globin gene in transgenic mice *Nature* 319, 685-689.
7. Peterson, K.R., Clegg, C.H., Huxley, C., Josephson, B.M., Haugen, H.S., Furukawa, T., and Stamatoyannopoulos, G. (1993) Transgenic mice containing a 248-Kb artificial chromosome carrying the human beta-globin locus display proper developmental control of human globin genes *Proc. Natl. Acad. Sci. USA* 90, 7593-7597.
8. Dillon, N., and Grosveld, F. (1993) Transcriptional regulation of multigene loci: multilevel control *Trends Genet.* 9, 134-137.
9. Engel, J.D. (1993) Developmental regulation of human beta-globin gene: a switch of loyalties? *Trends Genet.* 9, 304-309.
10. Forrester, W.C., Thompson, C., Edler, J.T., and Groudine, M. (1986) A developmentally stable chromatin structure in the human beta-globin gene cluster *Proc. Natl. Acad. Sci. USA* 83, 1359-1363.
11. Forrester, W.C., Takegawa, S., Papayannopoulou, T., Stamatoyannopoulos, G., and Groudine, M. (1987) Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids *Nucleic Acids Res.* 15, 10159-10177.
12. Tuan, D., Solomon, W., Li, Q., and London, I.M. (1985) The "beta-like globin" gene domain in human erythroid cells. *Proc. Natl. Acad. Sci. USA* 82, 6384-6388.
13. Forrester, W.C., Epner, E., Driscoll, M.C., Enver, T., Brice, M., Papayannopoulou, T., and Groudine, M. (1990) A deletion of the human beta-globin locus activation region causes a major alteration in chromatin structure and replication across the entire beta-globin locus *Genes &Dev.* 4, 1637-1649.
14. Ryan, T.M., Behringer, R.R., Townes, T.M., Palmiter, R.D., and Brinster, R.L. (1989) High-level erythroid expression of human alpha-globin genes in transgenic mice *Proc. Natl. Acad. Sci. USA* 86, 37-41.
15. Hanscombe, O., Vidal, M., Kaeda, J., Luzzatto, L. Greaves, D.R., and Grosveld, F. (1989) High-level, erythroid-specific expression of the human alpha-globin gene

in transgenic mice and the production of human hemoglobin in murine erythrocytes *Genes & Dev.* 3, 1572-1581.
16. Talbot, D., Collins, P., Antoniou, M., Vidal, M., Grosveld, F., and Greaves, D.R. (1989) A dominant control region from the human beta-globin locus conferring integration site-independent gene expression *Nature* 338, 352-355.
17. Forrester, W.C., Novak, U., Gelinas, R., and Groudine, M. (1989) Molecular analysis of the human beta-globin locus activation region *Proc. Natl. Acad. Sci. USA* 86, 5439-5443.
18. Fraser, P., Hurst, J., Collis, P., and Grosveld, F. (1990) DNaseI hypersensitive site 1, 2, and 3 of the human beta-globin dominant control region direct position-independent expression *Nucleic Acids Res.* 18, 3503-3508.
19. Caterina, J.J., Ryan, T.M., Pawlik, K.M., Palmiter, R.D., Brinster, R.L., Behringer, R.R., and Townes, T.M. (1991) Human beta-globin locus control region: analysis of the 5' DNase I hypersensitive site HS2 in transgenic mice *Proc. Natl. Acad. Sci. USA* 88, 1626-1630.
20. Ney, P.A., Sorrentino, B.P., McDonagh, K.T., and Nienhuis, A.W. (1990) Tandem AP-1-binding sites within the human beta-globin dominant control region function as an inducible enhancer in erythroid cells *Genes & Dev.* 4, 993-1006.
21. Talbot, D., Philipsen, S., Fraser, P., and Grosveld, F. (1990) Detailed analysis of the site 3 region of the human beta-globin dominant control region *EMBO J.* 9, 2169-2178.
22. Talbot, D., and Grosveld, F. (1991) The 5'HS2 of the globin locus control region enhances transcription through the interaction of a multimeric complex binding at two functionally distinct NF-E2 binding sites *EMBO J.* 10, 1391-1398.
23. Tuan, D., Kong, S., and Hu, K. (1992) Transcription of the hypersensitive site HS2 enhancer in erythroid cells *Proc. Natl. Acad. Sci. USA* 89, 11219-11223.
24. Tuan, D.Y.H., Solomon, W.B., London, I.M., and Pee, D.P. (1989) An erythroid-specific, developmental-stage-indipendent enhancer far upstream of the human "beta-like globin" genes *Proc. Natl. Acad. Sci. USA* 86, 2554-2558.
25. Ellis, J., Talbot, D., Dillon, N., and Grosveld, F. (1993) Synthetic human beta-globin 5'HS2 constructs function as locus control regions only in multicopy transgene concatamers *EMBO J.* 12, 127-134.
26. Kim, C.G., Epner, E.M., Forrester, W.C., and Groudine, M. (1992) Inactivation of the human beta-globin gene by targeted insertion into the beta-globin locus control region *Genes & Dev.* 6, 928-938.
27. Fiering, S., Kim, C.G., Epner, E.M., and Groudine, M. (1993) An "in-out" strategy using gene targeting and FLP recombinase for the functional dissection of complex DNA regulatory elements: analysis of the beta-globin locus control region *Proc. Natl. Acad. Sci. USA* 90, 8469-8473.
28. Fiering, S., Epner, E., Robinson, K., Zhuang, Y., Telling, A., Hu M., Ming, Martin, D.I.K., Enver, T., Ley, T.J., and Groudine, M. (1995) Targeted deletion of 5' HS2 of the murine beta-globin LCR reveals that it is not essential for proper regulation of the beta-globin locus*Genes & Dev.* 9, 2203-2213.
29. Pruzina, S., Hanscombe, O., Whyatt, D., Grosveld, F., and Philipsen, S. (1991) Hypersensitive site 4 of the human beta globin locus control region *Nucleic Acids Res.* 19, 1413-1419.
30. Hug, B.A., Moon, A.M., and Ley, T.J. (1992) Structure and function of the murine beta-globin locus control region 5' HS-3 *Nucleic Acids Res.* 20, 5771-5778.

31. Ellis, J., Tan-Un, K.C., Harper, A., Michelovich, D., Yannoutsos, N., Philipsen, S., and Grosveld, F. (1996) A dominant chromatin-opening activity in 5' hypersensitive site 3 of the human beta-globin locus control region *EMBO J.* 15, 562-568.
32. Dean, A., Ley, T., Humphries, R., Ferdis, M., and Schechter, A. (1983) Inducible transcription of five globin genes in K562 human leukemia cells *Proc. Natl. Acad. Sci. USA* 80, 5515-5519.
33. Philipsen, S., Pruzina, S., and Grosveld, F. (1993) The minimal requirement for activity in transgenic mice of HS3 of the beta globin LCR *EMBO J.* 12, 1077-1085.
34. Stamatoyannopoulos, J.A., Goodwin, A., Joyce, T., and Lowrey, C.H. (1995) NF-E2 and GATA binding motifs are required for the formation of DNase I hypersensitive site 4 of the human beta-globin locus control region *EMBO J.* 14, 106-116.
35. Papayannopoulou, T., Enver, T., Takegawa, S., Anagnou, N.P., and Stamatoyannopoulos, G. (1988) Activation of developmentally mutated human globin genes by cell fusion *Science* 242, 1056-1058.
36. Stanworth, S.J., Roberts, N.A., Sharpe, J.A., Sloane-Stanley, J.A., and Wood, W.G. (1995) Established epigenetic modifications determine the expression of developmentally regulated globin genes in somatic cell hybrids *Mol. Cell. Biol.* 15, 3969-3978.
37. Starck, J., Sarkar, R., Romana, M., Bhargava, A., Scarpa, A.L., Tanaka, M., Chamberlain, J.W., Weissman, S.M., and Forget, B.G. (1994) Developmental regulation of human gamma- and beta-globin genes in the absence of the Locus Control Region *Blood* 84, 1656-1665.
38. Arceci, R.J., King, A.A.J., Simon, M.C., Orkin, S.-H., and Wilson, D.B. (1993) Mouse GATA-4: a retinoic acid-inducible GATA-binding transcription factor expressed in endodermally derived tissues and heart *Mol. Cell. Biol.* 13, 2235-2246.
39. Kollias, G., Wrighton, N., Hurst, J., and Grosveld, F. (1986) Regulated expression of the human A-gamma, beta, and hybrid gamma/beta globin genes in trangenic mice: manipulation of the developmental expression patterns *Cell* 46, 89-94.
40. Magram, J., Chada, K., and Costantini, F. (1985) Developmental regulation of a cloned adult beta-globin gene in transgenic mice *Nature* 315, 338-340.
41. Townes, T.M., Lingrel, J.B., Chen, H.Y., Brinster, R.L., and Palmiter, R.D. (1985) Erythroid-specific expression of human beta-globin genes in transgenic mice *EMBO J.* 4, 1715-1723.
42. Raich, N., Enver, T., Nakamoto, B., Josephson, B., Papayannopoulou, T., and Stamatoyannopoulos, G. (1990) Autonomous developmental control of human embryonic globin gene switching in transgenic mice *Science* 250, 1147-1149.
43. Shih, D.M., Wall, R.J., and Shapiro, S.G. (1990) Developmentally regulated and erythroid-specific expression of the human embryonic beta-globin gene in transgenic mice *Nucleic Acids Res.* 18, 5465-5472.
44. Perez-Stable, C. (1994) Distinct negative regulatory mechanisms involved in the repression of human embryonic epsilon- and fetal gamma-globin genes intransgenic mice *J. Biol. Chem.* 269, 33109-33115.
45. Enver, T., Ebens, A.J., Forrester, W.C., Stamatoyannopoulos, G. (1989) The human beta-globin locus activation region alters the developmental fate of a human fetal globin gene in transgenic mice *Proc. Natl. Acad. Sci. USA* 86, 7033-7037.

46. Behringer, R.R., Ryan, T.M., Palmiter, R.D., Brinster, R.L., and Townes, T.M. (1990) Human gamma- to beta-globin switching in transgenic mice *Genes &Dev.* 4, 380-389.
47. Strouboulis, J., and Grosveld, F. (1992) Developmental regulation of a complete 70-Kb human beta-globin locus in transgenic mice *Genes & Dev.* 6, 1857-1864.
48. Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N., and Grosveld, F. (1991) Importance of globin gene order for correct developmental expression *Genes & Dev.* 5, 1387-1394.
49. Fraser, P., Pruzina, S., Antoniou, M., and Grosveld, F. (1993) Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression to the globin genes *Genes & Dev.* 7, 106-113.
50. Peterson, K.R., and Stamatoyannopoulos, G. (1993) Role of gene order in developmental control of human gamma- and beta-globin gene expression *Mol. Cell. Biol.* 13, 4836-4843.
51. Choi, O.-R., and Engel, J.D. (1988) Developmental regulation of beta-globin gene switching *Cell* 55, 17-26.
52. Foley, K.P., and Engel, J.D. (1992) Individual stage selector element mutations lead to reciprocal changes in beta- versus epsilon-globin gene transcription: genetic confirmation of promoter competition during globin gene switching *Genes & Dev.* 6, 730-744.
53. Gallarda, J.L., Foley, K.P., Yang, Z., and Engel, J. (1989) The beta-globin stage selector element factor is erythroid-specific promoter/enhancer binding protein NF-E4 *Genes & Dev.* 3, 1845-1859.
54. Enver, T., Raich, N., Ebens, A.J., Papayannopoulou, T., Costantini, F., and Stamatoyannopoulos, G. (1990) Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice *Nature* 344, 309-313.
55. Dillon, N., and Grosveld, F. (1991) Human gamma-globin genes silenced independently of other genes in the beta-globin locus *Nature* 350, 252-254.
56. Ptashne, M. (1988) How eukaryotic transcriptional activators work *Nature* 335, 683-689.
57. Travers, A.A. (1992) The reprogramming of transcriptional competence *Cell* 69, 573-575.
58. Wijgerde, M., Grosveld, F., and Fraser, P. (1995) Transcription complex stability and chromatin dynamics in vivo *Nature* 377, 209-213.
59. Jane, S.M., Ney, P.A., Vanin, E.F., Gumucio, D.L., and Nienhuis, A.W. (1992) Identification of a stage selector element in the human gamma-globin gene promoter that fosters preferential interaction with the 5' HS2 enhancer when in competition with the beta-promoter *EMBO J.* 11, 2961-2969.
60. Bottardi, S., and Santoro, C. (1995) CTF/NF-1 binds the stage selector element of the human gamma-globin gene promoter *Biochem. Bioph. Res. Comm.* 215, 874-880.
61. Gumucio, D.L., Heilstedt-Williamson, H., Gray, T.A., Tarlé, S.A., Shelton, D.A., Tagle, D.A., Slightom, J.L., Goodman, M., and Collins, F.S. (1992) Phylogenetic footprinting reveals a nuclear protein which binds to silencer sequences in the human gamma and epsilon globin genes *Mol. Cell. Biol.* 12, 4919-4929.
62. Jane, S.M., Nienhuis, A.W., and Cunningham, J.M. (1995) Hemoglobin switching in man and chicken is mediated by a heterodimeric complex between

the ubiquitous transcription factor CP2 and a developmentally specific protein *EMBO J.* 14, 97-105.
63. Gumucio, D.L., Rood, K.L., Blanchard-Mac Quate, K., Gray, T.A., Saulino, A., and Collins, F. (1991) Interaction of Sp1 with the human gamma-globin promoter: binding and transactivation of normal and mutant promoters *Blood* 78, 1853-1863.
64. Townes, T.M., and Behringer, R. (1990) Human globin locus activation region (LAR): role in temporal control *Trends Genet.* 6, 219-223.
65. Orkin, S.-H. (1995) Regulation of globin gene expression in erythroid cells *Eur. J. Biochem.* 231, 271-281.
66. Evans, T., and Felsenfeld, G. (1989) The erythroid-specific transcription factor Eryf-1: a new finger protein *Cell* 58, 877-885.
67. Martin, D.I.K., Tsai, S.-F., and Orkin, S.-H. (1989) Increased gamma-globin expression in a nondeletion HPFH mediated by an erythroid-specific DNA-binding factor *Nature* 338, 435-438.
68. Wall, L., de Boer, E., and Grosveld, F. (1988) The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein *Genes & Dev.* 2, 1089-1100.
69. Tsai, S.-F., Martin, D.I.K., Zon, L.I., D'Andrea, A.D., Wong, G.G., and Orkin, S.-H. (1989) Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells *Nature* 339, 446-451.
70. Omichinski, J.G., Trainor, C., Evans, T., Gronenborg, A.M., Clore, G.M., and Felsenfeld, G. (1993) A small single-"finger" peptide from the erythroid transcription factor GATA-1 binds specifically to DNA as a zinc or iron complex *Proc. Natl. Acad. Sci. USA* 90, 1676-1680.
71. Merika, M., and Orkin, S.-H. (1995) Functional sinergy and physical interactions of the erythroid transcription factor GATA-1 with the Kruppel family proteins Sp1 and EKLF *Mol. Cell. Biol.* 15, 2437-2447.
72. Crossley, M., Merika, M., and Orkin, S.-H. (1995) Self-association of the erythroid transcription factor GATA-1 mediated by its zinc finger domains *Mol. Cell. Biol.* 15, 2448-2456.
73. Yang, H.-Y., and Evans, T. (1995) Homotypic interactions of chicken GATA-1 can mediate transcriptional activation *Mol. Cell. Biol.* 15, 1353-1363.
74. Whyatt, D.J., deBoer, E., and Grosveld, F. (1993) The two zin finger-like domains of GATA-1 have different DNA binding specificities *EMBO J.* 12, 4993-5005.
75. Zon, L.I., Mather, C., Burgess, S., Bolce, M.E., Harland, R.M., and Orkin, S.-H. (1991) Expression of GATA-binding proteins during embryonic development in Xenopus laevis *Proc. Natl. Acad. Sci. USA* 88, 10642-10646.
76. Zon, L., Yamaguchi, Y., Yee, K., Albee, E.A., Bennet, J.C., Orkin, S.-H., and Ackerman, S.J. (1993) Expression of mRNA for the GATA-binding proteins in human eosinophils and basophils: potential role in gene transcription *Blood* 81, 3234-3241.
77. Martin, D.I.K., Zon, L.I., Mutter, G., and Orkin, S.-H. (1990) Expression of an erythroid transcription factor in megakaryocytic and mast cell lineages *Nature* 344, 444-447.
78.Romeo, P.-H., Prandini, M.-H., Joulin, V., Mignotte, V., Penant, M., Vainchenker, W., Marguerie, G., and Uzan, G. (1990) Megakaryocytic and erythrocytic lineages share specific transcription factors *Nature* 344, 447-449.

79. Ito, E., Toki, T., Ishihara, H., Ohtani, H., Gu, L., Yokoyama, M., Engel, J.D., and Yamamoto, M. (1993) Erythroid transcription factor GATA-1 is abundantly transcribed in mouse testis *Nature* 362, 466-468.
80. Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.-F., D'Agati, V., Orkin, S.-H., and Costantini, F. (1991) Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1 *Nature* 349, 257-260.
81. Simon, M.C., Pevny, L., Wiles, M.V., Keller, G., Costantini, F., and Orkin, S.-H. (1992) Rescue of erythroid development in gene targeted GATA-1 mouse embryonic stem cells *Nature Genetics* 1, 92-98.
82. Weiss, M.J., and Orkin, S.-H. (1995) Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis *Proc. Natl. Acad. Sci. USA* 92, 9623-9627.
83. Bottardi, S., Calligaris, R., Cogoi, S., Apezteguia, I., and Santoro, C. (1995) Alternative translation initiation site usage results in two functionally distinct forms of the GATA-1 transcription factor *Proc. Natl. Acad. Sci. USA* 92, 11598-11602.
84. Tsai, F.-Y., Keller, G., Kuo, F.C., Weiss, M., Chen, J., Rosenblatt, M., Alt, F.W., and Orkin, S.-H. (1994) An early haematopoietic defect in mice lacking the transcription factor GATA-2 *Nature* 371, 221-226.
85. Trainor, C.D., Evans,T., Felsenfeld, G., and Boguski, M.S. (1990) Structure and evolution of a human erythroid transcription factor *Nature* 343, 92-96.
86. Weiss, M.J., and Orkin, S.-H. (1995) GATA transcription factors: key regulators of hematopoiesis *Exp. Hematol.* 23, 99-107.
87. Orkin, S.-H. (1990) Globin gene regulation and switching: circa 1990 *Cell* 63, 665-672.
88. Yamamoto, M., Ko, L.J., Leonard, M.W., Beug, H., Orkin, S.-H., and Engel, J.D. (1990) Activity and tissue-specific expression of the transcriptional factor NF-E1 multigene family *Genes &Dev.* 4, 1650-1662.
89. Orkin,S.-H. (1992) GATA-binding transcription factors in hematopoietic cells *Blood* 80, 575-581.
90. Laverriere, A.C., MacNeill, C., Mueller, C., Polemann, R.E., Burch, J.E., and Evans, T. (1994) GATA-4/5/6, a subfamily of three transcription factors transcribed in developing heart and gut *J. Biol. Chem.* 269, 23177-23184.
91. Andrews, N.C., Erjument-Bromage, H., Davidson, M.B., Tempst, P., and Orkin, S.-H. (1993) Erythroid transcription factor (NF-E2) is a hematopoietic-specific basic-leucine zipper protein *Nature* 362, 722-728.
92. Andrews, N.C., Kutkow, K.J., Ney, P.A., Erdjument-Bromage, H., Tempst, P., and Orkin, S.-H. (1993) The ubiquitoous subunit of erythroid transcription factor NF-E2 is a small basic-leucin zipper protein related to the v-maf oncogene *Proc. Natl. Acad. Sci. USA* 90, 11488-11492.
93. Ney, P.A., Andrews, N.C., Jane, S.M., Safer, B., Purucker, M.E., Weremowicz, S., Morton, C.C., Goff, S.C., Orkin, S.-H., and Nienhuis, A.W. (1993) Purification of the human NF-E2 complex: cDNA cloning of the hematopoietic cell-specific subunit and evidence for an associated partner *Mol. Cell. Biol.* 13, 5604-5612.
94. Stamatoyannopoulos, J.A., Goodwin, A., Joyce, T., and Lowrey, C.H. (1995) NF-E2 and GATA binding motifs are required for the formation of DNase I hypersensitive site 4 of the human beta-globin locus control region *EMBO J.* 14, 106-116.

95. Shivdasani, R.A., and Orkin, S.-H. (1995) Erythropoiesis and globin gene expression in mice lacking the transcription factor NF-E2 *Proc. Natl. Acad. Sci. USA* 92, 8690-8694.
96. Feng, W.C., Southwood, C.M., and Bieker, J.J. (1994) Analyses of beta-thalassemia mutant DNA interactions with erythroid kruppel-like factor (EKLF), an erytroid cell-specific transcription factor *J. Biol. Chem.* 269, 1493-1500.
97. Miller, L.J., and Bieker, J.J. (1993) 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, 2776-2786.
98. Perkins, A.C., Sharpe, A.H., and Orkin, S.-H. (1995) Lethal beta-thalassemia in mice lacking the erythroid CACCC transcription factor EKLF *Nature* 375, 318-322.
99. Wijgerde, M., Gribnau, J., Trimborn, T., Nuez, B., Philipsen, S., Grosveld, F., and Fraser, P. (1996) The role of EKLF in human beta-globin gene competition *Genes & Dev.* 10, 2894-2902.
100. Perkins, A.C., Gaensler, K.M.L., and Orkin, S.-H. (1996) Silencing of human fetal globin expression is impaired in the absence of the adult beta-globin gene activator protein EKLF *Proc. Natl. Acad. Sci. USA* 93, 12267-12271.
101. Nuez, B., Michalovic, D., Bygrave, D., Ploemacher, R., and Grosveld, F. (1995) Defective hematopoiesis in fetal liver resulting from inactivation of the EKLF gene *Nature* 375, 316-318.
102. Bockamp, E.-O., McLaughlin F., Murrel, A., and Green, A.R. (1994) Transcription factors and the regulation of hemopoiesis: lessons from GATA and SCL proteins *BioEssays* 16, 481-488.
103. Green, A.R., Lints, T., Visvader, J., Harvey, R., and Begley, C.G. (1992) SCL is coexpressed with GATA-1 in hemopoietic cells but it is also expressed in developing brain *Oncogene* 7, 653-660.
104. Begley, C.G., Aplan, P.D., Denning, S.M., Haynes, B.F., Waldmann, T.A., and Kirsch, I.R. (1989) The gene SCL is expressed during early hematopoiesis and encodes a differentiation-related DNA-binding motif *Proc. Natl. Acad. Sci. USA* 86, 10128-10132.
105. Bash, R., Halls, S., Timmons, C., Crist, W., Amylon, M., Smith, R., and Baer, R. (1995) Does activation of the TAL1 gene occur in a majority of patients with T-cell acute lymphoblastic leukemia? A pediatric oncology group study *Blood* 86, 666-676.
106. Begley, C.G., Robb, L., Rockman, S., Visvader, J., and Green, A.R. (1994) Structure of the gene encoding the murine SCL protein *Gene* 138, 93-99.
107. Cheng, J.-T., Cobb, M.H., and Baer, R. (1993) Phosphorylation of the TAL-1 oncoprotein by the extracellular signal-regulated protein kinase ERK1 *Mol. Cell. Biol.* 13, 801-808.
108. Robb, L., Lyons, I., Li, R., Hartley, L., Harvey, R.P., Kontgen, F., and Begley, C.G. (1995) Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene *Proc. Natl. Acad. Sci. USA* 92, 7075-7079.
109. Shivdasani, R.A., Mayer, E.L., and Orkin, S.-H. (1995) Absence of blood formation in mice lacking the T-cell leukemia oncoprotein tal-1/SCL *Nature* 373, 432-434.
110. Robb, L., Elwood, N.J., Elefanty, A.G., Kontgen, F., Li, R., Barnett, L.D., and Begley, C.G. (1996) The scl gene product is required for the generation of hematopoietic lineages in the adult mouse *EMBO J.* 15, 4123-4129.

111. Aplan, P.D., Begley, C.G., Bertness, V., Nussmeier, M.A., Ezguerra, A., Coligan, J., and Kirsch, I.R. (1990) The SCL gene is formed from a transcriptionally complex locus *Mol. Cell. Biol.* 10, 6426-6435.
112. Aplan, P.D., Lombardi, D.P., Reaman, G.H., Sather, H.N., Hammond G.D., and Kirsch, I.R. (1992) Involvement of the putative hemopoietic transcription factor SCL in T-cell acute lymphoblastic leukemia *Blood* 79, 1327-1333.
113. Jackson, S.P., MacDonald, J.J., Lees-Miller, S., and Tjian, R. (1990) GC box binding induces phosphorylation of Sp1 by a DNA-dependent protein kinase *Cell* 63, 155-165.
114. Briggs, M.R., Kadonaga, J.T., Bell, S.P., and Tjian, R. (1986) Purification and biochemical characterization of the promoter-specific transcription factor, Sp1 *Science* 234, 47-52.
115. Courey, A.J., and Tjian, R. (1988) Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation domain *Cell* 55, 887-898.
116. Kadonaga, J.T., Carner, K.R., Masiarz, F.R., and Tjian, R. (1987) Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain *Cell* 51, 1079-1090.
117. Chan, J.Y., Han, X.-L., and Kan, Y.K. (1993) Cloning of Nrf1, an NF-E2-related transcription factor, by genetic selection in yeast *Proc. Natl. Acad. Sci. USA* 90, 11371-11375.
118. Caterina, J.J., Donze, D., Sun, C.-W., Ciavatta, D.J., and Townes, T.M. (1994) Cloning and functional characterization of LCR-F1: a bZIP transcription factor that activates erythroid-specific, human globin gene expression *Nucleic Acids Res.* 22, 2383-2391.
119. Moi, P., Chan, K., Asunis, I., Cao, A., and Kan, Y.W. (1994) Isolation of NF-E2-related factor 2 (Nrf2), a Nf-E2-like basic leucin zipper transcriptional activator that binds to the tandem NF-E2/AP-1 repeat of the beta-globin locus control region *Proc. Natl. Acad. Sci. USA* 91, 9926-9930.
120. Lam, L.T.L., and Bresnick, E. (1996) A novel DNA-binding protein, HS2NF5, interacts with a functionally important sequence of the human beta-globin locus control region *J. Biol. Chem.* 271, 32421-32429.
121. Berry, M., Grosveld, F., and Dillon, N. (1992) A single point mutation is the cause of the Greek form of hereditary persistence of fetal hemoglobin *Nature* 358, 499-502.
122. Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J., and Tjian, R. (1987) A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication *Cell* 48, 79-89.
123. Fischer, K.-D., Haese, A., and Nowock, J. (1993) Cooperation of GATA-1 and Sp1 can result in synergistic transcriptional activation or interference *J. Biol. Chem.* 268, 23915-23923.
124. Dorn, A., Bollekens, J., Staub, A., Benoist, C., and Mathis, D. (1987) A multiplicity of CCAAT box-binding proteins *Cell* 50, 863-872.
125. Cao, Q.P., Pitt, S., Laeszyk, J., and Baril, E.F. (1994) DNA-dependent ATPase from HeLa cells is related to human Ku autoantigen *Biochem.* 33, 8548-8557.
126. Chodosh, L.A., Olesen, J., Hahn, S., Baldwin, A.S., Guarante, L., and Sharp, P. (1988) A yest and a human CCAAT-binding protein have heterologous subunits that are functionally interchangeable *Cell*, 53, 25-35.

127. deBoer, E., Antoniou, M., Mignotte, V., Wall, L., and Grosveld, F. (1988) The human beta-globin promoter; nuclear protein factors and erythroid specific induction of transcription *EMBO J.* 13, 4203-4212.
128. vanHuijsduijnen, R.H., Li, X.Y., Black, D., Matthes, H., Benoist, C., and Mathis, D. (1990) Co-evolution from yeast to mouse: cDNA cloning of the two NF-Y (CP-1/NF-Y) subunits *EMBO J.* 9, 3119-3127.
129. Barberis, A., Superti-Furga, G., and Busslinger, M. (1987) Mutually exclusive interaction of the CCAAT-binding factor and of a displacement protein with overlapping sequences of a histone gene promoter *Cell* 50, 347-359.
130. Superti-Furga, G., Barberis, A., Schaffner, G., and Busslinger, M. (1988) The -117 mutation in Greek HPFH affects the binding of three nuclear factors to the CCAAT region of the gamma-globin gene *EMBO J.* 7, 3099-3107.
131. Shi, Y., Seto, E., Chang, L.-S., and Shenk, T. (1991) Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and relief of repression by adenovirus E1A protein *Cell* 67, 377-388.
132. Park, K., and Atchinson, M.L. (1991) Isolation of a candidate repressor/activator, NF-E1 (YY1, δ), that binds to the immunoglobulin k3' enhancer and the immunoglobulin heavy-chain μE1 site *Proc. Natl. Acad. Sci. USA* 88, 9804-9808.
133. Hariharan, N., Kelley, D.E., and Perry, R. (1991) δ, a transcription factor that binds to downstream elements in several polymerase II promoters, is a functionally versatile zinc finger protein *Proc. Natl. Acad. Sci. USA* 88, 9799-9803.
134. Flanagan, J.R., Becker, K.G., Ennist, D.L., Gleason, S.L., Driggers, P.H., Levi, B.-Z., Appella, E., and Ozato, K. (1992) Cloning of a negative transcription factor that binds to the upstream conserved region of moloney murine leukemia virus *Mol. Cell. Biol.* 12, 38-44.
135. Guo, B., Odgren, P.R., van Wisnen, A.J., Last, T.J., Nickerson, J., Penman, S., Lian, J.B., Stein, J.L., and Stein, G. (1995) The nuclear matrix protein NMP-1 is the transcriptional factor YY1 *Proc. Natl. Acad.Sci. USA* 92, 10526-10530.
136. Yant, S.R., Zhu, W., Millinoff, D., Slightom, J.L., Goodman, M., and Gumucio, D.L. (1995) High affinity YY1 binding motifs: identification of two core types (ACAT and CCAT) and distribution of potential binding sites within the human beta globin cluster *Nucleic Acids Res.* 23, 4353-4362.
137. Hyde-DeRuyscher, R., Jennings, E., and Shenk, T. (1995) DNA binding sites for the transcriptional activator/repressor YY1 *Nucleic Acids Res.* 23, 4457-4465.
138. Collins, F.S., and Weissman, S.M. (1984) The molecular genetics of human hemoglobin *Prog. Nucleic Acid Res. Mol. Biol.* 31, 315-462.
139. Bauknecht, T., Angel, P., Royer, H.D., and Hausen, H.Z. (1992) Identification of a negative regulatory domain in the human papillomavirus type 18 promoter: interaction with the transcriptional repressor YY1 *EMBO J.* 11, 4607-4617.
140. May, M., Dong, X.-P., Beyer-Finkler, E., Stubenrauch, F., Fuchs, P.G., and Pfister, H. (1994) The E6/E7 promoter of extrachromosomal HPV16 DNA in cervical cancer escapes from cellular repression by mutation of target sequences for YY1 *EMBO J.* 13, 1460-1466.
141. Gumucio, D.L., Shelton, D.A., Bailey, W.J., Slightom, J.L., and Goodman, M. (1993) Phylogenetic footprinting reveals unexpected complexity in trans factor binding upstream from the epsilon-globin gene *Proc. Natl. Acad. Sci. USA* 90, 6018-6022.

142. Weatherall, D.J. (1994) The thalassemias in *The molecular basis of blood diseases* (Stammatoyannopoulos, G., Nienhuis, A., Majerus, P., and Varmus, P., eds) pp 157-206.
143. Lee, T.-C., Shi, Y., and Schwartz, R.J. (1992) Displacement of BrdUrd-induced YY1 by serum response factor activates skeletal alfa-actin transcritpion in embryonic myoblasts *Proc. Natl. Acad. Sci. USA* 89, 9814-9818.
144. Lee, T.-C., Zhang, Y., and Schwartz, R. (1994) Bifunctional transcriptional properties of YY1 in regulating muscle actin and c-myc gene expression during myogenesis *Onccogene* 9, 1047-1052.
145. Gualberto, A., lePage, D., Pons, G., Mader, S.L., Park, K., Atchison, M.L., and Walsh, K. (1992) Functional antagonism between YY1 and serum response factor *Mol. Cell. Biol.* 12, 4209-4214.
146. Raich, N., Clegg, C.H., Groft, J., Romeo, P.-H., and Stamatoyannopoulos, G. (1995) GATA-1 and YY1 are developmental repressors of the human epsilon-globin gene *EMBO J.* 14, 801-809.
147. Peters, B., Merezhinskaya, N., Diffley, J.F.X., and Noguchi, C.T. (1993) Protein-DNA interactions in the epsilon-globin gene silencer *J. Biol. Chem.* 268, 3430-3437.
148. Yost, S.E., Shewchuk, B., and Hardson, R. (1993) Nuclear protein-binding sites in a transcriptional control region of the rabbit alfa-globin gene *Mol. Cell. Biol.*13, 5439-5449.
149. Ye, J., Cippitelli, M., Dorman, L., Ortaldo, J.R., and Young, H.A. (1996) The nuclear factor YY1 suppresses the human gamma interferon promoter through two mechanisms: inhibition of AP1 binding and activation of a silencer element *Mol. Cell. Biol.* 16, 4744-4753.
150. Ye, J., Zhang, X., and Dong, Z. (1996) Characterization of the human granulocyte-macrophage colony-stimulating factor gene promoter: an AP1 complex and an Sp1-related complex transactivate the promoter activity that is suppressed by a YY1 complex *Mol. Cell. Biol.* 16, 157-167.
151. Lu, S.-Y., Rodriguez, M., and Liao, W.S.-L. (1994) YY1 represses rat serum amyloid A1 gene transcription and is antagonized by NF-κB during acute-phase response *Mol. Cell. Biol.* 14, 6253-6263.
152. Chang, L.-S., Shi, Y., and Shenk, T. (1989) Adeno-associated virus P5 promoter contains an adenovirus E1A-inducible element and a binding site for the major late transcription factor *J. Virol.* 63, 3479-3488.
153. Lewis, B.A., Tullis, G., Seto, E., Horikoshi, N., Weinmann, R., and Shenk, T. (1995) Adenovirus E1A proteins interact with the cellular YY1 transcription factor *J. Virol.* 69, 1628-1636.
154. Shrivastava, A., Saleque, S., Kalpana, G.V., Artandi, S., Goff, S.P., and Calame, K. (1993) Inhibition of transcriptional regulator Yin-Yang-1 by association with c-Myc *Science* 262, 1889-1892.
155. Inouye, C.J., and Seto, E. (1994) Relief of YY1-induced transcriptional repression by protein-protein interaction with the nucleolar phoshpoprotein B23 *J. Biol. Chem.* 269, 6506-6510.
156. Riggs, K.J., Saleque, S., Wong, K.-K., Merrel, K.T., Lee, J.-S., Shi, Y., and Calame, K. (1993) Yin-Yang 1 activates the c-myc promoter *Mol. Cell. Biol.* 13, 7487-7495.
157.Ye, J., Young, H.A., Ortaldo, J.R., and Ghosh, P. (1994) Identification of a DNA binding site for the nuclear factor YY1 in the human GM-CSF core promoter *Nucleic Acids Res.* 22, 5672-5678.

158. Seto, E., Lewis, B., and Shenk, T. (1993) Interaction between transcription factors Sp1 and YY1 *Nature* 365, 462-464.
159. Lee, J.-S., Galvin, K.M., and Shi, Y. (1993) Evidence for physical interaction between the zinc-finger transcription factors YY1 and Sp1 *Proc. Natl. Acad. Sci. USA* 90, 6145-6149.
160. Natesan, S., and Gilman, M. (1995) YY1 activates the association of serum response factor with the c-fos serum response element *Mol. Cell. Biol.* 15, 5975-5982.
161. Finnie, N.J., Gottlieb, T.M., Blunt, T., Jeggo, P.A., and Jackson, S.P. (1995) DNA-dependent protein kinase activity is absent in xrs-6 cells: implications for site-specific recombination and DNA double-strand break repair *Proc. Natl. Acad. Sci. USA* 92, 320-324.
162. Seto, E., Shi, Y., and Shenk, T. (1991) YY1 is an initiator sequence-binding protein that directs and activates transcription in vitro *Nature* 354, 241-245.
163. Usheva, A., and Shenk, T. (1994) TATA-binding protein-independent initiation: YY1, TFIIB, and RNA Polymerase II direct basal transcription on supercoiled template DNA *Cell* 76, 1115-1121.
164. Chiang, C.-M., and Roeder, R. (1995) Cloning of an intrinsic human TFIID subunit that interacts with multiple transcriptional activators *Science* 267, 531-536.
165. Carter, T. Vancurova, I., Sun, I., Lou, W., and DeLeon, S. (1990) A DNA-activated protein kinase from HeLa cell nuclei *Mol. Cell. Biol.* 10, 6460-6471.
166. Mimori, T., Akizuki, M., Yamagata, H., Yoshida, S., and Homma, M. (1981) Characterization of a high molecular weight acidic nuclear protein recognized by autoantibodies in sera from patients with polymyositis-scleroderma overlap syndrome *J. Clin. Invest.* 68, 611-620.
167. de Vries, E., van Driel, W., Bergsma, W.G., Arnberg, A.C., and van der Vliet, P. (1989) HeLa nuclear protein recognizing DNA termini and translocating on DNA forming a regular DNA-multimeric protein complex *J. Mol. Biol.* 208, 65-78.
168. Griffith, A.J., Bliert, P.R., Mimori, T., and Hardin, J.A. (1992) Ku polypeptides synthesized in vitro assemble into complexes which recognize ends of double-stranded DNA *J. Biol. Chem.* 267, 331-338.
169. Reeves, W., and Sthoeger, Z.M. (1989) Molecular cloning of cDNA encoding the p70 (Ku) lupus autoantigen *J. Biol. Chem.* 264, 5047-5052.
170. Mimori, T., Ohosone, Y., Hama, N., Suwa, A., Azizuki, M., Homma, M., Griffith, A.J., and Hardin, J. (1880) Isolation and characterization of cDNA encoding the 80-KDa subunit protein of the human autoantigen Ku (p70/p80) recognized by autoantibodies from patients with scleroderma-polymyositis overlap syndrome *Proc. Natl. Acad. Sci. USA* 87, 1777-1781.
171. Yaneva, M., Wen, J., Ayala, A., and Cook, R. (1989) cDNA-derived amino acid sequence of the 86-KDa subunit of the Ku autoantigen *J. Biol. Chem.* 264, 13407-13411.
172. Mimori, T., and Jardin, J.A. (1986) Mechanism of interaction between Ku protein and DNA *J. Biol. Chem.* 261, 10375-10379.
173. Blier, P.R., Griffith, A.J., Craft, J., and Hardin, J.A. (1993) Binding of Ku protein to DNA *J. Biol. Chem.* 268, 7594-7601.
174. Falzon, M., Fewell, J.W., and Kuff, E.L. (1993) EBP-80, a transcription factor closely resembling the human autoantigen Ku, recognizes single- to double-strand transitions in DNA *J. Biol. Chem.* 268, 10546-10552.

175. Paillard, S., and Strauss, F. (1991) Analysis of the mechanism of interaction of simian Ku protein with DNA *Nucleic Acids Res.* 19, 5619-5624.
176. Gottlieb, T.M., and Jackson, S.P. (1993) The DNA-dependent protein kinase: requirement for DNA ends association with Ku antigen *Cell* 72, 131-142.
177. Zhang, W.-W., and Yaneva, M. (1992) On the mechanisms of Ku protein binding to DNA *Biochem. Bioph. Res. Comm.* 186, 574-579.
178. Messier, H., Fuller, T., Mangal, S., Brickner, H., Igarashi, S., Gaikwad, J., Fotedar, R., and Fotedar, A. (1993) p70 lupus autoantigen binds the enhancer of the T-cell receptor beta-chain gene *Proc. Natl. Acad. Sci. USA* 90, 2685-2689.
179. Allaway, G.P., Vivino, A.A., Kohn, L.D., Notkins, A.L., and Prabhakar, B.S. (1990) Characterization of the 70KDa component of the human Ku autoantigen expressed in insect cell nuclei using a recombinant baculovirus vector *Biochem. Bioph. Res. Comm.* 169, 747-755.
180. Chou, C.-H., Wanf, J., Knuth, M.W., and Reeves, W.H. (1992) Role of a major autoepitope in forming the DNA binding site of the p70 (Ku) antigen *J. Exp. Med.* 175, 1677-1684.
181. Wang, J., Satoh, M., Chou, C.-H., Reeves, W.H. (1994) Similar DNA binding properties of free p70 (Ku) subunit and p70/p80 heterodimer *FEBS* 351, 219-224.
182. Mimori, T., and Hardin, J.A. (1992) Ku polypeptides synthesized in vitro assemble into complex which recognizes ends of double-stranded DNA *J. Biol. Chem.* 267, 331-338.
183. Wu, X., and Lieber, M.R. (1996) Protein-protein and protein-DNA interaction regions within the DNA end-binding protein Ku70-Ku86 *Mol. Cell. Biol.* 16, 5186-5193.
184. Getts, R.C., and Stamato, T.D. (1994) Absence of a Ku-like DNA end binding activity in the xrs double-strand DNA repair-deficient mutant *J. Biol. Chem.* 269, 15981-15984.
185. Taccioli, G.E., Gottlieb, T.M., Blunt, T., Priestley, A., Demengeot, J., Mizuta, R., Lehmann, A.R., Alt, F.W., Jackson, S.P., and Jeggo, P.A. (1994) Ku 80: product of the XRCC5 gene and its role in DNA repair and V(D)J recombination *Science* 265, 1442-1445.
186. Smider, V., Rathmell, W.K., Lieber, M.R., and Chu, G. (1994) Restoration of X-ray resistance and V(D)J recombination in mutant cells by Ku cDNA *Science* 266, 288-291.
187. Knuth, M.W., Gunderson, S.I., Thompson, N.E., Strasheim, L.A., and Burgess, R.R. (1990) Purification and characterization of proximal sequence element-binding protein 1, a transcription activating protein related to Ku and TREF that binds the proximal sequence element of the human U1 promoter *J. Biol. Chem.* 265, 17911-17920.
188. Gunderson, S.I., Knuth, M.W., and Burgess, R.R. (1990) The human U1 snRNA promoter correctly initiates transcription in vitro and is activated by PSE1 *Genes & Dev.* 4, 2048-2060.
189. Roberts, M.R., Han, Y., Fienberg, A., Hunihan, L., and Ruddle, F.H. (1994) A DNA-binding activity, TRAC, specific for the TRA element of the transferrin receptor gene copurifies with the Ku autoantigen *Proc. Natl. Acad. Sci. USA* 91, 6354-6358.
190. Roberts, M.R., Miskimins, W.K., and Ruddle, F.H. (1989) Nuclear proteins TREF1 and TREF2 bind to the transcriptional control element of the transferrin receptor gene and appear to be associated as a heterodimer *Cell Regul.* 1, 151-164.

191. Niu, H., and Jacob, S.T. (1994) Enhancer 1 binding factor (E1BF), a Ku-related protein, is a growth-regulated RNA polymerase I transcription factor: association of a repressor activity with purified E1BF from serum-deprived cells *Proc. Natl. Acad. Sci. USA* 91, 9101-9105.

192. Hoff, C.M., Ghosh, A.K., Prabhakar, B.S., and Jacob, S.T. (1994) Enhancer 1 binding factor, a Ku-related protein, is a positive regulator of RNA polymerase I transcription initiation *Proc. Natl. Acad. Sci. USA* 91, 762-766.

193. Okumura, K., Takagi, S., Sakaguchi, G., Naito, K., Minoura-Tada, N., Kobayashi, H., Mimori, T., Hinuma, Y., and Igarashi, H. (1994) Autoantigen Ku protein is involved in DNA binding proteins which recognize the U5 repressive element of human T-cell leukemia virus type I long terminal repeat *FEBS* 356, 94-100.

194. Kim, D., Ouyang, H., Yang, S.-H., Nussenzweig, A., Burgman, P., and Li, G.C. (1995) A constitutive heat shock element-binding factor is immunologically identical to the Ku autoantigen *J. Biol. Chem.* 270, 15277-15284.

195. Li, G.C., Yang, S.-H., Kim, D., Nussenzweig, A., Ouyang, H., Wei, J., Burgman, P., and Li, L. (1995) Suppression of heat-induced hsp70 expression by the 70-KDa subunit of the human Ku autoantigen *Proc. Natl. Acad. Sci. USA* 92, 4512-4516.

196. May, G., Sutton, C., and Gould, H. (1991) Purification and characterization of Ku-2, an octamer-binding protein related to the autoantigen Ku *J. Biol. Chem.* 266, 3052-3059.

197. Giffin, W., Torrance, H., Rodda, D.J., Prefontaine, G.G., Pope, L., and Hache, R.J.G. (1996) Sequence-specific DNA binding by Ku autoantigen and its effects on transcription *Nature* 380, 265-268.

198. Kuhn, A., Stefanovsky, V., and Grummt, I. (1993) The nucleolar transcription activator UBF relieves Ku antigen-mediated repression of mouse ribosomal gene transcription *Nucleic Acids Res.* 21, 2057-2063.

199. Toth, E.C., Marusic, L., Ochem, A., Patthy, A., Pongor, S., Giacca, M., and Falaschi, A. (1993) Interactions of USF and Ku antigen with a human DNA region containing a replication origin *Nucleic Acids Res.* 21, 3257-3263.

200. Lees-Miller, S.P., Chen, Y.-R., and Anderson, C.W. (1990) Human cells contain a DNA-activated protein kinase that phosphorylates simian virus 40 T Antigen, mouse p53, and the human Ku autoantigen *Mol. Cell. Biol.* 10, 6472-6481.

201. Schindler, C., Fu, X.-Y., Improta, T., Aebersold, R., and Darnell Jr., J.E. (1992) Proteins of transcription factor ISGF-3: one gene encodes the 91- and 84-KDa ISGF-3 proteins that are activated by interferon alfa *Proc. Natl. Acad. Sci. USA* 89, 7836-7839.

202. Higashiura, M., Shimizu, Y., Tanimoto, M., Morite, T., and Yagura, T. (1992) Immunolocalization of Ku-proteins (p80/p70): localization of p70 to nucleoli and periphery of both interphase nuclei and metaphase chromosomes *Exp. Cell Res.* 201, 444-451.

203. Yaneva, M., and Jhiang, S. (1991) Expression of the Ku protein during cell proliferation *Biochem. Bioph. Acta* 1090, 181-187.

204. Kühn, A., Gottlieb, T.M., Jackson, S.J., and Grummt, I. (1995) DNA-dependent protein kinase: a potent inhibitor of transcription by RNA polymerase I *Genes & Dev.* 9, 193-203.

205. Jacoby, D.B., and Wensinki, P.C. (1994) Yolk protein factor 1 is a Drosophila homolog of Ku, the DNA-binding subunit of a DNA-dependent protein kinase from humans *J. Biol. Chem.* 269, 11484-11491.
206. Stamatoyannopoulos, G., Josephson, B., Zhang, J.-W., and Li, Q. (1993) Developmental regulation of human gamma-globin genes in transgenic mice *Mol. Cell. Biol.* 13, 7636-7644.
207. Dvir, A., Peterson, S.R., Knuth, M.W., Lu, H., and Dynan, W. (1992) Ku autoantigen is the regulatory component of a template-associated protein that phosphorylates RNA polymerase II *Proc. Natl. Acad. Sci. USA* 89, 11920-11924.
208. Dvir, A., Stein, L.Y., Calore, B.L., and Dynan, W. (1993) Purification and characterization of a template-associated protein kinase that phosphorylates RNA Polymerase II *J. Biol. Chem.* 268, 10440-10447.
209. Lees-Miller, S.P., Godbout, R., Chan, D.W., Weinfeld, M., DayIII, R.S., Barron, G.M., and Allalunis-Turner, J. (1995) Absence of p350 subunit of DNA-activated protein kinase from a radiosensitive human cell line *Science* 267, 1183-1185.
210. Kirchgessner, C.U., Patil, C.K., Evans, J.W., Cuomo, C.A., Fried, L.M., Carter, T., Oettinger, M.A., and Brown, J.M. (1995) DNA-dependent kinase (p350) as a candidate gene for the murine SCID defect *Science* 267, 1178-1183.
211. Tuteja, N., Tuteja, R., Ochem, A., Taneja, P., Huang, N.W., Simoncsits, A., Susic, S., Rahman, K., Marusic, L., Chen, J., Zhang, J., Wang, S., Pongor, S., and Falaschi, A. (1994) Human DNA helicase II: a novel DNA unwinding enzyme identified as the Ku autoantigen *EMBO J.* 13, 4991-5001..
212. Pace, B., Li, Q., Peterson, K., and Stamatoyannopoulos, G. (1994) Alfa-amino butyric acid cannot reactivate the silenced gamma gene of the beta locus YAC transgenic mouse *Blood* 84, 4344-4353.
213. Liakopolou, E., Blau, C.A., Li, Q., Josephson, B., Wolf, J.A., Fournarakis, B., Raisys, V., Dover, G., Papayannopoulou, T., and Stamatoyannopoulos, G. (1995) Stimulation of fetal hemoglobin production by short chain fatty acids *Blood* 86, 3227-3235.
214. Zitnik, G., Peterson, K., Stamatoyannopoulos, G., and Papayannopoulou, T. (1995) Effects of butyrate and glucocorticoids on gamma- to beta-globin gene switching in somatic cell hybrids *Mol. Cell. Biol.* 15, 790-795.
215. McKnight, G.S., Hager, L., and R.T. Palmiter (1980) Butyrate and related inhibitors of histone deacetylation block the induction of egg white genes by steroid hormone *Cell* 22, 469-477.
216. Pace, B.S., Li, Q., and Stamatoyannopoulos, G. (1996) In vivo search for butyrate responsive sequences using transgenic mice carrying Agamma gene promoter mutants *Blood* 88, 1079-1083.
217. Mantovani, R., Malgaretti, N., Giglioni, B., Comi, P., Cappellini, N., Nicolis, S., and Ottolenghi, S. (1987) A protein factor binding to an octamer motif in the gamma-globin promoter disappears upon induction of differentiation and hemoglobin synthesis in K562 cells *Nucleic Acids Res.* 15, 9349-9364.
218. Perez-Stable, C., Costantini, F. (1990) Roles of fetal G-gamma-globin promoter elements and the adult beta-globin 3' enhancer in the stage-specific expression of globin genes *Mol. Cell. Biol.* 10, 1116-1125.
219. Stoming, T.A., Stoming, G.S., Lanclos, K.D., and Nute, P.E. (1989) An A-gamma type of nondeletional hereditary persistence of fetal hemoglobin with a T->C mutation at position -175 to the cap site of the A-gamma globin gene *Blood* 73, 329-333.

220. Surrey, S., Delgrosso, K., Malladi, P., and Schwartz, E. (1988) A single-base change at position -175 in the 5' flanking region of the G-gamma globin gene from black with G-gamma-beta+ HPFH *Blood* 71, 807-810.
221. Ottolenghi, S., Nicolis, R., Taramelli, R., Malgaretti, N., Mantovani, R., Comi, P., Giglioni, B., Longinotti, M., Dore, F., Oggiano, L., Pistidda, P., Serra, A., Camaschella, C., and Saglio, G. (1988) Sardinian G-gamma-HPFH: a T->C substitution in a conserved "octamer" sequence in the G-gamma-globin promoter *Blood* 71, 815-817.
222. Costa, F.F., Zago, M.A., Cheng, G., Nechtman, J.F., Stoming, T.A., and Huisman, T.H.J. (1990) The Brazilian type of nondeletional A-gamma-fetal hemoglobin has a C->G substitution at nucleotide -195 of the A-gamma-globin gene *Blood* 76, 1986-1987.
223. Giglioni, B., Casini, C., Mantovani, R., Merli, S., Comi, P., Ottolenghi, S., Saglio, G., Camaschella, C., and Mazza, U. (1984) A molecular study of a family with Greek hereditary persistence of fetal hemoglobin and beta-thalassemia *EMBO J.* 3, 2641-2645.
224. Gelinas, R., Bender, M., and Lotshaw, C. (1986) Chinese A-gamma fetal hemoglobin: C to T substitution at position -196 of the A-gamma gene promoter *Blood* 67, 1777.
225. Tate, V.E., Wood, W.G., and Weatherall, D.J. (1986) The British form of hereditary persistence of fetal hemoglobin results from a single base pair mutation adjacent to an S1 hypersensitive site 5' to the A-gamma globin gene *Blood* 82, 6384-6388.
226. Jane, S.M., Gumucio, D.L., Ney, P.A., Cunningham, J.M., and Nienhuis, A.W. (1993) Methylation-enhanced binding of Sp1 to the stage selector element of the human gamma-globin gene promoter may regulate developmental specificty of expression *Mol. Cell. Biol.* 13, 3272-3281.
227. Gilman, J., Mishima, N., Wen, X., Kutlar, F., and Huisman, T. (1988) Upstream mutations associated with a modest elevation of fetal hemoglobin expression in human adults *Blood* 72, 78-81.
228. Collins, F., Stoeckert, C., Sergeant, G., Forget, B., and Weissman, S. (1984) G-gamma beta+ hereditary persistence of fetal hemoglobin: cosmid cloning and identification of a specific mutation 5' to the G-gamma gene *Proc. Natl. Acad. Sci. USA* 81, 4894-4898.
229. Collins, F., Boehm, C., Waber, P., Stoeckert, C., Weissman, S., Forget, B., and Kazakian, H. (1984) Concordance of a point mutation 5' to the G-gamma globin gene with G-gamma beta+ hereditary persistence of fetal hemoglobin in the black population *Blood* 64, 1292-1296.
230. Ronchi, A., Nicolis, S., Santoro, C., and Ottolenghi, S. (1989) Increased Sp1 binding mediates erythroid-specific overexpression of a mutated HPFH gamma-globin promoter *Nucleic Acids Res.* 17, 10231-10241.
231. Fischer, K.-D., and Nowock, J. (1990) The T->C substitution at -198 of the A gamma-globin gene associated with the British form of HPFH generates overlapping recognition sites for two DNA-binding proteins *Nucleic Acids Res.* 18, 5685-5693.
232. Sykes, K., and Kaufman, R. (1990) A naturally occuring gamma globin gene mutation enhances Sp1 binding activity *Mol. Cell. Biol.* 10, 95-102.
233. Nicolis, S., Ronchi, A., Malgaretti, N., Mantovani, R., Giglioni, B., and Ottolenghi, S. (1989) Increased erythroid-specific expression of a mutated HPFH

gamma-globin promoter requires the erythroid factor NFE-1 *Nucleic Acids Res.* 17, 5509-5516.

234. Mantovani, R., Malgaretti, N., Nicolis, S., Ronchi, A., Giglioni, B. and Ottolenghi, S. (1988) The effects of HPFH mutations in the human gamma-globin promoter on binding of ubiquitous and erythroid specific nuclear factors *Nucleic Acids Res.* 16, 7783-7797.

235. Lloyd, J.A., Lee, R.F., and Lingrel, J.B. (1989) Mutations in two regions upstream of the Agamma globin gene canonical promoter affect gene expression *Nucleic Acids Res.* 17, 4339-4352.

236. Oner, R., Kutlar, F., Gu, L.-H., and Huisman, T.H.J. The Georgia type of nondeletional hereditary persistence of fetal hemoglobin has a C->T mutation at nucleotide -114 of the A-gamma-globin gene *Blood* 77, 1124-1128.

237. Fucharoen, S., Shimizu, K., and Fukumaki, Y. (1990) A novel C-T transition within the distal CCAAT motif of the G-gamma-gene in the Japanese HPFH: implication of factor binding in elevated fetal globin expression *Nucleic Acids Res.* 18, 5245-5253.

238. Ronchi, A.E., Bottardi, S., Mazzucchelli, C., Ottolenghi, S., and Santoro, C. (1995) Differential binding of the NFE-3 and CP1/NF-Y transcription factors to the human gamma- and epsilon-globin CCAAT boxes *J. Biol. Chem* 270, 21934-21941.

239. Ronchi, A., Berry, M., Raguz, S., Imam, A., Yannoutsos, N., Ottolenghi, S., Grosveld, F., and Dillon, N. (1996) Role of the duplicated CCAAT box region in gamma-globin gene regulation and hereditary persistence of fetal haemoglobin *EMBO J.* 15, 143-149.

240. Gelinas, R., Endlich, B., Pfeiffer, C., Yagi, M., and Stamatoyannopoulos, G. (1985) G to A substitution in the distal CCAAT box of the A-gamma-globin gene in Greek hereditary persistence of fetal hemoglobin *Nature* 313, 323-325.

241. Collins, F.S., Metherall, J.E., Yamakawa, M., Pan, J., Weissman, S.M., and Forget, B.G. (1985) A point mutation in the A-gamma-globin gene promoter in Greek hereditary persistence of fetal hemoglobin *Nature* 313, 325-326.

242. Ottolenghi, S., Nicolis, S., Bertini, C., Ronchi, A., Crotta, S., and Giglioni, B. (1990) Regulation of gamma-globin expression in hereditary persistence of fetal hemoglobin *Sixth Cooley's anemia symposium-Annals of the N.Y. Academy Sci.* 612, 191-195.

243. Chodosh, L.A., Baldwin, A.S., Carthew, R.W., and Sharp, P.A. (1988) Human CCAAT-binding proteins have heterologous subunits *Cell* 53, 11-24.

244. Gumucio, D.L., Rood, K.L., Gray, T.A., Riordan, M.F., Sartor, C.I., and Collins, F. (1988) Nuclear proteins that bind the human gamma-globin gene promoter: alterations in binding produced by point mutations associated with hereditary persistence of fetal hemoglobin *Mol. Cell. Biol.* 8, 5310-5322.

245. Gilman, J.G., Mishima, N., Wen, X.J., Stoming, T.A., Lobel, J., and Huisman, T.H.J. (1988) Distal-CCAAT box deletion in the A-gamma-globin gene of two black adolescents with elevated fetal A-gamma-globin *Nucleic Acids Res.* 16, 10635-10642.

246. Mantovani, R., Superti-Furga, G., Gilman, J., and Ottolenghi, S. (1989) The deletion of the distal CCAAT box region of the human A-gamma-gene in black HPFH abolishes the binding of the erythroid specific protein NFE-3 and of the CCAAT displacement protein *Nucleic Acids Res.* 17, 6681-6691.

247. Manley, J.L., Fire, A., Cano, A., Sharp, P.A., and Gefter, M.L. (1980) DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract *Proc. Natl. Acad. Sci. USA* 77, 3855-3859.
248. Kadonaga, J.T., and Tjian, R. (1986) Affinity purification of sequence-specific DNA binding proteins *Proc. Natl. Acad. Sci. USA* 83, 5889-5893.
249. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular cloning: a laboratory manual. *Cold Spring Harbor Laboratory Press*, Cold Spring Harbor, NY.
250. Li, X.-Y., van Huijsduijnen, R.H., Mantovani, R., Benoist, C., and Mathis, D. (1991) Intron-exon organization of the NF-Y genes *J. Biol. Chem.* 267, 8894-8990.
251. Mantovani, R., Pessare, U., Tronche, F., Li, X.-Y., Knapp, A.-M., Pasquali, J.L., Benoist, C., and Mathis, D. (1992) Monoclonal antibodies to NF-Y define its function in MHC class II and albumin gene transcription *EMBO J.* 11, 3315-3322.
252. Genersch, E., Eckerskon, C., Lottspeich, F., Herzog, C., Kuhn, K., and Poschl, E. (1995) Purification of the sequence-specific transcription factor CTCBF, involved in the control of human collagen IV genes: subunits with homology to Ku antigen *EMBO J.* 14, 791-800.
253. Indrak, K., Indrakova, J., Kutlar, F., Pospisilova, D., Sulovska, I., Baysal, E., and Huisman, T.H. (1991) Compound heterozygosity for a beta zero-thalassemia (frameshift codons 38/39; -C) and a nondeletional Swiss type of HPFH (A->C at nt -110, G gamma) in a Czechoslovakian family *Ann. Hematol.* 63, 111-115.
254. Milot, E., Strouboulis, J., Trimborn, T., Wijgerde, M., de Boer, E., Langeveld, A., Tan-Un, K., Vergeer, W., Yannoutsos, N., Grosveld, F., and Fraser, P. (1996) Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription *Cell* 87, 105-114.
255. Farmer, S.C., Sun, C.-W., Winnier, G.E., Hogan, B.L.M., and Townes, T.M. (1997) The bZIP transcription factor LCR-F1 is essential for mesoderm formation in mouse development *Genes & Dev.* 11, 786-798.
256. Aplan, P.D., Jones, C.A., Chervinsky, D.S., Zhao, X.F., Ellsworth, M.K., Wu, C., McGuire, E.A., and Gross, K.W. (1997) An scl gene product lacking the transactivation domain induces bony abnormalities and cooperates with LMO1 to generate T-cell malignancies in transgenic mice *EMBO J.* 16, 2408-2419.