THE UNIVERSITY OF MANITOBA

FACTORS INTERACTING WITH THE PROMOTER AND 3′ ENHANCER OF THE CHICKEN HISTONE H5 GENE

BY

GAIL PENNER

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULLFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

ISBN 0-315-92239-7

Name C. Gail Penner

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

Molecular Biology
SUBJECT TERM

0307
SUBJECT CODE

U·M·I

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS
Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION
General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523

Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS
Language
- General 0679
- Ancient 0289
- Linguistics 0290
- Modern 0291

Literature
- General 0401
- Classical 0294
- Comparative 0295
- Medieval 0297
- Modern 0298
- African 0316
- American 0591
- Asian 0305
- Canadian (English) 0352
- Canadian (French) 0355
- English 0593
- Germanic 0311
- Latin American 0312
- Middle Eastern 0315
- Romance 0313
- Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY
Philosophy 0422
Religion
- General 0318
- Biblical Studies 0321
- Clergy 0319
- History of 0320
- Philosophy of 0322

Theology 0469

SOCIAL SCIENCES
American Studies 0323
Anthropology
- Archaeology 0324
- Cultural 0326
- Physical 0327

Business Administration
- General 0310
- Accounting 0272
- Banking 0770
- Management 0454
- Marketing 0338

Canadian Studies 0385
Economics
- General 0501
- Agricultural 0503
- Commerce-Business 0505
- Finance 0508
- History 0509
- Labor 0510
- Theory 0511

Folklore 0358
Geography 0366
Gerontology 0351
History
- General 0578
- Ancient 0579
- Medieval 0581
- Modern 0582
- Black 0328
- African 0331
- Asia, Australia and Oceania 0332
- Canadian 0334
- European 0335
- Latin American 0336
- Middle Eastern 0333
- United States 0337

History of Science 0585
Law 0398
Political Science
- General 0615
- International Law and Relations 0616
- Public Administration 0617

Recreation 0814
Social Work 0452
Sociology
- General 0626
- Criminology and Penology 0627
- Demography 0938
- Ethnic and Racial Studies 0631
- Individual and Family Studies 0628
- Industrial and Labor Relations 0629
- Public and Social Welfare 0630
- Social Structure and Development 0700
- Theory and Methods 0344

Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES
Agriculture
- General 0473
- Agronomy 0285
- Animal Culture and Nutrition 0475
- Animal Pathology 0476
- Food Science and Technology 0359
- Forestry and Wildlife 0478
- Plant Culture 0479
- Plant Pathology 0480
- Plant Physiology 0817
- Range Management 0777
- Wood Technology 0746

Biology
- General 0306
- Anatomy 0287
- Biostatistics 0308
- Botany 0309
- Cell 0379
- Ecology 0329
- Entomology 0353
- Genetics 0369
- Limnology 0793
- Microbiology 0410
- Molecular 0307
- Neuroscience 0317
- Oceanography 0416
- Physiology 0433
- Radiation 0821
- Veterinary Science 0778
- Zoology 0472

Biophysics
- General 0786
- Medical 0760

EARTH SCIENCES
Biogeochemistry 0425
Geochemistry 0996
Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES
Environmental Sciences 0768
Health Sciences
- General 0566
- Audiology 0300
- Chemotherapy 0992
- Dentistry 0567
- Education 0350
- Hospital Management 0769
- Human Development 0758
- Immunology 0982
- Medicine and Surgery 0564
- Mental Health 0347
- Nursing 0569
- Nutrition 0570
- Obstetrics and Gynecology 0380
- Occupational Health and Therapy 0354
- Ophthalmology 0381
- Pathology 0571
- Pharmacology 0419
- Pharmacy 0572
- Physical Therapy 0382
- Public Health 0573
- Radiology 0574
- Recreation 0575
- Speech Pathology 0460
- Toxicology 0383

Home Economics 0386

PHYSICAL SCIENCES
Pure Sciences
Chemistry
- General 0485
- Agricultural 0749
- Analytical 0486
- Biochemistry 0487
- Inorganic 0488
- Nuclear 0738
- Organic 0490
- Pharmaceutical 0491
- Physical 0494
- Polymer 0495
- Radiation 0754

Mathematics 0405
Physics
- General 0605
- Acoustics 0986
- Astronomy and Astrophysics 0606
- Atmospheric Science 0608
- Atomic 0748
- Electronics and Electricity 0607
- Elementary Particles and High Energy 0798
- Fluid and Plasma 0759
- Molecular 0609
- Nuclear 0610
- Optics 0752
- Radiation 0756
- Solid State 0611

Statistics 0463

Applied Sciences
Applied Mechanics 0346
Computer Science 0984

Engineering
- General 0537
- Aerospace 0538
- Agricultural 0539
- Automotive 0540
- Biomedical 0541
- Chemical 0542
- Civil 0543
- Electronics and Electrical 0544
- Heat and Thermodynamics 0348
- Hydraulic 0545
- Industrial 0546
- Marine 0547
- Materials Science 0794
- Mechanical 0548
- Metallurgy 0743
- Mining 0551
- Nuclear 0552
- Packaging 0549
- Petroleum 0765
- Sanitary and Municipal 0554
- System Science 0790

Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY
General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

FACTORS INTERACTING WITH THE PROMOTER AND 3' ENHANCER OF THE CHICKEN HISTONE H5 GENE

BY

GAIL PENNER

A Thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Abstract

The histone H5 protein is a variant linker histone which is expressed in the nucleated erythrocytes of birds. The function of histone H5 is to repress the bulk of the chromatin of avian erythrocytes. Expression of the histone H5 gene in erythrocytes is under the control of a promoter region and a number of enhancer elements, including an enhancer located 3′ of the protein coding region. The promoter and 3′ enhancer region were examined for the ability to interact with erythroid nuclear proteins, through the use of *in vitro* assays. A number of erythroid-specific and ubiquitous factors were shown to be capable of binding to the regions. The factors which were able to bind these regions and which were studied most extensively were the Sp1 factor and GATA-1, an erythroid specific factor.

Sp1 was shown to be capable of binding with high affinity to a site in the promoter and with medium and weak affinity to four sites in the enhancer. The Sp1 binding sites were occupied by high molecular weight complexes which suggested formation of tetrameric and multiples of tetrameric complexes of Sp1 as had been demonstrated previously. Sp1 bound at distant sites has been shown to be capable of interacting resulting in a looping out of the intervening DNA gene chromatin in chicken erythrocytes in which Sp1 bound at the promoter interacts with Sp1 bound within the enhancer is suggested based on these data. This interaction allows the enhancer and promoter to be juxtapositioned, with a looping out of the intervening DNA. The enhancer bound factors are therefore able to interact with the promoter bound factors and thereby affect transcription. Some of the largest complexes bound at the Sp1 binding site may also have contained other

proteins in addition to Sp1, such as factors of the general transcription factor TFIID (Weinzierl *et al.*, 1993).

GATA-1 was shown to be capable of interacting with non-DNA binding proteins to form GATA-1 containing protein complexes. GATA-1 containing protein complexes were partially purified. These GATA-1 containing protein complexes may be involved in affecting transcription activity and/or the chromatin structure of the transcriptionally active histone H5 gene in chicken erythrocytes. The action of the GATA-1 containing protein complexes is expected to occur within the context of the looping model.

The transcription factor NF1 which has variants that are associated with the nuclear matrix was shown to have a binding site within the histone H5 gene (Sun *et al*, 1992). Therefore, this transcription factor may provide an anchorage site for the histone H5 gene. The function of the nuclear matrix association would be to allow the gene access to RNA polymerase II which is thought to be nuclear matrix associated.

Statement of Authenticity

The data presented in this thesis were done by myself except for Figures 30 through 36. These figures were generated in conjunction with/or by Mr. Jian-Min Sun. However, extracts prepared by myself were used to determine that the data was reproducible. In some of the cases, I had also generated data which demonstrated the same results presented in this thesis.

Acknowledgements

I would like to thank Dr. James Davie, who accepted me as a graduate student and directed my research. The people of my committee who also provided suggestions for my research, and who had the task of reading this thesis and examining me on it, deserve a special thank-you. These people include Dr. James Davie, Dr. Leigh Murphy, Dr.Frits Stevens, Dr. Peter Cattini and Dr. David Bazett-Jones. Jian-Min Sun was also involved in this project. He provided much of the data with respect to the histone H5 promoter region and some of the enhancer region data.

Table of Contents

List of Figures

List of Tables

Abbreviations

A	adenosine
alpha G2	mouse alpha globin promoter GATA binding site
AP-1	Activator Protein 1
ATF	Activating transcription factor
ATP	adenosine triphosphate
BFU-E	burst forming cells of erythroid lineage
bp	basepairs
BPV	bovine papilloma virus
C	cytosine
CDP	CCAAT displacement protein
CBF	CCAAT binding factor
CFU-E	colony forming cells of erythroid lineage
CFU-M	colony forming cells of macrophage lineage
cGATA-1	chicken GATA-1
μCi	microCurie
clBF	chicken initiation factor
clBR	chicken initiation binding repressor
CP2	CCAAT protein 2
CSF	colony stimulating factor
Dde I	DNA endonuclease
deltaN250	C-terminal truncated version of 250 kDa *Drosophila* TATA binding protein associated factor
DH	DNAse I hypersensitive

DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxynucleotide triphosphate
DS1	DNA fragment used in research (Figure 6)
DS1a-L	DNA fragment used in research (Figure 6)
DS1a-U	DNA fragment used in research (Figure 6)
DS2	DNA fragment used in research (Figure 6)
$dTAF_{110}$	*Drosophila* 110 kDa TATA binding protein associated factor
$dTAF_{250}$	*Drosophila* 250 kDa TATA binding protein associated factor
dTBP	*Drosophila* TATA binding protein
dTTP	deoxythymidine triphosphate
E1A	adenovirus early region 1a transcription factor
E1B	adenovirus early region 1b gene promoter
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
Epo	erythropoietin
EpoR	erythropoietin receptor
eUSF	erythroid upstream stimulatory factor
F.T.	flowthrough
G	guanine
GATA-1	erythroid transcription factor which binds GATA sequence

Gal4	yeast transcription factor *gal* gene
GC box	binding site for transcription factor Sp1
H4TF2	histone H4 transcription factor 2
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
Hha I	DNA endonuclease
Hind III	DNA endonuclease
Hpa II	DNA endonuclease
HPFH	hereditary persistence of fetal hemoglobin
$hTAF_{125}$	125 kDa human TATA binding protein associated factor
$hTAF_{250}$	250 kDa human TATA binding protein associated factor
hTBP	human TATA binding protein
IAC	iodoacetamide
IE	immature erythrocyte
IL-3	interleukin 3
kb	kilobases
kDa	kilodalton
LCR	locus control region
M	molar
MAR	matrix associated region
ME	mature erythrocytes
MEL	murine erythroleukemia
MLP	major late promoter

mM	millimolar
MMTV	mouse mammary tumor virus
MNase I	micrococcal nuclease I
MSB-1	T-lymphoblastoma cells
MvaI	DNA endonuclease
NE	nuclear extract
NF-E2	erythroid nuclear factor 2
NF-E4	erythroid nuclear factor 4
NF1	Nuclear Factor I
NP-40	nonidet P-40
NT	non-template
NTP	nucleotide triphosphate
^{32}P	radioactive phosphorous
PAGE	polyacrylamide gel electrophoresis
PAL	chicken erythrocyte palindromic binding protein
PBGD	porphobilinogen deaminase gene
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PHO5	*S. cerevesiae* acid phosphatase gene
PMSF	phenylmethanesulphonyl fluoride
poly dA.dT	poly deoxyadenosine deoxythymine
poly dI.dC	poly deoxyinosine deoxycytosine
Pvu II	DNA endonuclease
RBC	red blood cell

RNA	ribonucleic acid
RNA pol II	RNA polymerase II
RSB	reticulocyte standard buffer
SAR	scaffold associated region
Sau3A1	DNA endonuclease
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
Sma I	DNA endonuclease
Sp1	transcription factor initially partially purified over Sephacryl S-300 and phosphocellulose columns
SSE	stage specific element
T	template
TAF	TATA binding protein associated factor
TAF_{110}	110 kDa TATA binding protein associated factor
TATA	binding element for TFIID
TBE	Tris borate EDTA buffer
TBP	TATA binding protein
TE	Tris EDTA buffer
TEF-2	transcriptional enhancer factor
TF	transcription factor
TFIIA-TFIIJ	basic transcription factors of RNA polymerase II
T	thymine

TNE	Tris sodium EDTA buffer
UAS	upstream activating sequence
UNE	upstream negative element of histone H5 promoter
UPE	upstream promoter element
US1a	DNA fragment used in research (Figure 6)
UTP	uridine triphosphate
VP-16	herpes simplex virus transcription factor
w/v	weight/volume
Xba I	DNA endonuclease

Introduction

Role of transcription in tissue specificity of genes expressed within tissues

Science has begun to understand how diverse cell types of a multicellular organism are generated. In a multicellular organism all cells contain essentially the same genetic material as the zygote; one maternal and one paternal set of chromosomes. However, despite the identical genotype, cells display widely divergent phenotypes. Different cell phenotypes are the result of differences in their pattern of gene expression. This causes cell-specific protein expression, which then affects cell function and phenotype.

Gene expression is controlled at a number of levels including transcription, transcript processing, stability of the transcript, translation to protein, stability of the protein and post-translational modifications. Although each of these processes is important in its own right for the phenotype of the cell, transcription is arguably the most important. Because initiation of transcription is the first step in expression of proteins, it is the controlling step at which the cell can decide to expend energy to express the gene. This may be the reason there is such an intricate control of transcription by a plethora of tissue specific, stage specific and ubiquitous transcription factors.

This thesis is concerned with the factors that bind to the histone H5 gene of chicken erythrocytes. Histone H5 is expressed only in nucleated erythrocytes, an environment in which most genes are silenced. Therefore the transcription factors involved in H5 gene expression need to overcome the factors that are important for silencing the majority of the genes. H5 gene expression is different from many of

the erythroid-specific genes in that its expression begins earlier than most, during erythroid differentiation. The results of this research centre on three major themes: 1) The interaction of the erythroid-specific factor GATA-1 with other proteins which are believed to have an important effect on transcription; 2) The interaction of Sp1 or Sp1-like proteins, to form high molecular weight complexes, which are thought to be important for bringing the 3′ enhancer region of H5 into juxtaposition with the promoter; and 3) the other factors binding to the 3′ enhancer and to the promoter region of H5, which are also important to H5 gene expression. Transcription *in vivo* takes place in the presence of chromatin, and since chromatin and transcription factors each have an effect on the other, this thesis will begin with a general discussion of chromatin. After the discussion of chromatin and the changes in transcriptionally active chromatin, the introduction will move to transcription initiation. The introduction will close with erythropoiesis, the factors involved in transcription during erythropoiesis and finally the specifics of histone H5 gene expression.

Chromatin Domain Organization

In higher eukaryotes, gene regulation is controlled to such an extent that a gene which is switched off is transcriptionally suppressed by as much as 10^9 fold relative to its fully active state. Generally gene expression is regulated at two different levels (Weintraub, 1985; Mitchel and Tjian, 1989; Felsenfeld, 1992). First, a general mechanism unfolds the chromatin domain to allow the transcription machinery access to the gene. The second mechanism includes the interaction of *trans*-acting factors with *cis*-acting regulatory DNA sequences such as promoters

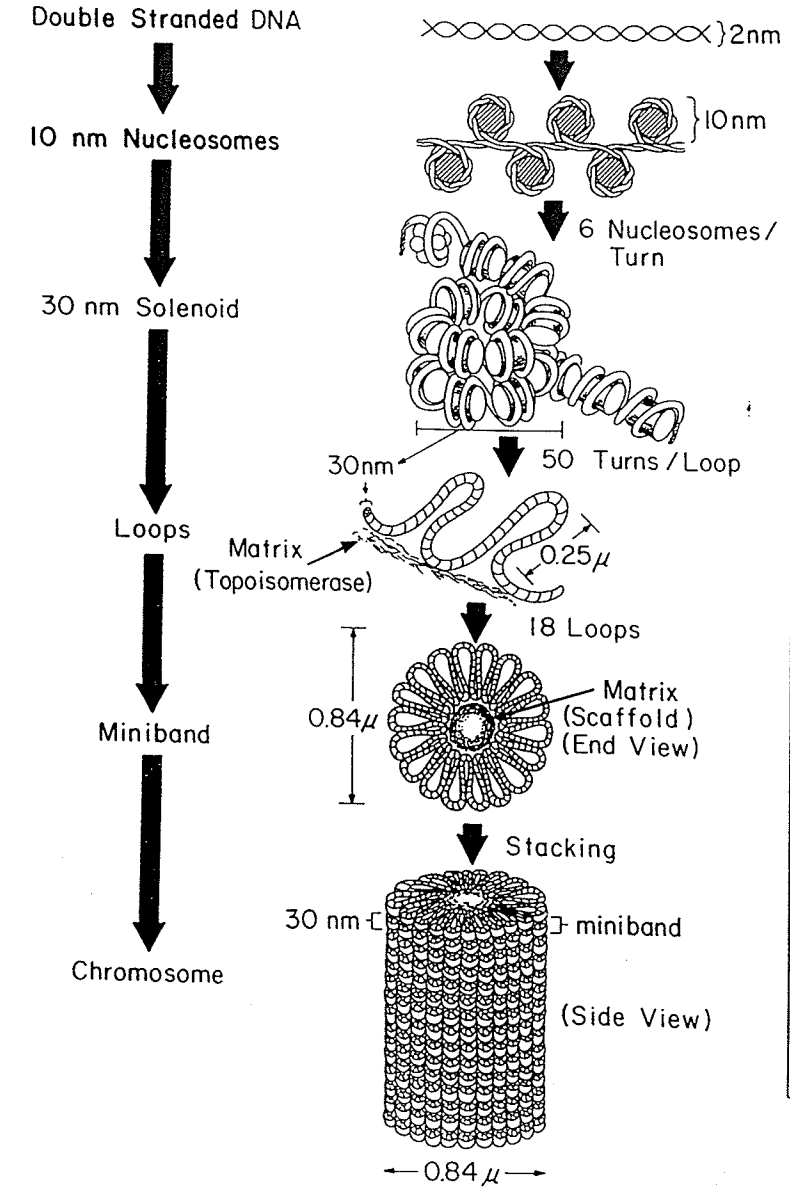
and enhancers (Dynan, 1989; Mitchell and Tjian, 1989). The two mechanisms do not act in isolation. For example, there is evidence that transcription factors alter the structure of chromatin, which in turn changes accessibility for these and other *trans*-acting factors. Therefore a discussion of transcription would not be complete without a discourse on chromatin domain organization within the cell.

The chromatin structure of the DNA within a cell allows an immense amount of DNA to be organized in such a way that individual genes are readily accessible. The chromatin structure compacts DNA and yet allows those genes to be expressed whose products are needed in a particular tissue. Thus chromosome structure and function has the overall effect of reducing the accessible genome by removing from the environment, through nucleosome assembly and higher order chromatin organization, DNA which would compete for transcription factors. In some cases, it has been shown that chromatin structure is required for transcription factors to be able to exert their effect (Laybourn and Kadonaga, 1992).

As an example of the extent of folding of DNA, the nuclei of most human cells contains about 6 X 10^9 bp of DNA, which is roughly 2 meters in length when extended. This DNA is folded about 12,000-fold and is condensed to a 10 μm diameter sphere thereby effecting a final packing ratio of 200,000-fold (Getzenberg *et al.*, 1991). There is a hierarchy of chromatin structures which is responsible for this folding (Figure 1). At the first level of chromatin folding, DNA binds histones, to form a long 10 nm fiber of nucleosomes which is the "beads-on-a-string" appearance of the chromatin when it is spread across an electron microscope grid at low ionic strength. The beads correspond to nucleosome core particles, the basic repeating structural units in chromatin, and the stretch of DNA that joins

Figure 1. Schematic of the levels of organization within a chromatid of a chromosome. Approximately 160 bp of 2 nm DNA helix is wound twice around the histone octamers to form the 10 nm nucleosomes. These nucleosomes form a "beads on a string" fiber, which winds in a solenoid fashion with 6 nucleosomes per turn to form the 30 nm chromatin filament. The 30 nm filament forms the 60 kilobase DNA loops that are attached at their bases to the nuclear matrix structure. The loops are then wound into the 18 radial loops that form a miniband unit or 1 turn on the chromatid. Reproduced from Getzenberg *et al.* (1991).

The Formation Of The Radial Loop Chromosome



Base Pairs per Turn	Packing Ratio
10 b.p.	1
80 b.p (160 b.p. per 2 turns)	6 - 7
1,200 b.p. (per turn)	40
60,000 b.p. (per loop)	680
1.1×10^6 b.p. (per miniband)	1.2×10^4
18 loops/ Miniband	1.2×10^4

them is termed the linker. The core particle has a disc-like shape, is 11 nm in diameter and 5.6 nm in height. The nucleosome consists of 146 base pairs of DNA wrapped in 1.75 turns of left-handed super helix around a protein core composed of two copies of the four histones H2A, H2B, H3 and H4, arranged as a $(H3\text{-}H4)_2$ tetramer and two H2A-H2B dimers. The length of the linker DNA may differ among genes in the same tissue, as well as between tissues and species, from a minimum of 15 to 20 base pairs in yeast to about 80 base pairs in sea urchin sperm (van Holde, 1988).

In the nucleus the 10 nm chromatin fiber is usually compacted further into the 30 nm fiber, also called the solenoid, which is the interphase state of chromatin. The chromatosome is a term used to include the nucleosome together with the linker histone and also includes about 167 basepairs of DNA required for these histones to bind. Linker histone (H1, H5, H1°) is required for stabilization of the 30 nm fiber (Thoma *et al.*, 1979; Renz *et al.*, 1977; Allan *et al.*, 1981; Thoma and Koller, 1981; McGhee *et al.*, 1983). The solenoid model for the 30 nm fibre has six nucleosomes per turn of the 'beads on the string' (Finch and Klug, 1976; Sperling and Klug, 1977; Renz *et al.*, 1977; Widom and Klug, 1985; Felsenfeld and McGhee, 1986; Butler, 1988). There have been many models proposed to describe the 30 nm filament, all of which predict a packing ratio of approximately 35-50 for the DNA within the filament.

A further level of chromatin structure involves the folding of the 30 nm fiber into loops or domains, and complex arrays of chromatin loops form chromosomes (Paulson and Laemmli, 1977; Warren and Cook, 1978). The interphase nucleus, consists of 30 nm filaments which form approximately 50,000 loop domains, each

of which is a unit of gene regulation, and are attached at their base to the inner portions of the nuclear matrix (Pardoll *et al.*, 1980; Rodi and Saverbier, 1989; Huberman and Riggs, 1968). Loop sizes range from a few to several hundreds of kilobasepairs (Gasser and Laemmli, 1987). The conformational state of the chromatin (e.g., supercoiling, compaction, accessibility to protein factors) is controlled for each loop individually.

Structure of Histones

Each of the four core histone proteins contains three domains; a long extended hydrophilic amino-terminal tail, containing a number of positively charged amino acids; a globular, hydrophobic core; and a very short hydrophilic carboxy-terminal tail. The hydrophobic cores are involved largely in the histone-histone interactions of the octamer. Enzymatic removal of the amino-terminal tails has only a minor effect on stability or assembly of the core particle (Whitlock and Stein, 1978; Ausio *et al.*, 1989), which suggests that their principal function is likely to be in histone/protein (Johnson *et al.*, 1990) or histone/DNA (Hill and Thomas, 1990) interactions outside of the core particle. The N-terminal tails of the nucleosomal histones are also thought to contribute to the stability of higher order chromatin structures by interacting with the DNA of the neighbouring nucleosome.

Linker histones contain three domains, a hydrophobic, globular, central core, a short N-terminal tail and a long C-terminal tail both of which are hydrophilic and unstructured and have a high content of basic residues, especially lysine (Hartmann *et al.*, 1977; Aviles *et al.*, 1978; Strickland *et al.*, 1980; Puigdomenech *et al.*, 1980). The linker histones are arranged head to tail in the 30 nm chromatin (Noll

and Kornberg, 1977; Lennard and Thomas, 1985). The globular domain of linker histones interacts at the nucleosome dyad and seals the two turns of the nucleosomal DNA by binding with a 10-bp nucleotide region at both ends of nucleosomal DNA, thereby protecting the chromatosomal 166 bp of DNA (Simpson, 1978b; Thoma *et al.*, 1979; Allan *et al.*, 1980; Gaubatz and Chalkley, 1977). The H1 or H5 carboxyl tails interact with linker DNA (Segers *et al.*, 1991; Allan *et al.*, 1980; Thomas and Wilson, 1986; Drew and McCall, 1987; van Holde, 1988) and act to condense the chromatin fiber in the physiological range of ionic strength (Allan *et al.*, 1980).

Poised/Transcribed Chromatin differs from Repressed chromatin

Transcription, and in particular initiation of transcription, is affected by chromatin structure. This is due in part to the effect of chromatin on transcription factor access to DNA. Within chromatin, general mechanisms which repress gene activity are nucleosome and linker histones which stabilize condensed chromatin structure and act as a barrier to transcription factor access to DNA (Weintraub, 1985; Croston *et al.*, 1991; Laybourn and Kadonaga, 1991; Finch and Klug, 1976; Renz *et al.*, 1977). The experiments of Croston *et al.* (1991) and Laybourn and Kadonaga (1991) demonstrated this. Binding of H1 to naked DNA or DNA reconstituted with core histones inhibited transcription of a test gene containing flanking binding sites for GAL4-VP16, Sp1 or GAGA factor *in vitro* (Croston *et al.*, 1991). Addition of the appropriate factor relieved the inhibition by H1. Histone H1 binds to DNA less tightly than do the core histones and so the order of addition is less important for histone H1 than for core histones: GAL4-VP16, even if added

after H1, was able to reverse the inhibition. In this *in vitro* system, there is irreversible repression by core histones, semi-reversible inhibition by H1, and activation above the basal level by GAL4-VP16 on chromatin (Laybourn and Kadonaga, 1991; Croston *et al.*, 1991). Therefore sequence-specific factors such as Sp1 and GAL4-VP16 both counteract the chromatin-mediated repression (antirepression) and facilitate transcription (true activation) (Laybourn and Kadonaga, 1992). Dusserre and Mermod (1992) demonstrated antirepression and transactivation mediated by cofactors of the transcription factor CTF.

Most expressed genes are organized as nucleosomes that are more susceptible to digestion by nucleases, including deoxyribonuclease I (DNase I) than the bulk of the genome (Gross and Garrard, 1987; Elgin, 1988; Weisbrod, 1982). The DNase I sensitivity, which is used as a biochemical tool indicates that the 30 nm fiber in the neighbourhood of transcribed genes is at least partially disrupted (Kimura *et al.*, 1983; Fisher and Felsenfeld, 1986). *In vivo*, factors that contribute to the variant 30 nm fiber are reversible post-synthetic histone modifications (e.g. histone "tail" acetylation, methylation, ubiquitination and linker histone phosphorylation), histone variants or subtypes and nonhistone chromosomal proteins (van Holde, 1988; Davie and Nickel, 1987; Davie and Saunders, 1981; Ball *et al.*, 1983). Among the effects of these modifications is the disruption of H1-H1 contacts essential for the cooperative H1 histone binding. There is some depletion of H1 (or H5) in transcriptionally competent chromatin (Postnikov *et al.*, 1991; Kamakaka and Thomas, 1990; Davie and Nickel, 1987). The DNase I sensitivity of active genes is not restricted to the coding region of the gene but extends far into adjacent non-transcribed sequences (Gross and Garrard, 1987). It is thought that

the DNase I sensitive domain is a chromosomal loop. For example, the 35 kb DNase I sensitive ß-globin domain of adult chicken erythrocytes consists of four genes, rho- $ß^H$-, $ß^A$- and ϵ-globin (Rocha *et al.*, 1984).

Function of H5

During the transition from the erythrocyte colony forming unit (CFU-E) to the mature erythrocyte, histone H5 gradually increases four-fold in nuclei with little concomitant displacement of the H1 histones in chickens (Affolter *et al.*, 1987). The highest relative increase in H5 content occurs before the cell loses its proliferative potential, between the CFU-E and the erythroblast stage. This results in a 70% net increase in linker histone (H1 plus H5) content, causing the chromatin to be highly condensed. During this process maturing red blood cells (RBCs) cease to replicate and most genes, except for erythroid-specific genes such as H5 and ß-globin, become transcriptionally inactive (Affolter *et al.*, 1987; Ringertz and Bolund, 1974; Williams, 1972). The accumulation of histone H5 in the nucleus is thought to cause inactivation of the bulk of the genome (Moss *et al.*, 1973; Appels and Wells, 1972; Bergman *et al.*, 1988; Ringertz *et al.*, 1985; Ruiz-Carrillo *et al.*, 1974; Brasch *et al.*, 1974; Dardick and Setterfield, 1976; Weintraub, 1978). The increased stability of the higher-order structure is thought to be due to the higher affinity of H5 for chromatin than that of H1 (Torres-Martinez and Ruiz-Carrillo, 1982; Mura and Stollar, 1984; Wu *et al.*, 1986). Another factor contributing to the chromatin condensation is that chicken erythrocytes have on average a higher proportion of linker histone than most other cell types. Erythrocytes have on average 1.2 linker histones (0.9 H5, 0.3 H1) per nucleosome compared to less than

0.7 for most cells (Bates and Thomas, 1981). Thus regions of the chicken erythrocyte chromatin have up to 2 linker histones per nucleosome (Thomas *et al.*, 1985; Olins *et al.*, 1976; Urban *et al.*, 1980; Weintraub, 1978; Allan *et al.*, 1981). The nucleosomes in this chromatin show an increased tendency toward interaction with neighbouring nucleosomes, and thus chromatin is compact even at low salt concentration (Nelson *et al.*, 1979).

Evidence that H5 is a general repressor of chromatin is that when chicken erythrocyte nuclei are reactivated by fusion with HeLa cells, H5 disappears, the chromatin decondenses and RNA is synthesized (Appels *et al.*, 1974). Further support is obtained by microinjection of H5 into proliferating L6 rat myoblasts (a cell line that maintains the potential to differentiate into myotubes) or transfection and transient expression of the H5 gene in rat sarcoma XC cells, which do not express H5. These treatments were linked to inhibition of transcription and replication (Bergman *et al.*, 1988; Sun *et al.*, 1989; Sun *et al.*, 1990). The selectivity of these effects suggests that H5 plays an active role in the control of DNA replication and cell proliferation.

Linker Histone Phosphorylation

In early dividing erythroblasts, newly synthesized H5 may have up to nine phosphate groups added. As the erythrocyte matures histone H5 becomes dephosphorylated (Sung *et al.*, 1977) and this correlates with genomic inactivation and chromatin condensation (Briand *et al.*, 1980; Sung and Freelender, 1978; Affolter *et al.*, 1987; Mura *et al.*, 1982). The phosphorylation of linker histone, through neutralization of charges, produces a protein that has less affinity for DNA.

The modified amino acids of histone H5 are located in the carboxyl-terminal region which is the region that contributes most to chromatin condensation. Therefore, through phosphorylation of histone H5, chicken erythrocytes are able to handle a high amount of H5 without causing condensation and hence inaccessibility to the transcription apparatus. Other evidence supporting a role for histone H5 phosphorylation in contributing to transcriptional activity comes from studies using "normal" chicken and quail fibroblast (L6 and XC) cells and transformed cells derived from these lines (6C2 and QT6) (Aubert *et al.*, 1991). H5 expressed in transformed cells is phosphorylated in contrast to the more normal cells, which show a greater decrease in transcription. *In vivo*, chicken embryonic erythrocytes become senescent and the erythroid condenses at 5 to 6 days and at 17 to 19 days, and this is correlated with histone H5 dephosphorylation (Pikaart *et al.*, 1991).

Histone H1, on the other hand, has a phosphorylation peak during mitosis, at a time when chromatin is highly condensed. Roth and Allis (1992) present a model to explain the function of linker histone tail phosphorylation in chromatin condensation. When linker histone is phosphorylated, the tails are unable to interact with DNA. This causes a charge repulsion within adjacent chromatin brought on by the positively charged linker histone tails. Therefore a domain specific decondensation would allow factors to access the DNA. The DNA could then be reprogrammed, so that either it would be condensed or, if the binding factors are transcription factors, the DNA could become transcriptionally active. Therefore, phosphorylation of both of the linker histones H1 and H5 is responsible for maintaining decondensed chromatin. Although in somatic cells, histone H1 is

phosphorylated throughout the cell cycle, there is a peak in histone H1 phosphorylation at G2/M. This transient phosphorylation may allow factors that cause condensation of the chromatin during mitosis access to the chromatin. The condensation of chromatin, when linker histone is phosphorylated, may still be dependent on core histone deacetylation. In addition, opposite charges between deacetylated core histone and phosphorylated linker histone tails could facilitate interactions between these tails to allow a high degree of condensation (Roth and Allis, 1992).

Acetylation of Core histones

Reversible histone acetylation occurs at the epsilon amino group of specific lysine residues located in the basic N-terminal domains of core histones (Allfrey *et al.*, 1964; Allfrey, 1971; Allfrey, 1977; Doenecke and Gallwitz, 1982; Reeves, 1984; Matthews and Waterborg, 1985; Matthews, 1988). Histone acetyltransferases link the acetyl group of acetyl-CoA to specific lysine residues and this neutralizes positive charges within the N-terminal part of the histone molecule. This modification can be reversed by the action of deacetylases.

It is thought that histone acetylation is involved in transcription (van Holde, 1988; Ramanathan and Smerdon, 1989). Active c-*fos* and c-*myc* genes, and histone H2A and H4 gene chromatin in S phase contain highly acetylated histones H4, H3 and H2B in HeLa cells (Allegra *et al.*, 1987; Sterner *et al.*, 1987). The activation, repression and superinduction of murine fibroblast proto-oncogenes c-*fos* and c-*myc* occurs with rapid and reversible histone acetylation (Chen and Allfrey,

1987). In chicken ß-globin gene switching, highly acetylated histones occur in both transcriptionally active and in genes which are poised for transcription but as yet inactive (Hebbes *et al.*, 1992). In growing yeast all the genes are active or potentially active and the entire chromatin is highly acetylated (Lohr and Hereford, 1979; Davie *et al.*, 1981). During bird erythrocyte maturation there is a decreasing steady-state level of histone acetylation (Ruiz-Carrillo *et al.*, 1974), during which time transcriptional activity decreases. In this system the transcriptionally active histone H5 gene is associated with acetylated nucleosomal histones (Ridsdale *et al.*, 1990; Hendzel *et al.*, 1991).

A positive role has been suggested for the effect of histone acetylation on transcription factor binding. The transcription factor TFIIIA binds and forms ternary complexes with complete nucleosome cores that are enriched in acetylated histones (Lee *et al.*, 1993). Tryptic removal of the histone tails also allows TFIIIA binding, suggesting that the positively charged tails of unacetylated core histones repress TFIIIA binding (Lee *et al.*, 1993). These studies suggest that TFIIIA can bind DNA which is packaged with complete nucleosome cores, and that the binding is facilitated by acetylation of the core histones.

Histone hyperacetylation of the nucleosomal histones has several effects on chromatin, including minor effects on the stability of the nucleosome core particle (Loidl, 1988; Loidl and Grobner, 1987; Simpson, 1978a; Ausio and van Holde, 1986; Libertini *et al.*, 1988; Bode *et al.*, 1983; Imai *et al.*, 1986). Hyperacetylation of chromatin also changes the 30 nm fiber. Histone acetylation neutralizes basic lysine residues in the N-termini of the core histones, thereby preventing interaction with the negatively charged DNA backbone, and causing

unfolding of chromatin fibres (McGhee and Felsenfeld, 1980). Core histone hyperacetylation also prevents the N-terminal tails from interacting with neighbouring nucleosomes (Reeves, 1984) and alters the capacity of linker histones (H1 and H5) to condense transcriptionally active histone H5 gene chromatin in chicken erythrocytes (Ridsdale *et al.*, 1990; Hendzel *et al.*, 1991). Electron microscopy of hyperacetylated interphase chromatin of HeLa cells after butyrate treatment shows that the chromatin forms thinner fibres with an average diameter of 20 nm under physiological conditions (Annunziato *et al.*, 1988). The DNA between hyperacetylated core particles is more DNase I sensitive which suggests a less constrained conformation. Therefore, core histone acetylation can cause a moderate unfolding of chromatin (Hansen and Ausio, 1992), although this effect is in conjunction with linker histone effect (Allan *et al.*, 1982).

DNase I Hypersensitivity within Chromatin

The generalized changes in chromatin structure, such as DNase I sensitivity which occur throughout the chromatin domain of a poised or transcriptionally active gene are accompanied by specific localized nucleosome-disrupted or nucleosome-free regions in chromatin structure which occur at defined places within the loop. These localized changes are important because they allow specific binding of transcription factors, which are important for basal and activated transcription of the gene. For example, histone octamer displacement at the start site for transcription is known to be required for RNA polymerase II activity (Laybourn and Kadonaga, 1991). The nucleosome-disrupted sites are detected by hypersensitivity to DNase I digestion or chemical modification (approximately 100 X more than bulk

chromatin) and are referred to as DNase I hypersensitive (DH) sites (Reeves, 1984). Therefore, changes in the overall chromatin structure of a gene, which result in general DNase sensitivity are associated with hypersensitivity at specific sites within the gene domain, both of which are associated with transcriptional activation.

Binding sites for *trans*-acting factors are necessary for generating hypersensitivity. DH sites often exhibit multiple "hot" spots *in vivo* which correspond to DNase I protected regions caused by the presence of bound *trans*-acting factors (Elgin, 1988). DH sites have been mapped at positions known to affect transcription, including promoters, upstream activation sequences, enhancers of active or inducible genes and silencers of transcription (Elgin, 1988; Gross and Garrard, 1988). Several classes of nuclear proteins have been found with transcription-associated hypersensitive sites, including topoisomerases I and II, short-range nucleosome phasing elements, RNA polymerase II, and transcription factors (Elgin, 1988; Gross and Garrard, 1988).

Generation of Hypersensitive Sites

Felsenfeld (1992) hypothesizes that there are two kinds of mechanisms for disruption of chromatin structure based on transcription activation experiments. The two mechanisms; "dynamic competition" and the "pre-emptive" model, vary with regard to whether a transcription factor is able to bind DNA located in chromatin or requires free DNA for binding. It should be noted that not all transcription factors fall at these two extremes. For example, nucleosome phasing affects the ability of the transcription factor TFIIIA to recognize its binding site

(Hayes and Wolffe, 1992). In the pre-emptive model, access by *trans*-acting factors occurs during replication, when core histones or compact higher-order structure which blocks factor binding sites, is disrupted at the replication fork before chromatin assembly (Wolffe and Brown, 1988). After the DH site is established, the absence of large pools of unbound histones and the presence of active assembly mechanisms are important to maintaining the site (Felsenfeld, 1992).

Consistent with the pre-emptive model, *in vitro* transcription is observed if a preinitiation complex is formed on an adenovirus type-2 major late promoter (MLP) before, but not after chromatin assembly (Knezetic *et al.*, 1988). Pre-binding of TFIID or yeast TATA binding protein to the TATA box protects the promoter against subsequent nucleosome binding (Meisterernst *et al.*, 1990; Workman and Roeder, 1987). Tissue-specific factors bound near a ß-globin promoter are able to exclude nucleosomes *in vitro*, but are not able to displace them (Emerson and Felsenfeld, 1984). Human heat-shock factor is unable to bind nucleosomal DNA (Taylor *et al.*, 1991; Workman and Kingston, 1992). Nuclear factor I does not bind with its site in the mouse mammary tumour (MMTV) promoter until after the glucocorticoid receptor has bound, as discussed below (Pina *et al.*, 1990; Archer *et al.*, 1991). *In vivo* evidence for this model is that many kinds of dividing differentiated cells can be reprogrammed (Blau and Baltimore, 1991; Miller and Nasmyth, 1984).

Dynamic competition does not require DNA replication since transcription factors and histone octamers compete for binding sites in the non-dividing cell. Factor binding causes nucleosome core destabilization and displacement. *Trans*-acting factors that can bind to DNA when incorporated into nucleosomes, probably

mediate the first steps in the disruption of repressive chromatin structures (Hayes and Wolffe, 1992) and establishment of occupancy of these regulatory sequences in chromatin (Felsenfeld, 1992).

In support of the dynamic competition model, evidence suggests that nucleosome displacement associated with transcription of certain genes does not require DNA replication. For example, c-*fos*, c-*myc* and the MMTV promoters are activated minutes after induction, which provides insufficient time for replication (Cordingley *et al.*, 1987). Replication-independent nucleosome disruption has been demonstrated for the yeast PHO5 promoter (Schmid *et al.*, 1992; Almer *et al.*, 1986; Fascher *et al.*,1990). The extent of chromatin disruption in yeast by estrogen receptor derivatives roughly correlates with the strength of the activation domain in transcription activation (Pham *et al.*, 1991). Taylor *et al.* (1991) showed that GAL4 and GAL4 derivatives can bind to DNA in chromatin, with an estimated affinity that is 100-fold lower (for a single site) or 10-fold lower (for five sites in tandem) than binding to the corresponding naked DNA template.

In vitro evidence for dynamic competition is that binding of certain transcription factors results in the formation of ternary complexes containing factors, histones and DNA. Studies of the MMTV promoter have demonstrated that the glucocorticoid receptor forms a ternary complex with a nucleosome positioned over the promoter (Perlmann and Wrange, 1988; Pina *et al.*, 1990; Archer *et al.*, 1991) and disrupts the nucleosome by a "hit and run" mechanism (Rigaud *et al.*, 1991). It is proposed that this binding alters the nucleosome so that nuclear factor I (NFI) and TFIID can bind to this promoter and activate transcription (Pina *et al.*,1990; Archer *et al.*, 1991; Cordingley *et al.*, 1987).

Nucleosome displacement by Gal4 via an unstable intermediate ternary complex was demonstrated by Workman and Kingston (1992). Nucleosome core destabilization by GAL4 did not require a transcriptional activation domain, which indicates that this displacement depends on the nature of the DNA binding domain (Taylor *et al.*, 1991). After binding, the activation domains of regulatory factors play a crucial role in a second competition for occupancy of the core promoter by nucleosomes or general transcription factors and RNA polymerase II (Workman *et al.*, 1991). Accessory factors might be involved in removing histones from factor-nucleosome complexes (Workman and Kingston, 1992). For example, GAL4-Vp16 requires a factor present in a yeast transcription extract to alleviate repression due to nucleosome cores present at the transcription start site or the TATA box of a yeast promoter (Lorch *et al.*, 1992).

In vivo evidence for the dynamic competition model comes from the fusion of heterogenous cells to form heterokaryons containing separate nuclei. In the absence of DNA replication, the cells were reprogrammed to produce distinct gene products (Chiu and Blau, 1984; Blau *et al.*, 1983; Baron and Maniatis, 1986; Bergman and Ringertz, 1990). In these experiments, *trans*-acting factors from one cell nucleus exerted a dominant effect on expression within the other nucleus. Activation of transcription was specific and not the result of a general activation of transcription (Baron and Maniatis, 1986).

A highly specific transient deacetylation, which would create a localized net positive charge in histones, has been proposed for the binding of regulatory proteins with acidic motifs to DNA (Lopez-Rodas *et al.*, 1993). The deacetylation would allow an interaction between the acidic region of the transcription factor and the

positive charge within the N-termini of the core histones. This binding is proposed to occur concomitant with a sequence specific recognition of DNA sites by regulatory proteins. The model attributes a structural role to the general histone acetylation within a chromatin loop and a regulatory function to the transient, localized histone deacetylation in the early transcriptional activation of genes.

Locus Control Regions

Locus control regions (LCRs) are *cis*-acting elements that make the expression of associated genes independent of their position within the genome and proportionate to the number of genes integrated into the chromatin (Grosveld *et al.*, 1987; Forrester *et al.*, 1987; Evans *et al.*, 1990; Townes and Behringer, 1990). Both human and chicken ß-globin LCR domains have been identified, some of which do not function as enhancers in transient expression assays, but only when stably integrated, suggesting that the activity is associated with chromatin structure and is not due to enhancer activity (Tuan *et al.*, 1989; Talbot *et al.*, 1990; Philipsen *et al.*, 1990; Talbot and Grosveld, 1991; Reitman *et al.*, 1990; Fraser *et al.*, 1990; Caterina *et al.*, 1991). Felsenfeld (1992) suggests that a primary role of the LCR is to enter into cooperative binding interactions that keep the promoter free of histones during replication. A similar role has been proposed by Wolffe (1990) for enhancers. Felsenfeld (1992) suggests that this is done through an 'overkill' in the multiple factor sites within the LCR.

All proteins that have been identified as binding to LCR sequences also bind to promoters and enhancers. In the human ß-globin LCR 5' HS 4, the minimal sequence needed to form a DH site in murine erythroleukemia (MEL) cells binds NF-

E2/AP-1, Sp1 and GATA-1 (Lowrey *et al.*, 1992). The overall activity of erythroid LCRs on linked gene expression appears to be dependent on three core DNA motifs: GATA, CACCC (GGTGG), and TGAGTCA (an AP-1 motif) (Talbot and Grosveld, 1991; Talbot *et al.*, 1990; Strauss and Orkin, 1992; Philipsen *et al.*, 1990). No single element of human ß-globin LCR 5' hypersensitive site 2 is necessary for LCR function (Enver *et al.*, 1990) and isolated NF-E2 sites, single or multimerized, do not have LCR activity (Talbot *et al.*, 1990). The conclusion is that the LCR hypersensitive site function is due to the collection of transcription factors binding to this region (Reitman *et al.*, 1993).

Nuclear Matrix

The nuclear matrix organizes chromatin into three dimensional loop domains (Ward and Coffey, 1990; Kalandadze *et al.*, 1990; Vogelstein *et al.*, 1980; Pardoll *et al.*, 1980). In the context of gene expression, the nuclear matrix brings nuclear activities that are responsible for transcription, RNA processing and transport together (van Driel *et al.*, 1991). Jackson and Cook (1985) provided evidence that transcription complexes were attached to the nucleoskeleton. Active genes have been found to be associated with the nuclear matrix only in cell types in which they are expressed (Getzenberg *et al.*, 1991). Transcriptionally active histone H5 and ß-globin, but not the transcriptionally inactive ϵ-globin, chromatin of chicken erythrocytes is associated with nuclear material found in the low salt residue (Delcuve and Davie, 1989; Hendzel *et al.*, 1991).

DNA sequences attached to the nuclear matrix are referred to as matrix (or scaffold) associated regions (MARs or SARs) (Gasser *et al.*, 1989; Gasser and

Laemmli, 1987; van Holde, 1988). MARs are usually about 200 base pairs in length and are A + T rich (Chou *et al.*, 1990). Most MARs contain consensus binding sequences for topoisomerase II, a matrix-associated protein (Earnshaw *et al.*, 1985; Berrios *et al.*, 1985). Several nuclear matrix associated sequence-specific DNA-binding proteins have been identified that bind to MAR elements (Gasser *et al.*, 1989; Dworetzky *et al.*, 1992), including transcription factors such as the *myc* protein, the large T antigen of the SV40 virus, and E1A from adenovirus (Eisenmann *et al.*, 1985; Staufenbiel and Deppert, 1983; Sarnow *et al.*, 1982; Waitz and Loidl, 1991). Hendzel *et al.* (1991) and Hendzel and Davie (1992) suggest that nuclear matrix compartmentalization of transcription processes is accomplished by the interaction of proteins of the nuclear matrix (including transcription factors, histone acetyltransferase and deacetylase) and MARs of DNA in transcriptionally active chromatin.

MARs are located, at least in some cases, close to enhancer elements and at the borders of chromatin domains (Cockerill and Garrard, 1986; Gasser and Laemmli, 1987; Zenke *et al.*, 1990; Bennet *et al.*, 1989; Phi-Van and Stratling 1988; Bode and Maass, 1988; Levywilson, and Fortier, 1989; Mirkovitch *et al.*, 1984; Jarman and Higgs, 1988). Similar to LCRs, MARs dampen the variations in gene expression level due to position effects at the integration site in the genome, and increase transcription levels after stable integration in the genome (Klehr *et al.*, 1991; Phi-Van *et al.*, 1990; Stief *et al.*, 1989). It is not known what the relationship is between LCRs and MARs. It may be that the shielding effect of MARs is due to their ability to bind to the nuclear matrix.

Looping, Long Range Activation of Transcription and Chromatin Structure

The basic idea of DNA looping is that a remote enhancer and a promoter interact physically with each other via proteins bound to these elements, thereby looping out the intervening spacer DNA and allowing a gathering together of recognition complexes (Serfling *et al.*, 1985; Picard and Schaffner, 1985; Ptashne, 1986; Ptashne, 1988). The looping model explains activation over large distances, the orientation independence of enhancers, and provides a mechanism whereby the effect of upstream promoter and enhancer elements can be funnelled through the basic transcription machinery. There are examples of *trans*-acting factors bound at widely separated sites which can interact with each other to form homodimers or heterodimers and induce loop formation in DNA. For example, homodimers of progesterone receptor (Theveny *et al.*, 1987), homotetramers Sp1 (Su *et al.*, 1991) or a heterodimer of Sp1 and the BPV enhancer E2 protein are able to form (Li *et al.*, 1991). Evidence from the chicken β^A-globin gene, suggests a physical association of the enhancer/LCR and promoter. Neither promoter nor enhancer/LCR is sufficient individually to open chromatin, but when both are present, each becomes hypersensitive and both are essential for activation (Reitman *et al.*, 1993). Felsenfeld (1992) envisions a physical promoter-enhancer/LCR activation and then recognition and activation of the closed promoter.

DNA sequences present in the linker DNA of chromatin, but separated by >1500 bp, will be superimposed in a 30 nm solenoidal structure. Thus when nucleosomes are present on the intervening DNA, it seems likely that transcription factors bound to proximal promoters may interact with those bound to distal sequence elements. Long-range activation of transcription by GAL4-VP16 protein

located 1300 base pairs upstream of the RNA start site was dependent on packaging of the template into histone H1-containing chromatin (Laybourn and Kadonaga, 1992). These results indicate that long-range activation of transcription can be achieved *in vitro* and is dependent on packaging of the template into chromatin.

Transcription

Basal Apparatus

Initial chromatographic fractionations and reconstitution assays demonstrated that the selectivity of RNA polymerase II (RNA pol II) initiation in crude extracts requires several components termed transcription factors (Matsui *et al.*, 1980; Samuels *et al.*, 1982; Dynan and Tjian, 1983; Price *et al.*, 1987; Tsai *et al.*, 1981; Lewis and Burgess, 1982). These general transcription factors which are required for basal transcription, were initially separated by chromatography on phosphocellulose. They were designated TFIIA for the flow-through factor, TFIIB for the factor in the 0.5 M KCl step, and TFIIC and TFIID for the two components present in the 1 M KCl fraction (Greenleaf *et al.*, 1976; Matsui *et al.*, 1980; Samuels *et al.*, 1982; Davison *et al.*, 1983). Component TFIIB was later shown to contain two activities, designated TFIIB and TFIIE respectively (Dignam *et al.*, 1983a) and TFIIE has been separated into TFIIE and TFIIF (Burton *et al.*, 1986; Flores *et al.*, 1988).

RNA Polymerase II

RNA pol II is a 500-600 kDa multisubunit enzyme, and is composed of two

large polypeptides and a collection of smaller proteins, the subunit composition of which is not clearly established (Sawadogo and Sentenac, 1990). Immunological cross-reactivity and DNA sequencing of cloned proteins show that the polypeptide composition seems highly conserved in lower eukaryotes, plants, insects, and vertebrates (Chambon, 1975; Roeder, 1976; Lewis and Burgess, 1982; Sentenac, 1985). Many of the different activities are highly interchangeable between species. This property has been useful in biochemical studies.

The largest subunit of RNA pol II is 220-240 kDa and is probably involved in binding the DNA template (Horikoshi *et al.*, 1983; Chuang and Chuang, 1987; Gundelfinger, 1983; Breant *et al.*, 1983; Carroll and Stollar, 1982; Horikoshi *et al.*, 1985). It is also involved in binding the nascent RNA chain (Gundelfinger, 1983; Bartholomew *et al.*, 1986). This subunit has a highly conserved, carboxy-terminal domain (CTD) which is subject to phosphorylation (Corden, 1990; Young, 1991), and consists of tandem repeats of the consensus heptapeptide sequence (Tyr-Ser-Pro-Thr-Ser-Pro-Ser). The tailpiece accounts for 10-20% of the subunit mass. Deletion analysis has shown that approximately half of the repeats in yeast, *Drosophila*, and mice are required for cell viability (Nonet *et al.*, 1987; Allison *et al.*, 1988; Bartolomei *et al.*, 1988; Zehring *et al.*, 1988).

The second largest subunit (140-150 kDa) has conserved immunological determinants. This subunit is probably involved in binding DNA (Gundelfinger, 1983), the nascent RNA chain (Bartholomew *et al.*, 1986) and the nucleotide substrates (Riva *et al.*, 1987; Grachev *et al.*, 1986). The smaller polypeptides of RNA pol II range in size from 10 to 40 kDa and vary in number from 6 to 10, depending on the investigators and the enzyme source.

Trans-Acting Factors

TFIIA from HeLa cells consists of three polypeptides of 34, 19, and 14 kDa (Samuels and Sharp, 1986; Usuda *et al.*, 1991; Cortes *et al*, 1992). The genes of the 32 and 13.5 kDa subunits of yeast have been isolated (Ranish *et al.*, 1992). Rat TFIIB is a single polypeptide of 35 kDa (Ha *et al.*, 1991; Conaway *et al.*, 1987). Both TFIIE and TFIIF have been cloned (Ohkuma *et al.*, 1991; Peterson *et al.*, 1991; Sumimoto *et al.*, 1991; Sopta *et al.*, 1989). Human TFIIF is a tetramer of two 74 kDa and two 30 kDa subunits, which prevent non-specific DNA-binding by RNA pol II. TFIIG is composed of TFIIH and TFIIJ. TFIIH has been shown to be composed of polypeptides of 92, 62 (cloned), 43, 40 and 25 kDa. TFIIH phosphorylates the CTD of RNA pol II. This kinase activity is stimulated by TFIIE. Human TFII-I is a 120 kDa protein.

TFIID

Native human and *Drosophila* TFIID is a multisubunit complex of ~700 kDa termed holo-TFIID composed of TATA-binding protein (TBP), an activity that binds specifically to the TATA box (Buratowski *et al.*, 1989; Van Dyke *et al.*, 1988; Pugh and Tjian, 1990; Tanese *et al.*, 1991; Dynlacht *et al.*, 1991; Conaway *et al.*, 1991) and several TATA-binding protein associated factors (TAFs), ranging in size from 28 to 250 kDa (TAF_{28} to TAF_{250}) (Zhou *et al.*, 1992). Yeast TFIID activity on the other hand, consists of only TBP (Buratowski *et al.*, 1988; Horikoshi *et al.*, 1989b). The various TBP genes have a distinct, bipartite structure with a highly conserved (80 - 90%) "core" domain of ~180 amino acids at the carboxyl-terminal end of the protein (amino acids 63-240) which is sufficient for binding to the TATA box and

functionally substitutes for the human TFIID fraction in directing basal transcription *in vitro* (Hahn *et al.*, 1989; Schmidt *et al.*, 1989b; Cavalinni *et al.*, 1989; Eisenmann *et al.*, 1989; Horikoshi *et al.*, 1989a; Peterson *et al.*, 1990; Horikoshi *et al.*, 1990; Pugh and Tjian, 1990). The amino-terminal end is variable in length with little or no sequence similarity among species.

Functions performed by holo-TFIID including activation of transcription by several classes of transcription factors and from RNA pol II promoters lacking a TATA box, is not supported by TBP alone, but requires the associated TAFs (Dynlacht *et al.*, 1991; Hoey *et al.*, 1993; Tanese *et al.*, 1991; Pugh and Tjian, 1990, 1991, 1992; Zhou *et al.*, 1992; Berger *et al.*, 1990; Kelleher *et al.*, 1990; Flanagan *et al.*, 1991; Meisterernst and Roeder, 1991; Meisterernst *et al.*, 1991). Thus, some of the TAFs are expected to function as coactivators and others may serve as bridging proteins that allow different coactivators to assemble into a complex containing TBP (Comai *et al.*, 1992; Taggart *et al.*, 1992; Lobo *et al.*, 1992; White and Jackson, 1992; Kassavetis *et al.*, 1992).

The components and structure of TFIID has been partially elucidated. Both human TBP (hTBP) and *Drosophila* TBP (dTBP) interact directly with the apparently homologous $hTAF_{250}$ and $dTAF_{250}$, respectively (Zhou *et al.*, 1993; Weinzierl *et al.*, 1993; Takada *et al.*, 1992; Ruppert *et al.*, 1993;). hTBP also binds to $hTAF_{125}$, but does not bind any of the smaller TAFs (Zhou *et al.*, 1993), similar to the results of Hoey *et al.* (1993) for dTBP and its TAFs. Therefore the remaining lower molecular weight TAFs likely interact with TBP indirectly, through TAF_{250} and TAF_{125}. On the basis of a comparison of the DNase I protected region produced by holo-TFIID on the adenovirus major late promoter to the DNase I protected region produced by

isolated TBP, Zhou *et al.* (1992) proposed a model in which TBP is at one side of the holo-TFIID complex.

Hoey *et al.* (1993) cloned the gene for *Drosophila* TAF_{110} ($dTAF_{110}$) and found that this protein interacts specifically with the activation domains of Sp1, but does not bind directly to dTBP. Weinzierl *et al.* (1993) assembled a partial TFIID complex containing a C-terminal truncated version of $dTAF_{250}$ (deltaN250), dTBP and $dTAF_{110}$. Their studies suggest that $dTAF_{II}250$ provides a direct link between the Sp1 coactivator $dTAF_{II}110$ and dTBP and thus might play a central role in coordinating and regulating the assembly of the TFIID complex *in vivo*. However, the Sp1-partial TFIID complex only slightly activates transcription when reconstituted with RNA pol II and the other basal factors in an *in vitro* assay. Weinzierl *et al.* (1993) speculate that activators such as Sp1, which contain multiple activation domains, may interact with more than one target or component of the TBP-TAF complex, and the other TAFs of TFIID may also be necessary for full coactivator activity. It may also be that the N-terminal sequences lacking in the deltaN250 truncated version of $dTAF_{II}250$ may be important for the assay.

Cis Elements

Cis elements are DNA sequences which are able to affect transcription of a gene. They are commonly determined through the use of such assays as CAT (chloramphenicol acetyl transferase) assays, in which DNA sequences of interest are subcloned into a plasmid containing the coding sequence for the CAT protein. The ability of the potential *cis* element to affect transcription of the CAT gene is then tested, by testing for activity of the final protein product, the CAT protein.

Cis elements are the specific DNA sequences to which transcription factors bind and as such are important, since their presence allows specific binding by transcription factors, which in turn affect transcription. As a research tool, *cis* elements are important since protein bound to the sites protect the region from DNase I digestion and can therefore be detected in DNase I protection assays. Alternatively, the DNA sequence of the *cis* element can be used for other experiments. One example is electrophoretic mobility shift assays in which binding of proteins to DNA affects the mobility of the DNA in a non-denaturing electrophoretic system. Most transcription factors bind to a consensus DNA sequence and therefore knowledge of the DNA sequence of an enhancer or promoter allows researchers to gain an understanding of the transcription factors that are acting on a gene.

The promoter is defined as the control region in the immediate vicinity of a transcription start site. Therefore this DNA element includes transcription factor binding motifs that are near the initiation site for RNA pol II. The operational definition of an enhancer is a region that regulates a promoter from a distance and in an orientation-independent fashion. Typically enhancers and promoters have one or more recognition sites for transcriptional activator proteins, each consisting of 7-20 bp of DNA. The TATAAA(A) consensus sequence or TATA box, which is typically located 25 to 30 base pairs upstream of the transcriptional start site or initiator element in many, but not all RNA pol II promoters is the binding site for TFIID and as such is a determinant of the site of initiation (Corden *et al.*, 1980; Breathnach and Chambon, 1981; Struhl, 1987; Nakajima *et al.*, 1988).

The TATA box is required for basal transcription activity in many genes.

Cis-acting DNA motifs, in addition to the TATA box, and typically upstream of the TATA box are important for creating a promoter that may be weak by itself but will respond strongly to a remote enhancer (Kuhl et a., 1987; Westin *et al.*, 1987). For example in many globin genes, motifs such as a CCAAT box or GC box, located upstream from the TATA box are required for transcription (Grosschedl and Birnstiel, 1980a, 1980b; Dierks *et al.*, 1983; McKnight and Tjian, 1986). Addition of several further sequence motifs upstream of the TATA box, however, can result in the build-up of an enhancer/promoter region that does not require the presence of a remote enhancer (Westin *et al.*, 1987).

Promoter and enhancers can overlap both physically and functionally. For example, many enhancer elements studied so far are also found to stimulate transcription when inserted directly upstream of a TATA box. Many transcription factors bind both to promoters and enhancers and may interact with the cellular transcriptional machinery in fundamentally the same way.

Formation of Basal Initiation Complex

After nucleosomes have been removed or modified so that transcription factors may bind, there are a number of steps required for transcription by RNA polymerase II. These steps are thought to include (1) the assembly of a preinitiation (complete closed) complex, which includes the polymerase, general transcription factors, sequence-specific transcription factors (other than the general transcription factors) and accessory proteins, on the RNA polymerase promoter site (Conaway and Conaway, 1991; Zawel and Reinberg, 1993); (2) two or more ATP-dependent events, including multiple phosphorylation of the CTD of bound RNA pol

II, by the kinase activity present in TFIIH (Bunick *et al.*, 1982; Sawadogo and Roeder, 1984; Rappaport and Weinmann, 1987; Peterson and Tjian, 1992) and unwinding or melting of the DNA around the start site to expose the template strand likely by TFIIF; this assembly is called the open complex (Wang *et al.*, 1992a); and (3) initiation and elongation of the RNA transcript in the presence of NTPs supported by TFIIF, to produce product. Concomitant with elongation there is a reclosing of the DNA around the start site.

In the absence of activator, the assembly of complete closed complexes on templates containing the TATA element involves the sequential addition of the general transcription factors IID, IIA, IIB, RNA pol II/IIF and IIE respectively (Buratowski *et al.*, 1989; Flores *et al.*, 1991; Greenblatt, 1991a; 1991b; Van Dyke *et al.*, 1988; Horikoshi *et al.*, 1988b; Reinberg *et al.*, 1987, Reinberg and Roeder, 1987; Roeder, 1991; Zawel and Reinberg, 1993). The first step, binding of TFIID to the TATA box may require additional factors in the absence of a TATA box (Nakajima *et al.*, 1988; Davison *et al.*, 1983; Fire *et al.*, 1984; Buratowski *et al.*, 1989). TFIID binding to the TATA elements results in template commitment (Buratowski *et al.*, 1988; Cavallini *et al.*, 1988; Conaway and Conaway, 1989; Reinberg *et al.*, 1987; Samuels and Sharp, 1986), which is resistant to challenge by poly (dI·dC) or other DNA templates (Davison *et al.*, 1983; Fire *et al.*, 1984; Hawley and Roeder, 1985; Carthew *et al.*, 1988) and may be stable through multiple rounds of transcription (Van Dyke *et al.*, 1988; 1989). The formation of stable complexes seems to be facilitated by TFIIA, specifically when the TFIID concentration is low (Davison *et al.*, 1983; Reinberg *et al.*, 1987; Samuels and Sharp, 1986; Matsui *et al.*, 1980; Egly *et al.*, 1984; Buratowski *et al.*, 1988). In

other experiments TFIID appears to act alone (Davison *et al.*, 1983; Conaway and Conaway, 1989; Sawadogo and Roeder, 1985; Van Dyke *et al.*, 1989). Recent evidence shows that there is a direct interaction between TFIIA and TBP and that TFIIA enhances TFIID binding to DNA by eliminating an otherwise inhibitory effect of the nonconserved N-terminus of TBP in *S. cerevisiae* (Lee *et al.*, 1992). Their data also suggest that TFIIA plays a role in allowing TFIID to recognize a wider variety of promoters. The effect or lack of effect of TFIIA on transcription as well as on the formation of stable complexes may also relate to the presence of contaminating repressors which compete with TFIIA for binding to the TFIID complex in the *in vitro* systems (Meisterernst and Roeder, 1991). In some systems TFIIA is replaced by other factors. For example, the TATA-less H5 promoter of duck erythrocytes requires the duck erythrocyte TF upstream stimulating factors (eUSF) (Bungert *et al.*, 1992a; 1992b). The factor TFII-I, which binds to an initiator element of Ad2MLP can substitute for TFIIA to direct TFIID binding to Ad2MLP (Roy *et al.*, 1991).

It is proposed that TFIID with help from TFIIA then recruits TFIIB, followed by RNA pol II, concomitant with TFIIF and finally TFIIE to form the complete preinitiation complex (Davison *et al.*, 1983; Fire *et al.*, 1984; Reinberg and Roeder, 1987; Buratowski *et al.*, 1989; Conaway *et al.*, 1990; Flores *et al.*, 1991). RNA pol II is brought into the complex as the non-phosphorylated form, in part, by an interaction of TBP with the CTD (Usheva et. al., 1992; Lu *et al.*, 1991). The interaction of TBP with the CTD may explain the requirement of TBP in transcription from TATA-less promoters (Smale *et al.*, 1990; Carcamo *et al.*, 1989; 1991; Pugh and Tjian, 1991). The complete preinitiation complex is in all likelihood identical to

the "rapid start" or "activated" complex (Fire *et al.*, 1984; Hawley and Roeder, 1985; Carthew *et al.*, 1988) identified by kinetic studies, or to the "complex O" (Dietrich *et al.*, 1986; Cai and Luse, 1987). Once the preinitiation complex is formed, RNA pol II is phosphorylated by a protein kinase present in TFIIH (Laybourn and Dahmus, 1989; 1990; Lu *et al.*, 1991), and this results in the dissociation of RNA pol II and associated factors from TFIID (Buratowski *et al.*, 1989; Maldonado *et al.*, 1990; Cortes *et al.*, 1992). This "disengagement" allows RNA polymerase II to leave the promoter as it begins to synthesize RNA (Kim and Dahmus, 1986; Cadena and Dahmus, 1987; Laybourn and Dahmus,1990). It should be noted that various interactions take place between TFIIB, TFIIE and RNA pol II in solution, although they are not stable enough to be detected by DNase I protection assays. This may indicate that the order of transcription factor addition to form the complete preinitiation complex may be somewhat flexible (Van Dyke *et al.*, 1988; Sawadogo and Sentenac, 1990).

Effect of Transcription Factors

As discussed above, general transcription factors are involved in establishing the multisubunit initiation complex and promoting basal level transcription. Proteins interacting with upstream promoter elements (UPEs) and/or enhancers are thought to stabilize and/or facilitate the formation of the active initiation complex, thereby increasing the level of transcription (Gralla, 1990; 1991).

Transcriptional activators are also thought to act at decision points between the formation of productive or nonproductive complexes. Fractions used for the *in vitro* studies may have contaminating negative cofactors which compete with TFIIA

to bind TFIID and repress transcription (Meisterernst and Roeder, 1991; Meisterernst *et al.*, 1991; Inostroza *et al.*, 1992; Wang *et al.*, 1992b). The negative cofactors might represent TAFs for RNA pol I and III, which also require TBP (Schultz *et al.*, 1992; Simmen *et al.*, 1991; Margottin *et al.*, 1991), and are thought to associate with distinct TAFs (Timmers and Sharp, 1991; Comai *et al.*, 1992). The role of activator and TFIIA would then be to prevent this binding, thereby promoting formation of a RNA pol II complex (Wang *et al.*, 1992b). Therefore, transcriptional activators, including Sp1 (and repressors) which affect the final extent of productive preinitiation complex formation, rather than the observed rate constant for formation of productive complexes are thought to influence whether the complex forms a productive or a non-productive complex (Carcamo *et al.*, 1989; Schmidt *et al.*, 1989a; Ayer and Dynan 1990; Katagiri *et al.*, 1990; Johnson and Krasnow, 1992; White *et al.*, 1992). However, it is not known whether there are branches during transcriptional initiation *in vivo* since the contaminating fractions thought to be present in the *in vitro* assays may not have access to the RNA pol II complexes *in vivo*.

The basal factor TFIID is thought to be a key link between promoter-specific transcription factors and the RNA pol II basal complex (Sawadogo and Roeder, 1985; Horikoshi *et al.*, 1988a; Horikoshi *et al.*, 1988b; Stringer *et al.*, 1990; Ptashne and Gann, 1990; Ingles *et al.*, 1991; Peterson *et al.*, 1990). Transcription factors like USF, AP-1 and CREB bind to the promoter cooperatively with TFIID (Sawadogo and Roeder, 1985; Sawadogo, 1988; Garcia *et al.*, 1987). In addition, yeast GAL4 and human ATF factors alter the interaction of human TFIID with the TATA box region (Horikoshi *et al.*, 1988a; 1988b; Hai *et al.*, 1988; Lin *et al.*,

1988). A herpes-virus (pseudorabies) immediate-early protein enhances transcription *in vitro* by facilitating TFIID binding to the promoter (Abmayr *et al.*, 1988; Workman *et al.*, 1988). Thus, in several systems, the interaction of TFIID with the TATA box has been implicated as a target for regulatory factors.

In addition to affecting the initial interaction of TFIID with the TATA box, activators may help to stabilize a TFIID-TFIIA-activator complex through multiple rounds of transcription. *In vitro*, the association of TFIID with the TATA box is slow and rate-limiting during transcription initiation, whereas assembly of an open complex from a TFIID-TFIIA-activator complex is rapid (Reinberg and Roeder, 1987; Schmidt *et al.*, 1989). This would allow subsequent rounds of assembly and transcription to also be rapid, as has been observed (Wang *et al.*, 1992b).

Addition of TFIIB also prevents diversion to repressed complexes. Thus, once TFIIA is bound, recruitment of TFIIB by activator (Lin *et al.*, 1991) could further protect against competition by negative cofactors (Meisterernst and Roeder, 1991; Meisterernst *et al.*, 1991). Activator may be involved at more than one step, which could account for its ability to activate transcription synergistically (Meiklejohn and Gralla, 1989; Carey *et al.*, 1990a, 1990b). For example one molecule of VP-16 which interacts with both TFIID and TFIIB could enhance the stability of TFIID binding and another could enhance the stability of TFIIB binding (Lin and Green, 1991; Lin *et al.*, 1991; Stringer *et al.*, 1990; Ingles *et al.*, 1991). Binding of TFIIB has also been proposed to be a rate-limiting step in the assembly pathway (Lin and Green, 1991).

Alternatively, distinct transcription factors may act synergistically to increase transcription. For example, within *cis*-acting elements there are frequently sites for

constitutive transcriptional activator proteins. These binding sites may be important for creating low levels of transcription. They may also help to create an environment where overall transcription is highly responsive to the presence of inducible proteins bound to nearby sites.

Coactivators

Current evidence indicates that enhancer and promoter-bound proteins may not interact directly with each other but require intermediary proteins, such as adaptors or coactivators, to effect transcriptional activation (Pugh and Tjian, 1992; Ham *et al.*, 1992). A coactivator is defined as a protein that acts with another transcription activator to directly interact with one of the basic transcription factors. An adaptor refers to a protein that binds to a sequence-specific DNA-binding protein and contains an activation region that can directly interact with one of the basic transcription factors (Martin, 1991). As discussed above, the TAFs of TFIID are thought to be coactivators.

The experiments of Berger *et al.* (1990) provided evidence for the existence of adaptors in activated transcription. They demonstrated that GAL4-VP16 is able to interfere with transcription of the heterologous dA:dT UAS-promoter sequence through two mechanisms. The first: *cis*-inhibition of basal-level transcription was probably due to a non-specific affinity for DNA by GAL4-VP16. *Trans*-inhibition was the second mechanism. In this experiment non-specific DNA added to the assay to interact with GAL4-VP16 prevented non-specific binding of GAL4-VP16 to the heterologous dA:dT UAS promoter. Under these conditions basal level, but not activated transcription was restored. An adaptor was invoked to explain this result.

The GAL4-VP16 in the system was thought to have "mopped up" coactivator required by the transcription factor which acts through the dA:dT UAS to increase transcription. Kelleher *et al.* (1990) found that the inhibition can be relieved *in vitro* by a crude yeast nuclear preparation but not by any of the known (purified) basal transcription factors, including TFIID.

Dusserre and Mermod (1992) purified cofactors of NF1 which were required for both antirepression and transactivation. Croston *et al.* (1992) found that in addition to the requirement by GAL4-VP16 for a cofactor to activate transcription, antirepression also required a cofactor. They suggest a model in which their cofactor acts as a histone H1 sink, since RNA can partially substitute for the activity. Therefore the cofactor is responsible for accepting H1, which then allows access of the basal transcription factors to the DNA template.

Development of Erythrocytes

During maturation of an organism, two distinct changes occur in the process of erythropoiesis; first, erythroid cells express different globin proteins at different developmental stages; and second, another level of control causes globin gene expression to be activated in a single cell lineage so that a pluripotent stem cell differentiates to form erythrocytes.

Developmental Switching

The developmental program of hematopoiesis is conserved throughout vertebrate evolution (Zon *et al.*, 1991). The site of blood formation changes sequentially during development from yolk sac in the early embryo, to the liver in

the fetus, and finally to bone marrow in the adult (Ingram, 1981; Tobin *et al.*, 1981; Groudine and Weintraub, 1981). The earliest events determining the hematopoietic program in vertebrates have yet to be defined. Hematopoietic commitment begins during early gastrulation, at about 20 to 23 hours well before blood formation is histologically evident (Groudine and Weintraub; 1981).

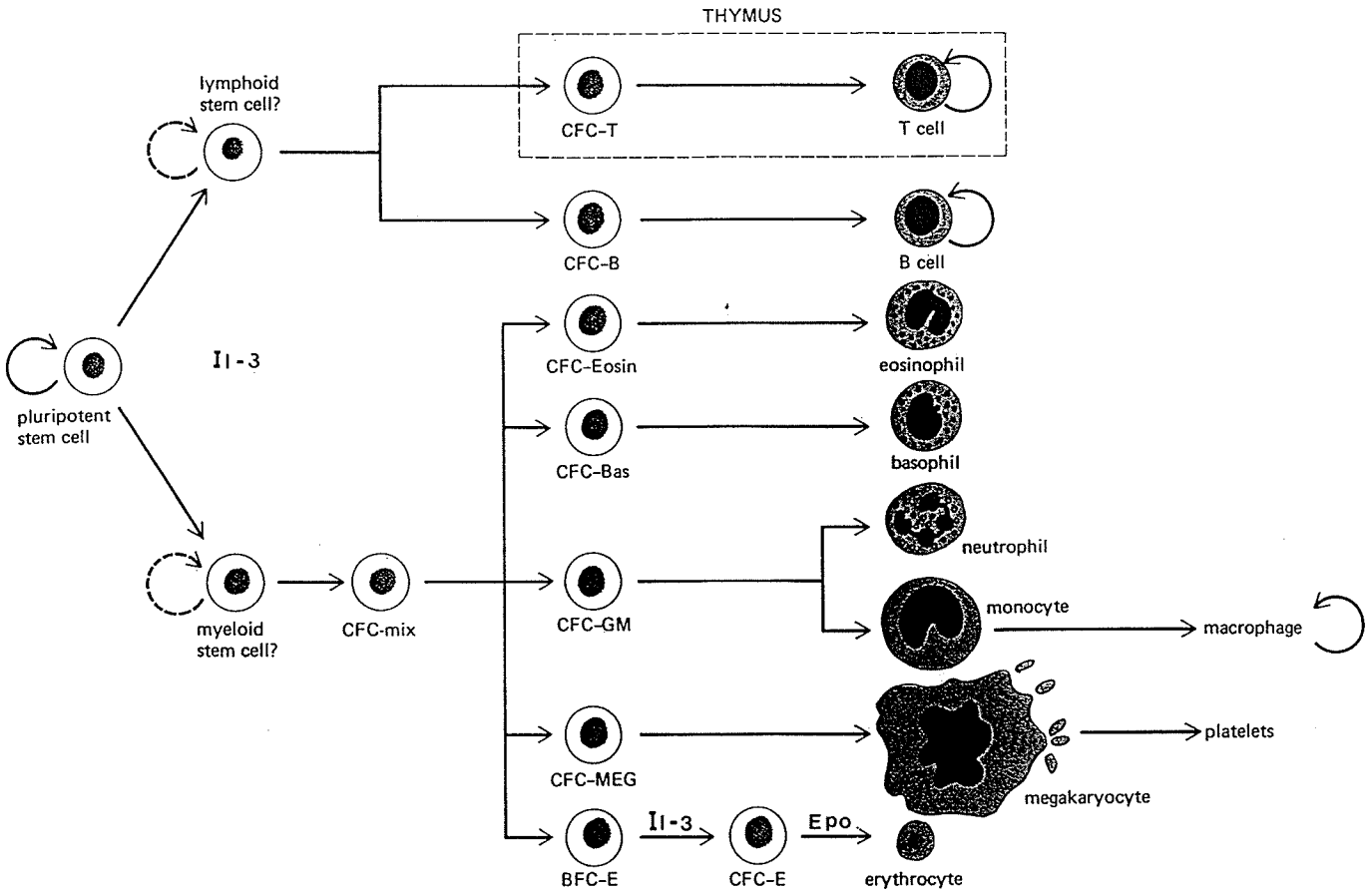
The first hemoglobinized cells appear in the blood islands in chicken embryos at 35 hours (Romanoff, 1960; Groudine and Weintraub, 1981), and these primitive cells produce embryonic hemoglobins; the alpha-globin genes pi, A, and D and the two embryonic ß-globin genes rho and ϵ (Bruns and Ingram, 1973). Beginning about 5 days after incubation, this primitive group of erythroid cells is replaced by definitive red blood cells, which synthesize adult hemoglobins: the alpha-globins A and D and the ß-globins H and A (Brown and Ingram, 1974). The transcription of the latter two genes in red blood cells arising from the bone marrow is maximal at about days 9 - 14 after fertilization. After hatching, a second type of definitive erythrocyte appears in which the $ß^H$-globin gene is silenced while $ß^A$-globin gene expression is maintained (Roninson and Ingram, 1983; Landes *et al.*, 1982; Chapman and Tobin, 1979). This changing pattern of globin isotypes during development is referred to as hemoglobin switching. The "switch" in hemoglobin synthesis in vertebrate organisms may occur at a particular stage of embryonic development in the pluripotent hematopoietic stem cell (Ingram, 1972; Weatherall *et al.*, 1976) or committed cells of the erythroid lineage may also be able to switch (Papayannopoulou *et al.*, 1986).

Differentiation

In both embryos and adults, the pluripotent stem cell gives rise to multipotential progenitors of the lymphoid and myeloid lineages (Figure 2). These lineages become progressively restricted in developmental potential and one of the lineages the myeloid stem cell gives rise to is the erythroid lineage (Figure 2) (Till *et al.*, 1964; Harrison 1982a,b; Dexter and Spooncer, 1987). There are at least two committed erythroid progenitor stages. The first, known as the burst-forming unit-erythroid (BFU-E) stage gives rise to the colony-forming unit-erythroid (CFU-E) stage. These progenitors are thought to divide rapidly for a limited number of times. Division occurs until the cells reach the erythroblast stages of terminal differentiation. With maturation they divide no further (Till and McCulloch, 1980). During the later stages of erythroid differentiation, various characteristic erythroid proteins, including hemoglobin are synthesized (Lazarides, 1987; Gasaryan, 1982).

The colony stimulating factor (CSF) interleukin 3 (IL-3), promotes the survival and proliferation of pluripotent stem cells and of most types of committed progenitor cells. In its presence, large erythroid colonies develop from single cultured BFU-Es of bone marrow. The BFU-E is distinct from the pluripotent stem cell in that it has a limited capacity to proliferate and gives rise to colonies that contain erythrocytes only, even under culture conditions that enable other progenitor cells to give rise to other classes of differentiated blood cells. Its progeny must go through as many as 12 divisions before they become mature erythrocytes. BFU-E is relatively insensitive to erythropoietin (Epo), and the proliferation of BFU-E does not depend on Epo, although the Epo receptor (EpoR) is expressed at low levels (Eaves and Eaves, 1984; Fukamachi *et al.*, 1987; Noguchi

Figure 2. Scheme of erythropoiesis. The haemopoietic stem cell gives rise to the lymphoid and myeloid lineages. The myeloid lineages gives rise to the erythrocyte via the intermediate progenitor cells (BFU-Es and CFU-Es) as summarized. Sites of action of interleukin-3 (Il-3) and erythropoietin (Epo) in the erythroid lineage are indicated. (Modified from Alberts *et al.* (1989).

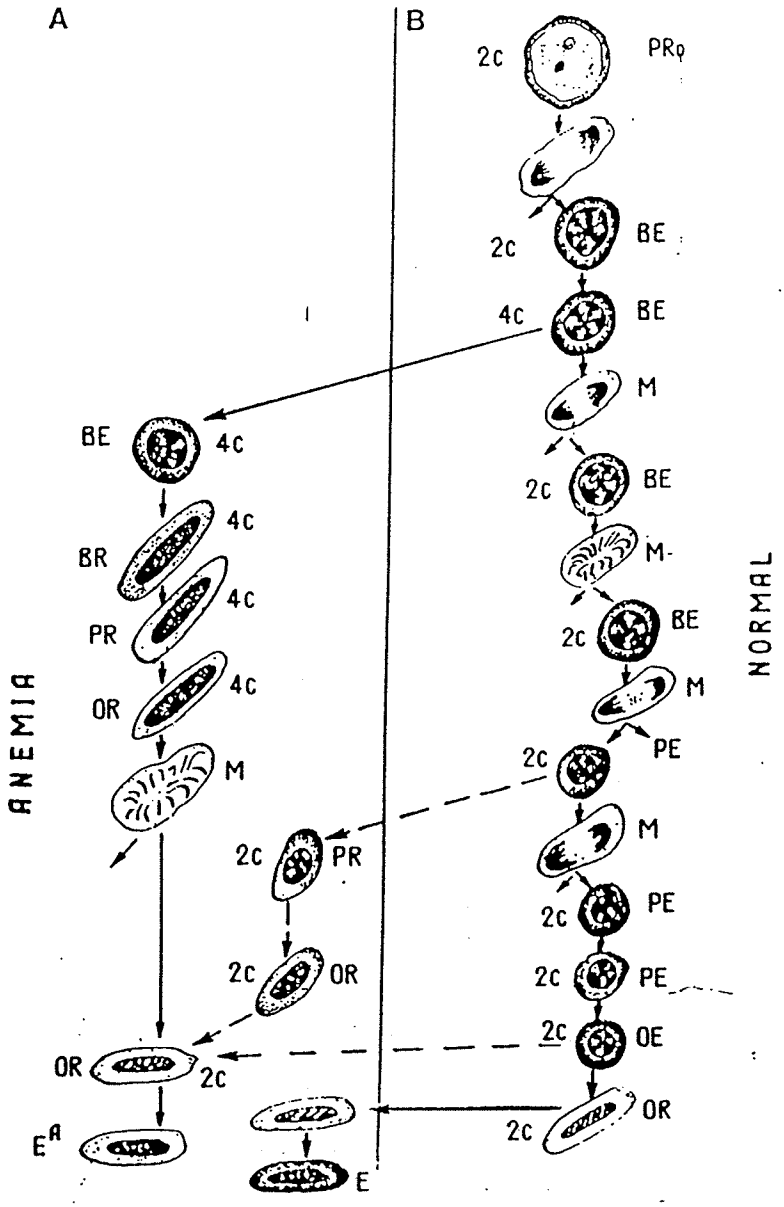

THYMUS
lymphoid stem cell?
CFC-T
T cell
CFC-B
B cell
CFC-Eosin
eosinophil
IL-3
pluripotent stem cell
CFC-Bas
basophil
neutrophil
myeloid stem cell?
CFC-mix
CFC-GM
monocyte
macrophage
platelets
CFC-MEG
megakaryocyte
IL-3
Epo
BFC-E
CFC-E
erythrocyte

et al., 1987).

At the next stage of erythroid development, the CFU-E is highly sensitive to Epo. In factor-dependent cell lines, the major regulatory step determining the erythroid-specific response to Epo is the efficiency of EpoR translocation to the cell surface (Miggliaccio *et al.*, 1991). CFU-E cells can be identified by culturing single bone marrow cells in a semisolid matrix in the presence of erythropoietin. In a few days, smaller colonies (than the BFU-E derived colonies) appear, which give rise to mature erythrocytes after about six division cycles. Physiologically, a lack of oxygen or a shortage of erythrocytes stimulates cells in the kidney to synthesize and secrete increased amounts of Epo into the bloodstream, thereby stimulating CFU-E maturation (Burgess and Nicola, 1983; Adamson, 1968). The CFU-Es do not yet contain hemoglobin, and must progress through the erythroblast stages (Figure 3).

An early effect of erythropoietin on the CFU-E is to stimulate the selective production of RNA synthesis. The histone H5 gene is already expressed at this stage (Affolter *et al.*, 1987). The mRNA molecules that code for globin appear later, indicating that hemoglobin formation is the end result of an involved differentiation process. Maturation of the erythroblast involves a limited number of cell divisions (Figure 3). During this time the cell produces erythroid-specific proteins and progresses through a number of stages as nuclear and cytoplasmic RNA amounts decrease (Schweiger, 1962; Sinclair and Brasch, 1975; Zentgraph *et al.*, 1975; Harrison *et al.*, 1988). The decrease in cytoplasmic RNA is seen in decreasing basophilic staining of maturing erythrocytes and lower amounts of polyribosomes and ribosomes in electron micrographs (Grasso and Woodard, 1966;

Figure 3. The pathways of erythroid differentiation in normal and acutely anemic adult pigeons. (A) Peripheral blood; (B) bone marrow. PRo, Proerythroblast; BE, basophilic erythroblast; PE, polychromatic erythroblast; OE, orthochromatic erythroblast; BR, basophilic reticulocyte; PR, polychromatic reticulocyte; OR, orthochromatic reticulocyte; E, erythrocyte (normal); E^A, erythrocyte formed in anemic condition; M, mitosis; C, ploidy (DNA equivalent). Reproduced from Gasaryan (1982).


A
B
ANEMIA
NORMAL
2c
PRo
2c
BE
4c
BE
M
BE
4c
2c
BE
4c
BR
M
4c
2c
BE
PR
M
4c
2c
PE
OR
M
2c
PR
M
2c
PE
2c
OR
2c
PE
2c
OE
OR
2c
2c
OR
E^A
E

Tooze and Davies, 1967; Small and Davies, 1972). Hence, early erythroblasts (proerythroblasts) have high RNA content and in sequence basophilic, polychromatic and orthochromatic erythroblasts have correspondingly less RNA at each stage. Mature erythrocytes do not show polyribosomes in electron micrographs (Gasaryan, 1982).

During normal erythropoiesis maturation of the red blood cell takes place in the bone marrow. At the last stage, the orthochromatic erythroblast becomes elongated before it is released from the bone marrow into blood circulation and becomes a mature erythrocyte. During anemia the basophilic erythroblast is released into the bloodstream and becomes elongated to form a basophilic reticulocyte (Figure 3; Scherrer *et al.*, 1966; Gasaryan, 1982). The cell continues to mature in a process known as "reserve erythropoiesis" (Gasaryan, 1982). "Reserve erythropoiesis" differs from normal erythropoiesis in that only one mitotic division takes place, and the final hemoglobin content of these cells is reduced by about 30% (Gasaryan, 1982). These differences between normal and "reserve erythropoiesis" suggest that gene activity may be somewhat perturbed in the anemic bird (Gasaryan, 1982).

The mechanisms responsible for decision-making in hematopoietic development likely consist of two interrelated pathways; one involving nuclear regulatory proteins (particularly GATA-1) which directly control gene expression, and a second involving signal transduction from growth factors (especially Epo) (Zon *et al.*, 1991b; Heberlein *et al.*, 1992; Chiba *et al.*, 1991; Tsai *et al.*, 1991). The two mechanisms do not act in isolation, each affects the activity of the other to form a positive autoregulatory loop. Relatively low levels of GATA-1 (or GATA-

2, discussed below), the amounts of which are enough to transactivate the EpoR gene, are triggered to express by unknown differentiation stimuli in the early stage of hematopoiesis. The presence of the GATA-1 causes activation of the EpoR promoter, which has GATA-1 binding sites and is strongly activated by GATA-1 in all species (Kuramochi *et al.*, 1990; Youssoufian *et al.*, 1990). Zon *et al.* (1991b) showed that the haematopoietic-specific transcription factor GATA-1 is necessary, and indeed sufficient as the sole cell-restricted regulator, for activation of the EpoR promoter in fibroblast transfection assays. There are also GATA-binding sites in the promoters of GATA-1 genes which provide a positive regulatory loop by which GATA-1 locks the promoter into an "on" state (Schwartzbauer *et al.*, 1992; Ho *et al.*, 1989; Hannon *et al.*, 1991; Zon *et al.*, 1991b; Heberlein *et al.*, 1992; Chiba *et al.*, 1991; Tsai *et al.*, 1991). Erythropoietin-mediated signals enhance GATA-1 gene expression in BFU-E and CFU-Es. Thus GATA-1 ensures continued survival of erythroid progenitors in the presence of Epo. The presence of GATA-1 also programs the progressive increase in mature protein content during cellular maturation. Hence, a cell-specific transcription factor and a cell-surface receptor effectively cooperate to promote survival and maturation of erythroid progenitors.

Erythroid Transcription Factors and Motifs

GATA-1

Structure of GATA-Binding Proteins

GATA-1 is a 37 - 39 kDa protein (Perkins *et al.*, 1989; Evans and Felsenfeld, 1989). GATA-1 may first be expressed at the multipotential CFU-GM stage, and then subsequently downregulated along the myeloid pathway, and

Figure 4. Amino acid sequence of GATA-1 (Evans and Felsenfeld, 1989). The zinc fingers are indicated in shadow. The three activation domains are indicated by double underline. The DNA binding polypeptide studied by Omichinski *et al.* (1993b) through NMR is from amino acids 156 to 218. The presumptive nuclear targeting signal is underlined.

Met Glu Phe Val Ala Leu Gly Gly Pro Asp Ala Gly 12

Ser Pro Thr Pro Phe Pro Asp Gly Ala Gly Ala Phe 24

Leu Gly Leu Gly Gly Gly Glu Arg Thr Glu Ala Gly 36

Gly Leu Leu Ala Ser Tyr Pro Pro Ser Gly Arg Val 48

Ser Leu Val Pro Trp Ala Asp Thr Gly Thr Leu Gly 60

Thr Pro Gln Trp Val Pro Pro Ala Thr Gln Met Glu 72

Pro Pro His Tyr Leu Glu Leu Leu Gln Pro Pro Arg 84

Gly Ser Pro Pro His Pro Ser Ser Gly Pro Leu Leu 96

Pro Leu Ser Ser Gly Pro Pro Pro Cys Glu Ala Arg 108

Glu Cys Val Asn Cys Gly Ala Thr Ala Thr Pro Leu 120

Trp Arg Arg Asp Gly Thr Gly His Tyr Leu Cys Asn 132

Ala Cys Gly Leu Tyr His Arg Leu Asn Gly Gln Asn 144

Arg Pro Leu Ile Arg Pro Lys Lys Arg Leu Leu Val 156

Ser Lys Arg Ala Gly Thr Val Cys Ser Asn Cys Gln 168

Thr Ser Thr Thr Thr Leu Trp Arg Arg Ser Pro Met 180

Gly Asp Pro Val Cys Asn Ala Cys Gly Leu Tyr Tyr 192

Lys Leu His Gln Val Asn Arg Pro Leu Thr Met Arg 204

Lys Asp Gly Ile Gln Thr Arg Asn Arg Lys Val Ser 216

Ser Lys Gly Lys Lys Arg Arg Pro Pro Gly Gly Gly 228

Asn Pro Ser Ala Thr Ala Gly Gly Gly Ala Pro Met 240

Gly Gly Gly Gly Asp Pro Ser Met Pro Pro Pro Pro 252

Pro Pro Pro Ala Ala Ala Pro Pro Gln Ser Asp Ala 264

Leu Tyr Ala Leu Gly Pro Val Val Leu Ser Gly His 276

Phe Leu Pro Phe Gly Asn Ser Gly Gly Phe Phe Gly 288

Gly Gly Ala Gly Gly Tyr Thr Ala Pro Pro Gly Leu 300

Ser Pro Gln Ile 304

maintained along the erythroid, megakaryocytic and mast cell lineages (Figure 2) (Orkin, 1992). Chicken GATA-1 was cloned by Evans and Felsenfeld (1989), who showed that it was a basic protein with a net positive charge of 12, which was able to increase expression (by 75%) of the chicken α^D-globin gene promoter containing a single GATA binding site. The exceedingly basic region (KGKRR) may be a nuclear localization signal (Dingwall and Laskey, 1991) (Figure 4). The protein has two related but nonidentical zinc fingers (Cys_{110} to Cys_{134} and Cys_{164} to Cys_{188}) of the form $CXNCX_4TXLWRRX_3GX_3CNAC$, each followed by a basic domain. The two finger regions have a high degree of similarity: 55% of the amino acids from residues 110-147 and 163-202 are identical (well beyond the cysteine cores). The region downstream of the cysteines (located on the same exon) is well conserved among species although it is not well conserved between fingers I and II (Yang and Evans, 1992). The pattern of amino acids conserved within or between the GATA fingers, does not resemble any of the conserved patterns found in other finger proteins, suggesting that GATA proteins are a unique form of zinc finger protein. Finger II and the basic downstream regions of it are required for DNA-binding activity (Yang and Evans, 1992). The C-terminal finger binds a GATA site with a specific association constant of 1.2×10^8 as compared with values of about 10^9 for the full-length GATA-1 protein (Omichinski *et al.*, 1993a). The methylation interference patterns for finger II are identical to that of full length GATA-1 (Omichinski *et al.*, 1993a).

Finger I may contribute to the discrimination between high- and low-affinity sites and play a role in specificity and stability of binding of both chicken (Yang and Evans, 1992) and mouse GATA-1 (Martin and Orkin, 1990). A finger II protein-

DNA complex shows a similar half-life regardless of the DNA probe, while the wild-type protein has a 1.7 -fold greater half-life for a higher-affinity α^D-globin site relative to a π-globin-derived probe. When GATA-1 is bound to appropriate double GATA-binding sites, DNA methylation interference assays reveal that the finger I protects one of the binding motifs but has little effect on the dissociation constant (Martin and Orkin, 1990; Tsai *et al.*, 1991).

GATA-1 binds DNA specifically as a zinc or iron complex (Yang and Evans, 1992; Omichinski *et al.*, 1993a). Zinc binds at a molar ratio of 1 to the carboxyl finger (Omichinski *et al.*, 1993a). The binding of DNA is dependent on the cysteine residues at the base of the fingers, which complex the zinc (Yang and Evans, 1992). The 3' C-X-X-C site of finger II makes critical contacts on the bottom strand of a GATA binding site which are postulated to be required for GATA-1 to activate transcription (Yang and Evans, 1992).

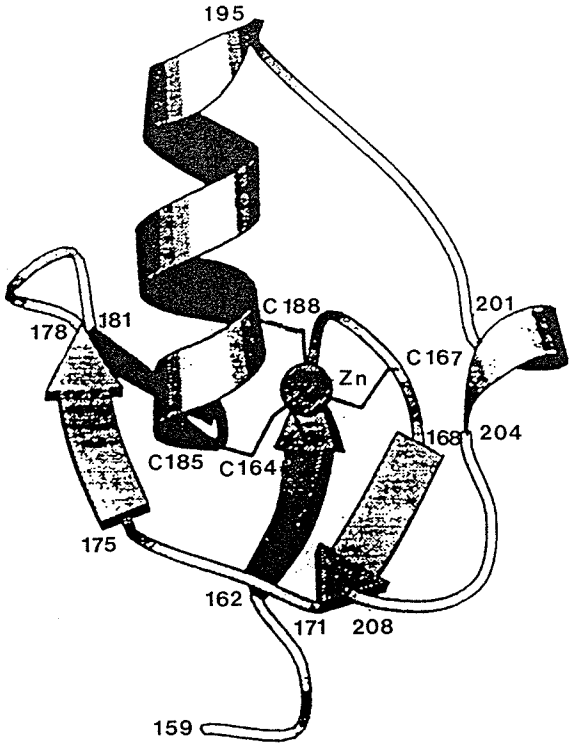
Omichinski *et al.* (1993b) used NMR to study the configuration of a polypeptide encoding finger II of GATA-1. Their data showed that the DNA bases pairs from -1 (sequence $^{-1}$A GAT $^{+1}$AACC^{+4}) to +4 were in contact with amino acids 162 to 216 (including finger II) of GATA-1 (Figure 5A). Residues 159 to 162 form a turn, followed by two short irregular antiparallel ß sheets, a helix (residues 185 to 195) and a long loop (residues 196 to 208; Figure 5B). ß strands 1 (residues 162 to 164) and 2 (residues 168 to 171) form the first ß sheet, and ß strands 3 (residues 175 to 178) and 4 (residues 181 to 184) form the second ß sheet. The core of finger II is dominated by the zinc which interacts with the S_{gamma} atoms of Cys^{164}, Cys^{167}, Cys^{185} and Cys^{188}. Cys^{164} is located in ß strand 1, Cys^{167} in the turn between ß strands 1 and 2 and Cys^{185} and Cys^{188} in the helix. Hence

Figure 5. Chicken GATA-1 finger II: Structure and interaction with DNA.

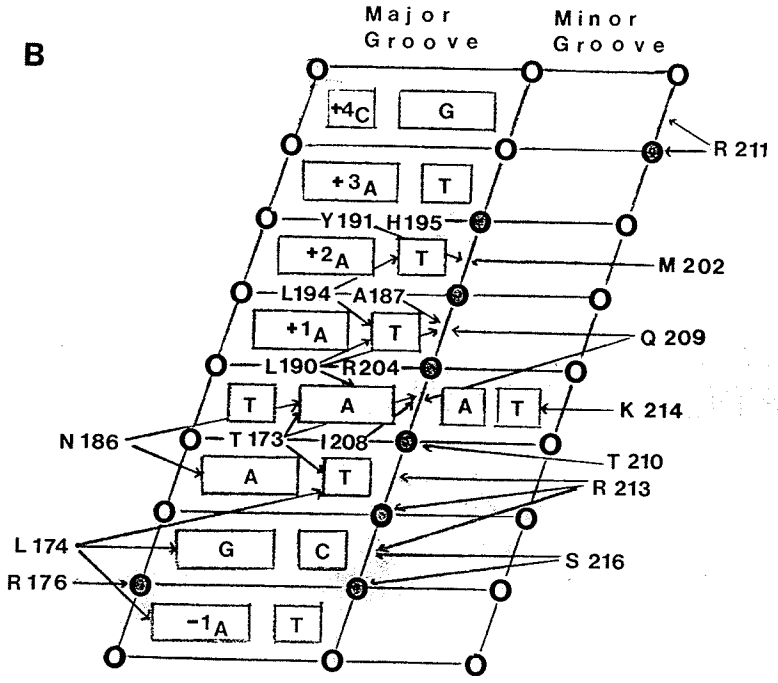
A. Schematic ribbon drawing of the core of the chicken GATA-1 DNA binding domain (residues 162 to 216). The description of the protein fragment is given in the text.

B. Sketch summarizing the contacts between the chicken GATA-1 DNA binding domain and DNA with the DNA represented as a cylindrical projection. Bases interacting with the protein are shaded; phosphates are represented as circles: filled circles indicate sites of interactions with sugar and phosphate or both in the major and minor grooves. Reproduced from Omichinski *et al.* (1993b). The numbering of the nucleotides of Omichinski *et al.* (1993b) were changed to correspond with the nucleotide numbering used throughout this project and are relative to the central GAT sequence. The amino acid residue numbering from Omichinski *et al.* (1993b) were changed to reflect the position of the amino acid residues within the full length GATA-1 protein.

A


195
178
181
C 188
201
C 167
Zn
168
204
C185
C164
175
162
171
208
159

B


Major Groove
Minor Groove
+4C
G
R 211
+3A
T
Y 191
H195
+2A
T
M 202
L194
A187
+1A
T
Q 209
L190
R204
T
A
A
T
K 214
N 186
T 173
I 208
T 210
A
T
R 213
L 174
G
C
S 216
R 176
-1A
T

the zinc is responsible for determining the orientation of the helix with respect to the first antiparallel ß sheet.

The DNA base pairs contacted by finger II were from -1 (sequence $^{-1}$A GAT $^{+1}$AACC^{+4}, where the superscript designates the position of the nucleotide relative to the core GAT sequence) to +4 (Figure 5B). The cGATA-1 DNA binding domain makes specific contacts with eight bases, seven in the major groove (A at -1 of the core GATA sequence; G and A of the core GAT sequence; T and A of the complementary strand within the core ATC sequence; and at the Ts complementary to position +1 and +2) and one in the minor groove (the T of the core GAT sequence). These data are in agreement with methylation interference experiments of the major and minor grooves which have implicated every position of the GATA-1 binding site from base pairs -2 to +2 (Martin and Orkin, 1990; Yang and Evans, 1992; Omichinski *et al.*, 1993a). The bases which are contacted in the major groove are through the helix and the loop connecting ß strands 2 and 3. The remaining contacts involve the sugar-phosphate backbone, the majority of which are located on the complementary strand to the GATA sequence. Salt bridges or hydrogen bonds or both are made by Arg^{177}, Arg^{205} and His^{196} with the phosphates of the G of the core GAT sequence, the A of the reverse ATC and the T on the opposite strand at +2 in the major groove. Arg^{209}, Thr^{208}, Arg^{211} and Arg^{214} contact the phosphates of the C at +4, the T and C of the reverse ATC sequence and the T on the reverse strand at the -1 position in the minor groove. The interactions of Arg^{209} and Arg^{211} above and below the polypeptide chain span the full length of the target site and are probably responsible for the bending of the DNA in the direction of the minor groove.

The helix and the loop connecting strands ß2 and ß3 (which is located directly beneath the helix) are located in the major groove of the DNA, and the carboxyl-terminal tails wrap around the DNA and lie in the minor groove, directly opposite the helix (Figure 5A). Omichinski *et al.* (1993b) suggest the structure of GATA-1 bound to DNA is analogous to a right hand holding a rope, with the rope representing the DNA, the palm and fingers of the hand the core of the protein, and the thumb the carboxyl-terminal tail. This results in a pincer-like configuration of the protein that causes a kink in the DNA in the direction of the minor groove in the DNA between base pairs +1 and +2.

Ko and Engel (1993) used random DNA sequences in EMSAs with GATA-1 and then PCR amplified those sequences which were shifted by GATA-1 in order to determine the sequences with which GATA-1 preferred to interact. It was found that at positions -3 and -2 there was no real nucleotide preference. At position -1, A occurred at a high frequency, but T and C were also allowed, whereas G appeared to be prohibited. The presence of a T at this position resulted in a high affinity binding site for GATA-1. The GAT sequence of the core was essential as determined by methylation interference studies (Wall *et al.*, 1988) with G being the most critical contact point (Schwartzbauer *et al.*, 1992). There was a strong preference for A in both position +1 and +2. The +2 position differed from the canonical sequence in that a G was allowed at this position in the canonical sequence whereas this was not seen in the PCR amplified sequences. There was no apparent preference at the +3 position, but this site did appear to depend on adjacent sequences. An A was preferred at the +4 position. They also found that the sequence of adjacent nucleotides often affected the affinity of GATA-1 for the

sequence. For example by chi-square analysis, the dependence of the identity of the +2 position upon the +1 identity exceeds the 0.005 level of significance when the two sites contain adenines (Ko and Engel, 1993).

Activation of transcription depends on DNA-binding and three activation domains in cGATA-1 which act synergistically to affect transcription (Yang and Evans, 1992). The first domain (act I) appears to be located within amino acids 1 to 71 but is difficult to map precisely because of protein instability; the second domain (act II) maps near or within finger I and seems to be critical (Yang and Evans, 1992; Figure 4). Deleting amino acids 227 to 280 reduces activity to 31% of wild-type and is the third activation domain (act III). Three qualitatively different activation domains were also mapped by structure-function studies on the mouse GATA-1 homolog by fusion of N- and C-terminal regions to a heterologous DNA-binding domains (Martin and Orkin, 1990). The N-terminal mouse domain (an acidic region) confers transcriptional activation in this assay, whereas the C-terminal portion does not. It should be noted that in these activation assays the constructs for GATA-binding did not include binding sites for other erythroid factors which may interact with GATA-1 (see below; Walters and Martin, 1992). Thus there may be as yet unmapped domains of GATA-1 which are important for interactions with other factors. Because finger I is well conserved among GATA family members, act II is a good candidate for a general GATA protein activation domain, and perhaps interacts with specific cellular proteins. This may account for the observation that avian or mammalian GATA-1 proteins both function when expressed in heterologous cell types (i.e. cGATA expressed in human HeLa cells or hGATA-1 expressed in primary chick fibroblasts), despite the lack of sequence conservation

between the factors outside the presumptive DNA-binding domain (Yang and Evans, 1992). Alternatively, a system of poorly conserved repeats among the first 50 and the last 50 residues of chicken, mouse and human GATA-1 proteins might be important for interacting with proteins of the transcription apparatus (Trainor *et al.*, 1990).

The activation domains of GATA-1 which do not contain obvious protypic domains such as an acidic domain may act to properly position the protein to interact with the DNA sequence. Walters and Martin (1992) suggest that subtle changes in the way GATA-1 interacts with its cognate site may have a dramatic affect on how other regions of the factor make protein-protein contacts. As an example, a mutant GATA-1 binding site leads to a very subtle alteration in the DNA-protein interaction but results in inappropriately high levels of globin gene expression (Martin *et al.*, 1989). This mutation results in a promoter, that is 4-5 times more active than the wild-type promoter when tested in erythroid cell (Martin *et al.*, 1989).

GATA-Binding Sites

Binding sites for the transcription factor GATA-1 formerly known as Eryf1, NF-E1 or GF-1 in human erythroid cell lines were first identified in the promoters of the alpha^{D}- and rho-globin genes (Kemper *et al.*, 1987; Evans *et al.*, 1988, Wall *et al.*, 1988, Martin *et al.*, 1989). Evans *et al.* (1988) reported that all chicken globin gene promoters contained a GATA-sequence conforming to the consensus [T/A(GATA)A/G]. There are no known erythroid-expressed genes that have been shown to be independent of GATA-motifs within their promoters, enhancers or

LCRs (Tsai *et al.*, 1991; Grosveld *et al.*, 1987; Orkin, 1990; Talbot and Grosveld, 1991; Talbot *et al.*, 1990; Strauss and Orkin, 1992; Strauss *et al.*, 1992; Jarman *et al.*, 1991; Philipsen *et al.*, 1990). This includes globin genes, heme-biosynthetic genes and all other genes which are expressed in erythrocytes, including the enhancer of the erythroid-specific chicken H5 gene (Trainor *et al.*, 1987; Rousseau *et al.*, 1989). The species examined are human, mouse and chicken. There are three GATA-binding proteins expressed in erythroid tissue which have been cloned, GATA-1, -2, and -3 respectively. GATA-2 and GATA-3 are strongly conserved among avians, amphibians, and mammals, whereas GATA-1 is not (Yamamoto *et al.*, 1990; Zon *et al.*, 1991). Another GATA-binding protein, F6 is a nuclear matrix protein and may mediate an interaction between the nuclear matrix and certain erythroid genes (Vassetzky *et al.*, 1993).

Chimeric mice derived in part from embryonic stem cells with a mutant GATA-1 gene, require GATA-1 expression for generation of a fully differentiated erythroid lineage (Pevny *et al.*, 1991) and *in vitro* for differentiation into erythroid cells (Simon *et al.*, 1992). Therefore, GATA-1 may be the primary determinant protein of erythroid lineage. The GATA-binding factors play a conserved role in erythropoiesis. The changes in the relative abundance and the expression pattern of all three GATA factors during erythroid differentiation are very similar across species (Yamamoto *et al.*, 1990). Therefore, it is expected that each of the GATA factors, which vary in binding specificity and in transactivation potential, play a specific role at the time that they are expressed. GATA-1 expression is high throughout erythroid differentiation and increases slightly as the cell matures; GATA-2 is down-regulated shortly after the induction of differentiation, during

which time GATA-3 is induced (Yamamoto *et al.*, 1990). For example, the glycophorin B promoter is not active in transfected T cells that contain abundant GATA-3, but is highly active in the GATA-1-rich erythroid environment. GATA-1 has greater transactivation potential than GATA-2 or GATA-3 (Martin and Orkin, 1990; Evans and Felsenfeld, 1991). Therefore, it is possible that the GATA proteins bind to the same target sequences (or a common subset) but due to different activation potentials have different effects within cells. For example, there is evidence that GATA-2 maintains erythroid progenitors in an immature state. Expression and activity of a GATA-2/estrogen receptor chimera inhibits terminal erythroid differentiation and concurrently promotes proliferation of immature erythroblasts in an estrogen dependent manner (Briegel *et al.*, 1993). GATA-2 is down-regulated shortly after induction of differentiation in ts-oncogene-transformed erythroblasts. The similar down-regulation of GATA-2 *in vivo* might be a key event required for terminal differentiation of erythrocytes. Briegel *et al.* (1993) suggest that the upregulation of GATA-3 during late maturation may be important for globin gene switching. cGATA-3 is not expressed in embryonic (primitive) or immature definitive erythroid cells (Yamamoto *et al.*, 1990).

Selected amino acid differences are characteristic of the different GATA proteins, such that the fingers of GATA-1 can be distinguished from GATA-2 and 3, and similarly GATA-2 and GATA-3 can be distinguished from each other and from GATA-1 (Zon *et al.*, 1991). Analysis of the DNA-binding characteristics of the GATA factors demonstrates that each factor binds to consensus (WGATAR) sites with similar affinities, but chicken GATA-2 and GATA-3 are also capable of binding to an alternative consensus sequence (AGATCTTA) with high affinity (Ko and

Engel, 1993). Therefore the GATA factors may bind to different transcriptional regulatory sequences, whose distinct target genes then elicit separate effects on the physiological program of erythroid cells, such as promoting self-renewal (GATA-2) or terminal differentiation (GATA-3).

It is thought that the complexity in erythroid control elements correlates to the function of the gene product. For example, genes that are required early in erythropoiesis have promoters which can be directly *trans*-activated by GATA-1 in non-erythroid cells (Zon *et al.*, 1991b). Examples are the GATA-1 promoter itself, the erythropoietin receptor gene and the alpha^{A} globin gene which is expressed throughout erythroid development and has a simple enhancer which appears to only contain GATA binding sites (Chiba *et al.*, 1991; Evans and Felsenfeld, 1991; Hannon *et al.*, 1991; Zon *et al.*, 1991b; Knezetic and Felsenfeld, 1989). In other cases, GATA-1 does not function in heterologous systems to transactivate erythroid-specific promoters, implying the importance of other factors. The ß-globin family, which goes through a series of gene switching, contain in addition to GATA-binding motifs, sites for the binding of other factors.

There is, as well, complexity in GATA binding sites with respect to the sequence of the site, whether the site occurs as a single or double copy and in the case of double copy sites, the orientation with respect to each other and distance apart including overlapping sites (Tsai *et al.*, 1991; Martin and Orkin, 1990; Evans and Felsenfeld, 1991). This complexity affects the strength of binding and the transcription efficiency. The human gamma-globin (Martin *et al.*, 1989; Martin and Orkin, 1990) contains a slightly higher-affinity site consisting of two copies of the GATA-binding sequence, both contacted by a single molecule of GATA-1. The

GATA motif occurs as isolated single copies in number of globin enhancers and promoters. The $ß^A/\epsilon$-globin enhancer contains two adjacent GATA binding sites in opposite orientation which binds two molecules of GATA-1 noncooperatively (Evans *et al.*, 1988). The GATA-1 binding site in α^D promoter of chicken, which contains a double GATA motif, binds one molecule of GATA-1 that recognizes primarily the downstream GATA motif. The presence of the first motif however causes GATA-1 to bind with a two-fold higher affinity, relative to the second motif alone and this facilitates a four-fold higher level of *trans*-activation (Evans and Felsenfeld, 1991). GATA-1 binding site clusters, such as in the chicken GATA-1 promoter which has 3 sites and may bind 2-3 molecules of GATA-1, can be strong positive *cis*-acting elements (Hannon *et al.*, 1991; Evans *et al.*, 1988; Martin *et al.*, 1989; Evans and Felsenfeld, 1991).

GATA-1 may serve to mediate an interaction between the 3'-enhancer and a noncanonical TATA-box (GATAAA) in certain erythroid genes (Fong and Emerson, 1992). It is thought that at these motifs GATA-1 binds first and establishes an interaction between the 3' enhancer and the promoter. GATA-1 is subsequently displaced by TFIID in conjunction with adaptor proteins to form an active initiation complex. Thus a critical step in the tissue-specific regulation of the ß-globin gene is the establishment of enhancer-promoter interaction mediated, in part, by cGATA-1 bound at the -30 position. Fong and Emerson (1992) propose that interaction of cGATA-1 with the distal 3' enhancer and the specialized TATA box confers erythroid specificity to the initiation complex by mediating promoter-enhancer communication. Proximal promoters of several erythroid-expressed genes (erythropoietin receptor, pyruvate kinase, and glycophorin B) also contain a GATA

motif in lieu of a traditional TATA box (Orkin, 1990). In contrast to more complex promoters (such as those of the adult ß-globin genes), they are directly *trans*-activated in heterologous cells by coexpression of GATA-1.

CACCC motif/GC box

The CACCC and GC motifs are grouped together because most factors that bind to one motif also bind to the other, albeit more weakly. Sp1, NEF-2 and other erythroid factors bind these elements. Which if any regulates erythroid genes however is unknown (Gumucio *et al.*, 1991; Davidson *et al.*, 1988; Emerson *et al.*, 1989). Sp1 is a ubiquitous transcription factor that is modified by O-glycosylation and phosphorylation (Jackson and Tjian, 1988). The glycosylation enhances the ability of Sp1 to activate transcription and Sp1 is phosphorylated after binding DNA by a DNA dependent kinase (Jackson *et al.*, 1990). Transcriptional activation by Sp1 depends on a three zinc finger structure which is responsible for DNA binding (Kadonaga *et al.*, 1987). There are four activation domains; 2 glutamine rich domains (domains A + B), a weakly basic domain (C), all located amino-terminal to the DNA-binding domain and the last domain (D) which has no outstanding features is located at the carboxy terminus (Pascal and Tjian, 1991). Sp1 is able to activate transcription synergistically when more than one DNA-binding site is present (Anderson and Freytag, 1991; Pascal and Tjian, 1991). Synergism requires the glutamine-rich domains A and B and domain D. Sp1 forms a tetramer which is subsequently assembled into multiple tetramers and Sp1 proteins bound at DNA sites separated by 1.8 kb will interact, resulting in the looping out of the intervening sequences in a manner dependent on at least one glutamine rich domain (Su *et al.*,

1991; Mastrangelo *et al.*, 1991; Pascal and Tjian, 1991). This information may be important since it would provide a mechanism whereby Sp1 bound at distant sites could bring the sites into juxtaposition and allow proteins bound at the sites to interact, and if they are in contact with a promoter, to affect transcription. Human Sp1 increases the number of productive transcription complexes formed *in vitro* on the adenovirus E1b promoter without apparently affecting TFIID binding or promoting transcription reinitiations (Schmidt *et al.*, 1989a). Sp1 action probably prevents diversion of initiation complexes into non-productive complexes as discussed above.

Sp1 binds to the sequence (G/T)(G/A)GGC(G/T)(G/A)(G/A)(G/T) (Faisst and Meyer, 1992). Sp1 also binds to sequences where the central C is changed to T or A, although these changes reduce the affinity of the sequences for Sp1 binding (Letovsky and Dynan, 1989; Lobanenkov *et al.*, 1990; Yu *et al.*, 1991). CACCC factors also bind the Sp1 consensus sequence (Jackson *et al.*, 1989). CACCC sequences are present in the human 3′ ß-globin enhancer (Giglioni *et al.*, 1989) and chicken $ß^A/\epsilon$ LCR/enhancer (Emerson *et al.*, 1987) and are important for ß-globin expression (Lin *et al.*, 1987; Myers *et al.*, 1986). Binding sites for Sp1, besides the CACCC element have been identified in the α-globin promoters of chicken (Kemper *et al.*, 1987), mouse (Barnhart *et al.*, 1988), human (Whitelaw *et al.*, 1989) and in the chicken $ß^A$ promoter which has been shown to bind human or avian Sp1 (Evans *et al.*, 1988; Lewis *et al.*, 1988). The $ß^A$ promoter site is critical for transcriptional activity (Jackson *et al.*, 1989; Emerson *et al.*, 1989). Sp1 binds to the human zeta promoter, but may be displaced *in vitro* by erythroid-specific factors that bind to overlapping sites (Yu *et al.*, 1990).

Sp1 also binds the stage selector element (SSE) found in the chicken β^A-globin promoter required in definitive erythroid cells for both the expression of the adult gene and the suppression of a linked embryonic ϵ gene (Choi and Engel, 1988; Lewis *et al.*, 1988; Jackson *et al.*, 1989; Emerson *et al.*, 1989). A 65 kDa protein (NF-E4), also binds the purine-rich RAGAGGRGG motif (R = A or G) and may be the relevant protein *in vivo*, since it is present in mature adult erythrocytes, but not in primitive erythrocytes that express the ϵ gene (Gallarda *et al.*, 1989). The mode of action of NF-E4 may be to mediate a physical interaction between the adult-stage specific promoter and the LCR/enhancer located between the epsilon and ß-globin genes (Nickol and Felsenfeld, 1988; Choi and Engel, 1988). Thus there is a competition for the shared enhancer by the promoters of each of the genes and after the NF-E4 protein begins to be expressed in maturing erythrocytes the adult ß-globin promoter is able to "win" the competition as the chicken matures. In humans a similar mechanism may be present in which an embryonic factor performs the role of NF-E4 (Jane *et al.*, 1992).

CCAAT Motif/ NF1 [(T/C) GG (A/C) N_{5-6} GCCAA

There are at least five distinct CCAAT binding factors (CBFs) or families of factors three of which have been implicated in erythroid maturation that recognize the CCAAT motif, a region commonly protected in DNase I protection assays of erythroid genes (Chodosh *et al.*, 1988; Dorn *et al.*, 1987a; 1987b; Superti-Furga *et al.*, 1988; deBoer *et al.*, 1988; Li *et al.*, 1992; Santoro *et al.*, 1988). The factors are distinguishable by sequences peripheral to the conserved CCAAT core, the strength of binding and by homology among proteins. The ubiquitous factor CBF

(NF-Y/CP1) is composed of 2 subunits, which are highly conserved in vertebrates (Li *et al.*, 1992) and has the consensus sequence (C/T)NNNNNN(A/G)(A/G)CCAATCAN(C/T(G/T) (Chodosh *et al.*, 1988). CBF appears to be the factor that interacts with highest affinity to CCAAT elements in the promoters of the human α, ß, and gamma genes and mouse ß-globin (Chodosh *et al.*, 1988; deBoer *et al.*, 1988; Gumucio *et al.*, 1988; Superti-Furga *et al.*, 1988). A mutation that increases the similarity of a CCAAT sequence to the consensus binding site for CP1 (Chodosh *et al.*, 1988) creates a stronger binding site for the human gamma-globin promoter (Gumucio *et al.*, 1988), which correlates with increased expression of this gene. CP2 interacts preferentially with the human zeta-globin promoter (Yu *et al.* 1990). The CP2 consensus sequence is (C/T)AG(C/T)NNN(A/G)RCCAATCNNNR and the protein appears to be a heterodimer (Chodosh *et al.*, 1988).

Nuclear factor 1 (NF1), also known as CTF is a family of proteins that bind to the sequence (T/C)GG(A/C)N_{5-6}GCCAA and some of the proteins also bind the CCAAT motif (Kawamura *et al.*, 1993). Four NF1 genes in chicken have been identified; NF1-A, -B, -C, -X (Rupp *et al.*, 1990; Kruse *et al.*, 1991), which code for proteins of 46 to 62 kDa. Differential splicing and protein modification (O-glycosylation) also contribute to the heterogeneity in the NF1 family (Rupp *et al.*, 1990). A mouse NF-1 factor that recognizes primarily the NF-1-binding sequence and not the CCAAT box is phosphorylated *in vitro* by cdc2 kinase, although the significance of this result was not determined (Kawamura *et al.*, 1993). Human CTF-2, which is homologous to chicken NF1-C2 (Rupp *et al.*, 1990) regulates transcription via a proline-rich activation domain (Mermod *et al.*, 1989). Cell

specific differences in the amount and forms of NF1 have been observed (Goyal *et al.*, 1990).

CTF/NF1 from HeLa cells has been shown to bind to a CCAAT box and stimulate transcription from the human α promoter (Jones *et al.*, 1987). CTF/NF1 related proteins are capable of inhibiting binding of positive transcription factors to an adjacent site in the promoters of the chicken β^H- (Plumb *et al.*, 1986) and the chicken β^A-globin (Jackson *et al.*, 1989). PAL, the chicken β^A inhibitor is a mature erythrocyte protein, the appearance of which correlates with the shut off of globin expression in mature erythroblasts (Jackson *et al.*, 1989). This protein binds in an anti-cooperative manner with a positively acting CACCC factor. A CTF/NF1 activity related to the CCAAT displacement protein (CDP) of sea urchins (Barberis *et al.*, 1987) binds within the human ß globin genes (deBoer *et al.*, 1988). It also binds to the duplicated CCAAT elements of the human gamma promoter and inhibits binding of CP1 (Superti-Furga *et al.*, 1988). A similar situation may exist in the mouse α promoter (Barnhart *et al.*, 1988).

cIBR

cIBR (chicken initiation binding repressor) is a 75 kDa DNA binding glycoprotein from mature chicken erythrocytes. It protects the sequence -10 to +19 of the transcription start site of the H5 gene in DNase I protection assays (Gomez-Cuadrado *et al.*, 1992). cIBR represses transcription from the H5 promoter *in vitro* by interfering with assembly of the initiation complex. Human TFIID bound to the TATA element prevents subsequent binding of cIBR, although cIBR does not affect binding of TFIID. It was found that cIBR did not affect transcription when

bound upstream of the promoter. Based on these data it was suggested that cIBR directly interferes with general transcription factors other than TFIID, possibly TFIIB. Transcriptionally active, immature erythrocytes contain cIBF (chicken initiation binding factor) instead of cIBR. Purified cIBF is a non-glycosylated 68-70 kDa DNA binding protein(s) which protects a slightly larger region (-8 to +23) of the H5 gene in DNase I protection assays. The two proteins however make the same DNA contacts with a shorter region (-6 to +4), as inferred from methylation interference studies.

AP-1 [TGA (G/C) T (C/A) A] /NF-E2 [TGACTCAG]

AP-1/NF-E2 binding sites are present in control elements of a number of erythroid specific genes, such as the chicken ß-globin enhancer, the promoter of PBGD and in the LCRs of alpha and ß globin genes some of which are tandem AP-1 sites (Emerson *et al.*, 1987; Mignotte *et al.*, 1989b; Ney *et al.*, 1990; Philipsen *et al.*, 1990; Disela *et al.*, 1991). AP-1/NF-E2 elements may be composite hormone response elements which may respond to a number of AP-1 transcription factors (Walters and Martin, 1992). There are several AP-1/NF-E2 binding factors present in erythroid tissue, including *fos/jun*, the thyroid hormone receptor c-*erb*A and NF-E2, an erythroid-specific protein with a distribution pattern similar to GATA-1 (Schule *et al.*, 1990; Zenke *et al.*, 1990; Privalsky *et al.*, 1990; Talbot and Grosveld, 1991); Mignotte *et al.*, 1989b). The major component of murine NF-E2 is a 45 kDa protein that belongs to the basic-region-leucine zipper family of transcription factors (Andrews *et al.*, 1993). Mice homozygous for a mutant form of NF-E2 are severely anaemic (Peters *et al.*, 1993). Normal human and murine

erythroblasts express a high level of c-*fos* mRNA throughout their differentiation (Caubet *et al.*, 1989), and it is suggested the *Fos* protein has a role in the regulation of the transcription of genes switched on during the terminal erythroid differentiation (Sambucetti and Curran, 1986; Rauscher *et al.*, 1988a; 1988b; Franza *et al.*, 1988).

V-*erb*A regulates the erythroid-specific proteins; anion transporter band 3, carbonic anhydrase II, and the enzyme aminolevulinate synthase through AP-1 sites (Disela *et al.*, 1991). C-*erb*A blocks differentiation in a hormone-dependent manner (Zenke *et al.*, 1990). The promoter of the chicken GATA-1 has a potential binding site for members of the steroid hormone-receptors superfamily. This region is protected in DNase I protection assays when non-erythroid (but not erythroid) extracts are used (Hannon *et al.*, 1991). Therefore, it may represent a negative regulator of GATA-1 gene activation.

Synergy Between Erythroid Transcription Factors

There is evidence for synergy among some erythroid factors, which may reflect a physical interaction between factors. In erythroid tissue it has been noted that ß-globin promoters often have a TATA box at about -30, a CCAAT box near -70 to -90 and a CACCC motif often from -95 to -120 (McDonagh *et al.*, 1991; Lin *et al.*, 1987; Myers *et al.*, 1986; Yu *et al.*, 1991). Gong *et al.* (1991) noted that in addition to these factors there is a conserved GATA motif at about -165 in mammalian epsilon globin genes. In the human epsilon-globin promoter, this site is required for interaction with an enhancer and in the absence of the enhancer, the site does not affect transcription (Gong *et al.*, 1991). Further research showed that

the binding of an AP-1-like factor to the enhancer was required for enhancer activity (Gong and Dean, 1993). Gong and Dean (1993) suggest a multimeric interaction may be responsible for this effect.

Additionally, it has been noted that CACCC elements are found close to GATA motifs in erythroid elements (Collis *et al.*, 1990). Activity of the porphobilinogen deaminase gene (PBGD) promoter is lowered when the spacing between GATA-1 and CACCC elements is increased (Frampton *et al.*, 1990). The HPFH-associated changes in GATA-1 activity (discussed above), are expressed in nonerythroid cells only when the site is linked to either a CACCC or a AP-1/NF-E2 element (Martin and Orkin, 1990). In transient transfection assays none of the elements GATA, CACCC, and AP-1/NF-E2 alone created an active erythroid promoter, whereas combinations of two elements created active promoters in mouse and human erythroid leukemia cell lines (Walters and Martin, 1992). Therefore GATA-1 motifs linked to either CACCC or AP-1/NF-E2 elements form strong, inducible erythroid promoters. This result suggests that GATA-1 requires erythroid-specific factors in order to function in erythroid cells. These data imply that globin gene regulation is achieved by the combined action of several transcription factors (Walters and Martin, 1992). Thus GATA-1 may interact with factors binding to the CACCC and/or AP-1/NF-E2 elements.

Minie *et al.* (1992a; 1992b) suggest that a combination of the changing concentrations of transcription factors together with the variability in binding affinities of Sp1 and GATA-1 to control sites of individual erythroid-specific genes is important for differential regulation of these genes. In the developing chicken embryo, the concentrations of GATA-1 and Sp1 decrease by almost an order of

magnitude as the primitive cells are replaced by definitive cells at day 5 (Minie *et al.*, 1992a; 1992b). Therefore the number of doubly occupied sites in the rho promoter which contains a GATA and an Sp1-binding site are estimated to decrease by two orders of magnitude (Minie *et al.*, 1992a; 1992b). Thus regulation of rho-globin expression can be controlled by factors that are not confined to a particular lineage. A similar concentration-dependent mechanism (involving additional factors) appears to account for lineage-specific expression of the alpha^{pi}-globin gene (Minie *et al.*, 1992a; 1992b).

Introduction to Histone H5

Histone H5 is a variant of the H1 family (Aviles *et al.*, 1979; Neelin *et al.*, 1964) found in nucleated erythrocytes of birds (Neelin *et al.*, 1964; Shannon *et al.*, 1985). H5 shows structural (Aviles *et al.*, 1979) and sequence (Yaguchi *et al.*, 1977) similarities to histone H1. The expression pattern of histone H5 is similar to that of the adult ß-globin gene in that both are expressed at higher levels in erythrocytes than in early erythroblasts (Beug *et al.*, 1982; Neelin, 1968; Sotirov and Johns, 1972). However, histone H5 is transcribed in avian erythroblastosis virus-transformed CFU-E erythroid cells, at an earlier stage than the ß-globin gene (Beug *et al.*, 1979; Zenke *et al.*, 1988). Before the maturing cells lose their proliferative potential, transcription of the H5 gene increases by up to a factor of 6. As the cells enter the terminal G1 phase, the rate of H5 gene transcription remains high. Although the transcription rates of H5 and ß^{A}-globin genes are similar, the level of ß^{A}-globin mRNA increases 30-170-fold over that of H5 mRNA as the cells reach the erythroblast stage. Moreover, while the level of ß^{A}-globin mRNA is

maintained, that of H5 mRNA drops during maturation, probably due to a lack of mRNA transport to the cytoplasm (Affolter *et al.*, 1987).

Expression of Chicken Histone H5 Gene

Unlike most histone genes which are replication dependent or partially replication dependent, histone H5 belongs to the class of histone genes which code for replication-independent variants. The H5 gene is unique, and the polyadenylated H5 mRNA lacks the terminator element typical of other histone messengers (Krieg *et al.*, 1983). H5 does not have any introns and is unlinked to either H1 or core histones over 41 kilobases (Krieg *et al.*, 1983; Ruiz-Carrillo *et al.*, 1983). The site of polyadenylation is at +768, which is about 880 bases from the putative cap site at -109, in good agreement with the estimate of 900 bases for mature H5 mRNA (Krieg *et al.*, 1983). The H5 gene TATA box has the sequence TTAAAT located at -22 to -17 (Ruiz-Carrillo *et al.*, 1983). The 5' region of the H5 gene does not have a 'CAAT' box which is observed 30 to 60 bases upstream from the TATA box in many eukaryote genes (Benoist *et al.*, 1980).

Both the chicken β^A-globin gene and the histone H5 genes have downstream enhancers contain binding sites for GATA-1. Each of the enhancers also has sequence similarity within a 34 base pair region which is protected in DNase I protection assays in the chicken β^A-globin (DNase I protected region V; Trainor *et al.*, 1987). Both the histone H5 and the ß-globin gene are only expressed in an erythroid environment and therefore would be expected to be under the control of erythroid-specific factors. An important distinction between histone H5 gene expression and ß-globin expression is that the histone gene is expressed earlier than

the ß-globin gene and does not have a pattern of gene switching as does the ß-globin gene. This would suggest that despite similarities between the two genes, the control of transcription is different for the two genes.

DNase I Hypersensitive Sites of Chicken Erythroid Histone H5 Gene Chromatin

Renaud and Ruiz-Carrillo (1986) examined 14.4 kb around the chicken H5 gene for DH sites at several stages of erythroid differentiation. Taking the mRNA cap site (+1) as the reference, the DH sites map at -2190 (DH 1); -1499 to -1411, (DH 2); -1296 to -1216, (DH 3); -840 (DH 4); -117(±6) (DH 5); -25(±6) (DH 5_L); 28(±7) (DH 6); 1003(±12) (DH 7_L); 1095(±12) (DH 7_U) and 2118 (DH 8), the latter being the only site that is not tissue specific (Renaud and Ruiz-Carrillo, 1986). This thesis is concerned with the region of promoter which includes DH site 5_L and the 3' enhancer which includes DH sites 7_L and 7_U. In immature erythrocytes, micrococcal nuclease (MNase) cleaves at -111(±6) (DH 5), 1016(±12) (DH 7_L) and 1211(±12) (DH 7_U). DH site 7 contains stretches of relatively unprotected DNA interspersed with more resistant structures. The regions sensitive to the exonucleolytic activity of MNase are about 59 bp (DH 7_L) and 66 bp (DH 7_U). Digestion with MNase I resulted in a nuclease-resistant region of 50 to 70 basepairs bordered by DH sites 7_L and 7_U. S_1 nuclease maps DH site 7 to 1026(±12).

A decline in sensitivity of the DH 5_L, which maps close to the TATA box, correlates with a decrease in transcription initiation (Ruiz-Carrillo *et al.*, 1983). DH site 5 also maps near a tissue-specific upstream promoter element (Rousseau *et al.*, 1989; Trainor and Engel, 1989). The histone H5 gene promoter has two positive elements: a GC-box (-83 to -75) and an upstream activating sequence (UPE, -51

to -33) (Rousseau *et al.*, 1989). Deletion of either of these elements reduces the strength of the promoter (Rousseau *et al.*, 1989; Trainor and Engel, 1989; Shannon *et al.*, 1985; Wigley and Wells, 1987; Trainor *et al.*, 1987). An upstream negative element mapped to -95 to -115 (Rousseau *et al.*, 1989).

DH sites 7_U and 7_L map with a 3′ erythroid-specific enhancer element which is located in the region +851 to +1185 (Trainor *et al.*, 1987; Rousseau *et al.*, 1989). In transient transfection studies using H5 expressing cells, the H5 3′ enhancer requires the presence of H5 promoter elements, including the GC box and most of the UPE (-38 to -90) to activate transcription (Rousseau *et al.*, 1989). This suggests that enhancer-bound proteins interact directly or indirectly (by way of coactivators or adaptors) with proteins associated with promoter elements that are localized between -38 to -90. The 3′ enhancer is less active in immature than mature erythroid cells (as represented by line HD6 or undifferentiated HD3) (Trainor *et al.*, 1987).

Significance of *In Vitro* Interactions Between Factors of Chicken Erythrocyte Nuclei and DNA Elements of the Histone H5 Gene

Knowledge of *in vitro* interactions between nuclear proteins of chicken erythroid cells and gene sequences of the chicken histone H5 gene should be instructive for understanding *in vivo* interactions and functions of factors binding to the histone H5 gene chromatin. Therefore, this project studied *in vitro* interactions of nuclear extracts of chicken erythroid cells with DNA elements of the chicken histone H5 gene using a number of techniques. Based on the data obtained, a looping model is presented for physical protein interactions between factors bound

at the promoter and the 3′ enhancer within the gene chromatin. We propose that this looping occurs *in vivo* and is important for transcriptional activity within the histone H5 gene chromatin. Possible functions of factors bound within the looping model are suggested as they relate to transcriptional activity of the gene.

Materials and Methods

Materials

Amersham

- deoxyadenosine 5′-α [^{35}S] thiotriphosphate
- α-[^{32}P] deoxyadenosine triphosphate
- α-[^{32}P] deoxyguanine triphosphate
- α-[^{32}P] deoxycytosine triphosphate
- ^{14}C-methylated Protein Standards

Bio-Rad

- Bio-Rad Protein assay
- acrylamide

Boehringer Mannheim

- Quick Spin™ G-25 sephadex columns

British Drug House

- casamino acids

Pharmacia

- calf intestinal alkaline phosphatase
- T4 DNA ligase
- poly dI·dC
- Klenow enzyme
- KpnI
- EcoRI
- HindIII
- HhaI

Promega

RQ1 Deoxyribonuclease I

Sigma

Cyanogen bromide activated Sepharose 4B beads

Deoxyribonuclease I

1-acetyl-2-phenylhydrazine

phenyl-agarose

heparin-agarose

S-sepharose

Q-sepharose

double-stranded DNA cellulose

single-stranded DNA cellulose

United States Biochemical Corporation

Sequenase DNA sequencing kit

MvaI

Worthington Biochemical Corporation

Micrococcal Nuclease

DNA Laboratory (University of Manitoba) and Regional DNA synthesis Laboratory (University of Calgary)

oligonucleotides

Buffers used:

20 x SSC: 3 M NaCl, 0.3 M Na Citrate, pH 7.0 with HCl

TE buffer: 10 mM Tris.HCl, 1 mM EDTA

Reticulocyte Standard Buffer (RSB): 10 mM Tris·HCl pH 7.5, 10 mM NaCl, 3 mM

$MgCl_2$, 10 mM sodium butyrate

TBE: 0.09 M Tris, 0.09 M borate 2 mM EDTA (pH 8.3 without adjustment)

TNE: 5 mM Tris·HCl pH 8, 1 mM EDTA, 15 mM NaCl

Buffer D: 20 mM HEPES pH 7.9 (NaOH), 20% glycerol, 0.1 M KCl, 0.2 mM EDTA, 1 mM dithiothreitol, 1 mM PMSF

Buffer E: 40 mM HEPES pH 7.9 (NaOH), 20% glycerol, 0.2 mM EDTA, 1 mM dithiothreitol, 1 mM PMSF (Sepharose column)

Buffer G: 20 mM HEPES·KOH, 5 mM $MgCl_2$, 0.1 mM EDTA, 1 mM dithiothreitol, 0.1% NP-40, 10% glycerol (double-stranded DNA-cellulose column) [$\pm(NH_4)_2SO_4$]

Buffer H: 20 mM HEPES·KOH, 12.5 mM $MgCl_2$, 1 mM dithiothreitol, 0.1% NP-40, 20% glycerol (single-stranded DNA-cellulose column) ($\pm$KCl)

Equilibration buffer: 20 mM HEPES pH 7.9 (NaOH), 20% glycerol, 5 mM $MgCl_2$, 1 mM dithiothreitol, 0.1 mM EDTA, 25 mM $(NH_4)_2SO_4$, 0.35% Brij-35 (Emerson and Felsenfeld, 1984)

Buffer Z: 100 mM KCl, 20 mM HEPES (pH 7.9), 12.5 mM $MgCl_2$, 1 mM dithiothreitol, 0.1% NP-40, 20 % glycerol (Evans and Felsenfeld, 1989)

Blood Collection buffer: 10 mM Tris·HCl pH 7.5, 75 mM NaCl, 25 mM EDTA

1 X EMSA binding buffer: 10 mM Tris·HCl pH 7.5, 40 mM NaCl, 1 mM EDTA, 1 mM ß-mercaptoethanol, 4 % glycerol (weight/volume)

2 X SDS sample buffer: 30 mM Tris·HCl pH 6.8, 5% glycerol (v/v), 1% SDS, 25 mM ß-mercaptoethanol, 6 X 10^{-4}% bromophenol blue

Protease inhibitor concentrations for column fractions and tissue preparations as specified in the methods are as follows: phenylmethanesulphonyl fluoride (PMSF), 1 mM; iodoacetamide (IAC), 1 μg/ml; aprotinin, 1 μg/ml and leupeptin, 0.5 μg/ml. NP-40 was used at a final concentration of 0.1% unless otherwise stated. Distilled, deionized water was used throughout. All culture media were autoclaved or filter sterilized. All procedures were done on ice, or at 4°C unless otherwise stated.

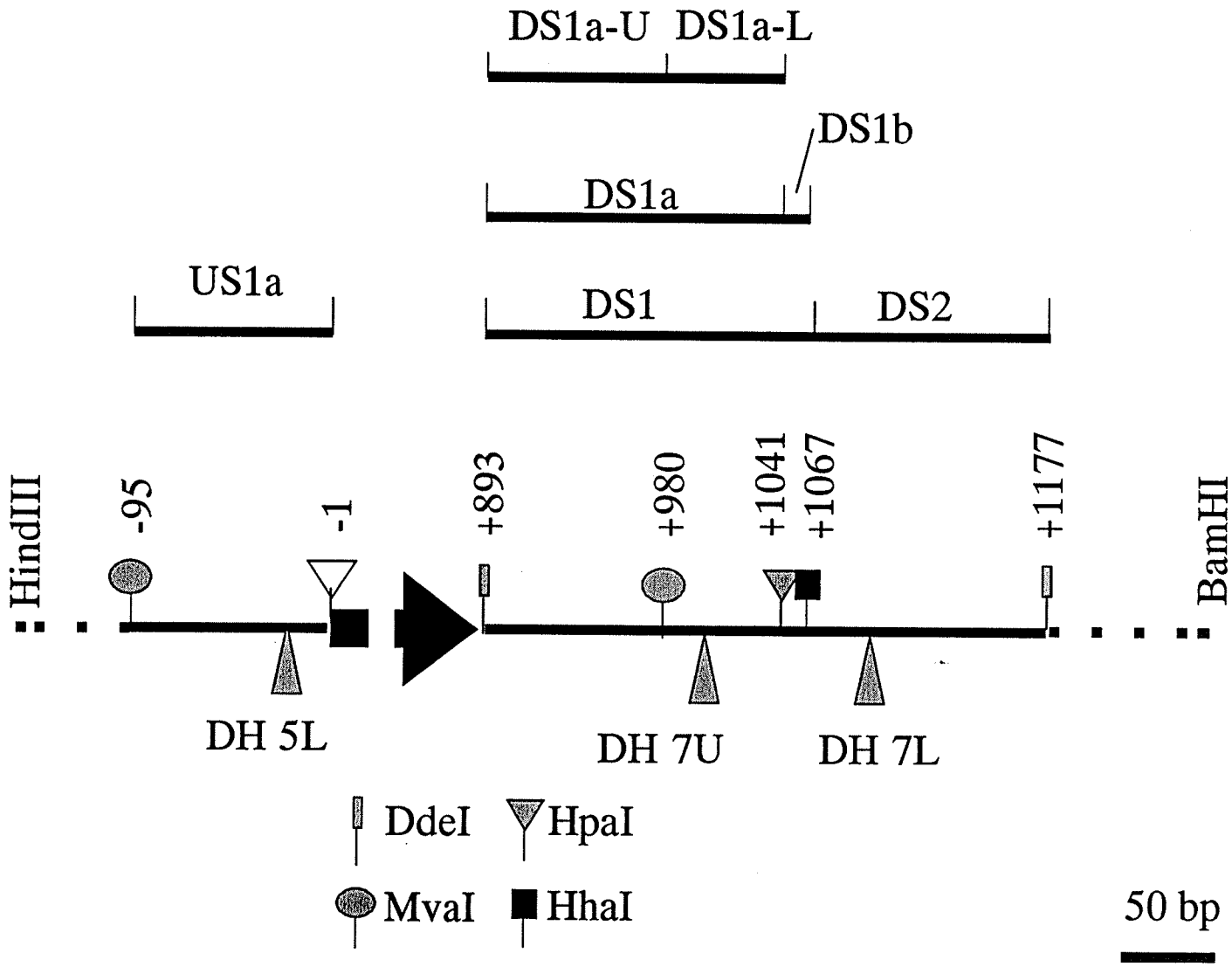
Methods

Most manipulations of DNA were done according to Ausubel *et al.* (1987). This manual was also used for some protein manipulations, including SDS-polyacrylamide gel electrophoresis.

Subcloning and Purification of DNA Fragments

Adolfo Ruiz-Carrillo kindly supplied a plasmid containing the region -777 to +1368 (relative to the start of transcription) of the chicken histone H5 gene. The chicken H5 gene Dde I restriction enzyme fragment from +893 to +1177 (DS1 + DS2) was cloned into the Sma I site in the multiple cloning region of pGEM-3 and was used as the source of DNA for subcloning the DS1 fragment. DS1 extends from the Dde I site at +893 to the HpaII site at 1068. DS1 was subcloned into pGEM-3 and the double-stranded DNA was sequenced. The insert was removed from vector by digesting with appropriate pairs of enzymes such as Xba I/EcoR I, BamH I/EcoR I, Hind III/EcoR I which cut within the multiple cloning region of pGEM-3, one on either side of the insert. Therefore the insert contained parts of

Figure 6. The H5 gene DNA fragments used in this study are indicated. US1a, DS1 and DS1 + DS2 were subcloned into pGEM-3, and the DNA fragments containing US1a, DS1a, DS1b, DS1a-U, DS1a-L and DS2 were excised from the appropriate clone. The arrowheads show the location of the DNase I hypersensitive (DH) sites in the erythroid histone H5 gene chromatin (Renaud and Ruiz-Carrillo, 1986).


DS1a-U
DS1a-L
DS1b
DS1a
US1a
DS1
DS2
HindIII
-95
-1
+893
+980
+1041
+1067
+1177
BamHI
DH 5L
DH 7U
DH 7L
DdeI
HpaI
MvaI
HhaI
50 bp

the pGEM multiple cloning site at either end. Other DNA sequences used were the chicken histone H5 5′ promoter (H5-5′ P) from Sau3A I (-350) to Sau3A I (+212), US1a (-95 to -1) and the ß-globin 3′ enhancer region (ß-globin-3′ E) from Pvu II (+1877) to Pvu II (+2178), each of which was subcloned into pGEM-3. The purified insert was prepared similar to DS1 and therefore also contained pGEM multiple cloning region sequence. DS1a (+893 to +1041) was isolated by digesting the pGEM-3 vector containing DS1 with HindIII and HhaI. DS1a-U (+893 to +980) and DS1a-L (981 to 1041) were isolated from DS1 which was digested with Mva I and Hha I, respectively (Figure 6). Therefore the plasmids described were digested appropriately to produce a sample containing vector and insert.

The samples were electrophoresed to separate insert from plasmid on a 1% agarose gel containing 5 μg/ml of ethidium bromide. When the insert was separated from the plasmid DNA, the band containing the insert was cut out of the gel and eluted electrophoretically as follows. A well was prepared which was lined with dialysis tubing which had been soaked in 1% (w/v) EDTA for 20 minutes, washed several times in water and then left in TBE until needed. The band containing the insert was placed in the well and the well was filled with TBE. The DNA was electroeluted into the TBE after which the TBE together with the DNA was moved to a 13 ml polypropylene tube. The TBE and DNA was extracted with equal volumes of isoamyl alcohol until the sample was reduced to 1 ml. The isoamyl alcohol both removes the aqueous phase and ethidium bromide which would cause strand breakage of the DNA. The DNA was ethanol precipitated and then resuspended in 100 μl TE.

A number of sequences containing GATA-binding sites of erythroid-specific

genes were synthesized. The H5-GATA sequence covers the region from +1045 to +1067 downstream of the start of transcription of the chicken H5 gene and covers the GATA-protein binding site (Rousseau *et al.*, 1989). The complementary sequences of this site are GAGGCTGGAGATAACAGTGCGG (H5-GATA-for) and GCCGCACTGTTATCTCCAGCCT (H5-GATA-rev). The -for or -rev noted in the naming of each strand was used to indicate coding and non-coding strand respectively. ß-globin 3' GATA contains the GATA sequence of the chicken ß-globin enhancer located from +1888 to +1910 downstream of the start of transcription (Evans *et al.*, 1988). The synthesized sequences are CAGGTTGCAGATAAACATTTTG (ß-globin 3' GATA-for) and GCAAAATGTTTATCT-GCAACCT (ß-globin 3' GATA-rev). The oligonucleotides encoding the chicken ß-globin promoter sequence, which has a GATA sequence in place of the TATA box were synthesized (Fong and Emerson, 1992). The sequences used were GGCGGAGGCGATAAAAGTGGGG (ß-globin promoter TATA/GATA-for) and TCCCC-ACTTTTATCGCCTCCGC (ß-globin promoter TATA/GATA-rev). The chicken rho-globin promoter sequence GGACAGCAAGATAAGGGCTGCT (rho-globin promoter-for) and CAGCAGCCCTTATCTTGCTGTC (rho-globin promoter-rev) were synthesized (-200 to -206; Minie *et al.*, 1992a). The mouse α-globin promoter sequence (located at -189 to -174) GGGGCAACTGATAAGGATTCCC (αG2 promoter-for) and TGGGAATCCTTATCAGTTGCCC (αG2 promoter-rev) were synthesized (Plumb *et al.*, 1989). The strands were annealed by mixing equal amounts of each pair of complementary DNAs in 7 X SSE, placing the tube in boiling water and allowing them to cool to room temperature.

Oligonucleotides containing the histone H5 GC-box (-92 to -67), which has

the sequence of a high affinity Sp1 binding site, and the UPE (-54 to -26) were synthesized. The following oligonucleotides that contained the histone H5 Sp1 motifs 5'-GATGCCTGCGGGGCGGGGCAGAGGGG-3' (Sp1-for) and 5'-TCCCCTCT-GCCCCGCCCCGCAGGCAT-3' (Sp1-rev), 5'-GGGTTGCGGTGGGCGGGTGGAGGA-CT-3' (PI-for) 5'-TAGTCCTCCACCCGCCCACCGCAACC-3' (PI-rev), 5'-GGGGGAG-GTAGGAGGGCTGTGAGAGC-3' (PII-for), 5'-TGCTCTCACAGCCCTCCTACCTCCCC-3' (PII-rev), 5'-GGAGGAGAGGGGACTCCTCCTTGTCC-3' (PII-for), 5'-TGGACAA-GGAGGAGTCCCCTC-TCCTC-3' (PIII-rev), 5'-GGGCCATTGGGGTGGGGTTGAACG-CT-3' (PV-for), 5'-TAGCGTTCAACCCCACCCCAATGGCC-3' (PV-rev). The histone H5 UPE-oligonucleotides were 5'-TAGGCAGTÇCTCCCCGCGGTCCGTGCCG-3' (UPE-for) and 5'-GCGGCACGGACCGCGGGGAGGACTGCCT-3' (UPE-rev) were synthesized. The double-stranded competition oligonucleotide Sp1 (5'-GATCGATCGGGGCGGGGCGATC-3') was obtained from the Stratagene HotFoot DNase I Footprinting kit. Oligonucleotides that contained the NF1 motif 5'-TCGAGGGCTTGGCACAGCCCCAAGACCA-3' (NF1-for) and 5'-GTGGTCTTGGGGC-TGTGCCAAGCCCTCG-3' (NF1-rev) were synthesized.

Preparation of [^{32}P]-labelled DNA

End-labelled oligonucleotide and plasmid insert DNA were prepared using 3 - 8 units of Klenow enzyme, 10-50 μCi of the appropriate α-[^{32}P] dNTP, $\pm$ 0.8 mM unlabelled dNTP as required and DNA in a buffer of 50 mM Tris.HCl pH 7.2, 10 mM $MgSO_4$, 0.1 mM dithiothreitol, 50 μg/ml bovine serum albumin in a final reaction volume of 25 μl (Ausubel *et al.*, 1987). Enzyme was added last, and the reaction was at 37°C for 15 minutes. The amount of DNA used for the reaction was 100

ng of insert DNA (e.g. DS1) or 10 ng of oligonucleotide. The reaction was stopped with 2.5 μl of 250 mM EDTA. The sample was diluted to 40 μl with TNE buffer. Unincorporated nucleotides were separated from DNA with a Quick Spin™ column. Incorporated radioactivity was determined by measuring counts per minute in a scintillation counter.

Primer Extension Labelling of Oligonucleotide

Primer extension labelling of oligonucleotides was used to label one strand of a double-stranded oligonucleotide (Ausubel *et al.*, 1987). The probe was prepared with the oligonucleotide H5-GATA-for hybridized to a 8 bp complementary primer (5'-CCGCACTG-3') or by hybridizing the oligonucleotide H5 GATA-rev to a 8 bp complementary primer (5'-AGGCTGGA-3'). The template and primer were hybridized by mixing the two in water, placing the sample in boiling water and allowing the water to cool slowly to room temperature. The DNA was used at 5 ng of oligonucleotide per labelling reaction. The oligonucleotide was made completely double-stranded by incubation with the Klenow polymerase in a reaction mix containing 0.05 mM of the non-radioactive deoxynucleotide triphosphates (dNTPs), 50 μCi of ^{32}P-labelled dNTP, 3-8 units of Klenow enzyme and hybridized template and primer in a buffer of 50 mM Tris·HCl pH 7.5, 10 mM $MgCl_2$, 1 mM dithiothreitol, and 50 μg/ml bovine serum albumin in a total volume of 25 μl. The enzyme was added last, and the reaction was allowed to proceed for 2 - 4 hours at room temperature. The uniformly labelled oligonucleotide was used for many of the EMSAs, since it allowed incorporation of a high amount of ^{32}P and hence produced a probe with high specific activity. In order to increase efficiency of ultraviolet light

cross-linking, UTP was used in place of dTTP (Ausubel *et al*, 1987). The labelling reaction was stopped by adding 2.5 μl of 250 mM EDTA. The mixture was diluted to 40 μl with TNE buffer. Unincorporated nucleotides were separated from DNA with a Quick Spin™ column. Incorporated radioactivity was determined by measuring counts per minute in a scintillation counter.

Preparation of GATA-Affinity Column

The two complementary oligonucleotides used for preparation of the column were 5'-GCTTCCCGGGTTGCAGATAAACATTTAT-3' and 5'-CCCGGGAAGCATAA-ATGTTTATCTGCAA-3'. The annealed fragments form the double-stranded binding site for GATA-binding proteins and is the sequence found within the chicken ß-globin gene region IV DNase I protected region as identified by Evans *et al.* (1988). Evans and Felsenfeld (1989) used this sequence for purifying GATA-1 from mature chicken erythrocytes. The method of Kadonaga (1990) was used for preparation of the DNA. Two batches of 220 ng of each complementary oligonucleotide were placed in separate eppendorf tubes in a buffer of 50 mM Tris·HCl pH 7.6, 10 mM $MgCl_2$, 5 mM dithiothreitol, 0.1 mM spermidine and 0.1 mM EDTA in 75 μl. The mixed oligonucleotides were incubated at 88°C for 2 minutes, 65°C for 10 minutes, 37°C for 10 minutes and at room temperature for 5 minutes to allow hybridization. ATP including 1 μCi of ^{32}P ATP was added to each tube to a final concentration of 3 mM. One hundred units of T4 polynucleotide kinase was added in a final volume of 100 μl. The phosphorylation reaction was allowed to proceed for 2 hours at 37°C. Fifty μl of 10 M ammonium acetate and 100 μl of water was added. The DNA was heated at 65°C for 15 minutes to inactivate the kinase and

the samples were cooled to room temperature. The DNA was ethanol precipitated, redissolved to 225 μl in TE, phenol/chloroform extracted and chloroform extracted. Twentyfive μl of 3 M sodium acetate was added and the DNA was ethanol precipitated again, followed by a 75% ethanol wash. The pellet was dried. Ligation of the oligonucleotides was done in a buffer containing 66 mM Tris·HCl pH 7.6, 10 mM $MgCl_2$, 15 mM dithiothreitol, 1 mM spermidine, 3 mM ATP and 30 Weiss units of T4 DNA ligase. The ligation mixes were incubated overnight at 15°C. Ligation was monitored by determining the size distribution of DNA when the ligated DNA was electrophoresed in a 1% agarose gel. After ligation, the DNA was phenol and phenol/chloroform extracted, ammonium acetate was added to 2 M and the DNA followed by an isopropanol precipitation. The DNA was dissolved in TE with sodium acetate added to 0.3 M and ethanol precipitated. The pellets were washed two times with 750 μl ethanol and then dried. The DNA was dissolved in 400 μl 10 mM potassium phosphate pH 8.2 with pooling of the DNA. One μl of the DNA was removed for determining radiation by scintillation counting.

Preparation of the column was according to Evans and Felsenfeld (1989). Preparation of the column was done in the cold room. The column was made using 1.2 grams of cyanogen bromide activated Sepharose 4B beads. The beads were hydrated in 20 ml of 1 mM HCl, poured into a sintered glass funnel and then washed with 150 ml of 1 mM HCl using a slight vacuum to draw the buffer through the funnel. The next steps were done rapidly. The column matrix was washed with 4 X 15 ml of cold 10 mM potassium phosphate pH 8.2. The Sepharose 4B was transferred to a 15 ml tube in 6 ml of 10 mM potassium phosphate pH 8.2. The ligated DNA was added to the tube, which was then capped and left to rotate

constantly at room temperature for 16 - 18 hours. The reaction was stopped by adding Tris·HCl pH 8 to 0.1 M. Washing of the column was at room temperature. The column was transferred back to the glass funnel and washed in succession with 10 ml of 0.1 M Tris·HCl pH 8, 3 X with 10 ml of 0.1 M potassium phosphate pH 8.2, 3 X with 10 ml of 1.5 M sodium chloride, 10 mM Tris·HCl pH 8 and 3 X with 10 ml of 0.1 M sodium chloride, 10 mM Tris·HCl pH 8, 1 mM EDTA. By determining the total radiation present in the first washes and subtracting this value from the total radiation present at the start of the reaction, the percent of radiation incorporated could be determined. Typically 70% of the counts were incorporated. The column was ready for use at this time.

Preparation of Peripheral Blood

Adult White Leghorn chickens were made anemic by intramuscular injections of 2.5% 1-acetyl-2-phenylhydrazine hydrochloride in 60% ethanol (Williams, 1972). The injection schedule was day 1) 0.7 ml, day 2) 0.7 ml, day 3) 0.6 ml, day 4) 0.4 ml, day 5) 0.8 ml, day 6) 0.8 ml. The proportion of immature reticulocytes increases from near zero to 95% over this time (Sung *et al.*, 1977). Ninety percent of the blood cells are late- or mid-polychromatic (Sung *et al.*, 1977). Peripheral blood from the chickens was collected on the seventh day. The chickens were anaesthetized and the carotid arteries were cut in order to bleed them.

Blood of adult chicken was collected into blood collection buffer at a final dilution of about 1:2 on ice. The blood was filtered through 3 layers of cheese cloth and then centrifuged in a GSA rotor for 10 minutes at 3,000 rpm. The blood was washed two more times in blood collection buffer with pelleting by

centrifugation. After the last centrifugation, the supernatant was aspirated, together with as many white blood cells as possible. The packed cells were either used fresh or were frozen to -80°C until further use.

Preparation of Nuclear Extract

Five ml of RSB and PMSF was added to 5 ml of frozen or fresh chicken erythrocytes (adult immature or mature peripheral erythrocytes or from developmentally staged embryos). The frozen erythrocytes were thawed at 37°C. A pasteur pipette with the tip removed was used to resuspend the cells. The resuspended cells were transferred to a homogenizer with 40 ml of RSB. PMSF, IAC and 500 μl 25% NP-40 were added with rapid mixing. The suspension was homogenized 3-5 strokes, then pelleted at 5,000 rpm for 10 minutes using a SS34 rotor. The supernatant was discarded and the nuclei were resuspended in 30 ml of RSB with PMSF and IAC and homogenized as above. A sample of the homogenized nuclei from this step was used to determine absorbance at 260 nm. To prepare the sample for reading the absorbance, 10 μl of nuclei was vortexed in 1 ml in 5 M urea, 2 M NaCl. The nuclei were pelleted by centrifugation at 4,000 rpm in an SS34 rotor for 10 minutes. The final nuclear pellet was suspended to 200 absorbance units at 260 nm. Four molar sodium chloride was slowly added with mixing to a final concentration of 0.3 M (later it was found that using a final concentration of 0.175 M NaCl resulted in as good GATA-1 and GATA-1 containing protein complexes). At the same time PMSF, IAC, aprotinin and leupeptin were added. The nuclei were extracted for 10 minutes at 37°C with mixing. After extraction, the nuclear debris was removed by centrifugation at 10,000 rpm for 10

minutes in an SS34 rotor. The supernatant was used immediately or glycerol was added to a final concentration of 20% (v/v). The protein was aliquoted and stored at -80°C. Final concentration of the crude nuclear extract was 5 - 15 mg/ml as determined using the Bio-Rad protein microassay. If the protein was to be concentrated, ammonium sulphate was added to a final concentration of 0.35 g/ml and the sample was centrifuged at 14,500 rpm in a SS34 (Sorvall) rotor for 20 minutes, or for 5 minutes in a microcentrifuge to pellet the protein. Initially crude nuclear extracts were dialysed against buffer D. Later it was found that better results were obtained when the protein was not dialysed (see Results). The Dignam *et al.* (1983b) method of preparation of crude nuclear protein extract was also used, and the gel mobility shift patterns seen with these extracts were not noticeably different from patterns obtained with nuclear extract prepared as described with dialysis. Larger or smaller scale preparations were scaled accordingly. Protein concentration was determined using the Bio-Rad protein microassay based on the Bradford method of determining protein concentration.

Uninduced and induced mouse erythroleukemia cells (MEL) were a gift from Dr. Mike Mowat. Extracts made from MEL cells were prepared according to the method of Greenberg and Ziff (1984). Cell were resuspended at 10^7 cells/ml in RSB, and allowed to swell on ice for 10 minutes. NP-40 was added to 0.5% (v/v) with PMSF. The cells were mixed and left on ice for 5 minutes. The nuclei were pelleted by microcentrifuging for 5 minutes. The nuclei were washed once with RSB in the presence of PMSF. The absorbance at 260 nm was determined as above and the cells were resuspended to 200 A_{260} absorbance units per ml. Four M NaCl was added to 0.3 M as above. The crude nuclear extract was dialysed

against buffer D over 5 hours at 4°C and frozen in aliquots to -80°C.

Collection of Erythrocytes From Chicken Embryos

Chicken embryos were staged according to Hamburger and Hamilton (1951). Preparation of nuclear extracts of chicken embryos was according to Minie *et al.* (1992a). The eggs were cracked open at the blunt end of the egg. Enough of the shell was removed so that the embryo was accessible. Day 5 embryos were put into PBS containing 1 mM PMSF. The embryos were minced to release as much blood as possible. The embryos were removed by filtering through cheesecloth. For later stage embryos, the blood was collected by venous puncture. The blood was pelleted by centrifugation and washed once with PBS. The blood was pelleted through lymphocyte separation media, washed once with PBS and then frozen to -80°C or the nuclei were purified and extracted as described above.

Electrophoretic Mobility Shift Assay

The electrophoretic mobility shift assay (EMSA) was an important tool for this research. A protein-DNA binding reaction was prepared by allowing radioactively labelled DNA to interact with protein. The mixture of bound and free DNA was separated electrophoretically and an autoradiogram was made of the gels. Typically, the binding reaction contained ^{32}P-labelled DNA (10,000 to 50,000 counts per minute), poly dI·dC and protein with a final concentration of 1 X EMSA binding buffer in a reaction volume of 10 μl. The amount of protein varied depending on the degree to which it had been purified. Crude nuclear extract was used at about 5 - 20 μg protein/ng of labelled DS1a or 5 - 20 μg protein/0.1 ng of

labelled oligonucleotide. Thus, the final ratio of crude nuclear extract to DNA was 1 to 4 μg of protein per femtamole of labelled DNA. Order of addition of protein to the binding reaction did not affect the complexes formed. Poly dI·dC was added at 100 ng to 3.6 μg per binding reaction (0.5 to 20 μg of crude nuclear extract). For binding reactions using partially purified protein extracts, bovine serum albumin was added to a final concentration of 1 μg/μl (Kozmik *et al.*, 1990). For these reactions the amount of partially purified protein and poly dI·dC added was determined empirically. For example, 10 ng of poly dI·dC with 5 μl of partially purified protein was adequate for prevention of non-specific protein-DNA binding. Addition of NP-40 to a final concentration of 0.05% greatly aided in the detection of shifted complexes, especially with the partially purified protein. The addition of the NP-40 may have prevented the protein from sticking to the eppendorf tube or to the pipette tip used for loading the gel, or it may have helped to stabilize the protein. The NP-40 had to be added before the DNA in order to be effective. The binding reaction was allowed to incubate on ice for 30 minutes prior to electrophoresis. Electrophoresis was through a 1 mm, 4% (20 acrylamide: 1 bis-acrylamide), 0.5 X TBE gel either at room temperature or at 4°C at 200 volts on a 16 cm mini slab apparatus (Idea Scientific) using a PS500X power supply (Hoefer Scientific Instruments). Time of electrophoresis was adjusted empirically so that the free ^{32}P-labelled DNA was just short of running off the gel. After electrophoresis, the gel was dried and an X-ray film was exposed to it in order to prepare an autoradiograph. For competition assays, competitor DNA was added with the radioactively labelled DNA.

Molecular weight determination of protein-DNA complexes were done as

detailed by Bading (1988). The standard binding reaction of DNA and protein was used. The gel was 5% acrylamide (stock 30 acrylamide: 0.8 bis-acrylamide), 0.5 X TBE, which had been pre-run for 1 hour at 160 volts. After loading the binding reactions onto the gel, it was run at 80 volts, for 9 hours when using the DS1 fragment. The molecular weight (M.W.) of the protein-DNA complexes was determined by the formula M.W. = (M/M'-1)K, where K is a constant determined by Bading (1988) to be 215 for a 5% gel, M is the mobility of free DNA and M' is the mobility of DNA shifted by protein complexes.

For the two-dimensional diagonal mobility shift assay (Schaufele *et al.*, 1990), a binding reaction was prepared in which the amount of protein and DNA (radioactive and poly dI·dC) present in 10 μl was scaled up so that 100,000 counts per minute was present. The samples were electrophoresed in the first dimension in a standard EMSA with 0.8 mm spacers. The lane was cut from the wet gel and overlaid to another 4% acrylamide, 0.5 X TBE gel, with 1 mm spacers instead of the 0.8 mm. One ml of a buffer of 1 X binding buffer, 24 μg poly dI·dC and 250 μg of bovine serum albumin was placed on top of the gel slice. A control lane was loaded in parallel with a protein-DNA bound mixture. The electrophoresis apparatus was the Bio-Rad Protean System. The gel was electrophoresed in the second dimension as it was in the first dimension.

Ultraviolet Light-Crosslinking of Protein and DNA

The DNA used for ultraviolet crosslinking was a primer extension labelled oligonucleotide with bromouridine incorporated in place of thymidine. A binding reaction was prepared in which the amount of protein and DNA (radioactive and

poly dI·dC) present in 10 μl was scaled up so that 100,000 counts per minute was present per sample. Duplicate reactions were prepared so that a control, which was not crosslinked by the ultraviolet light, could be prepared. The binding reaction was allowed to proceed for 30 minutes. The eppendorf tubes were covered with plastic wrap instead of the attached lids and the tubes were placed on ice for the crosslinking. Ultraviolet light-crosslinking was accomplished using a U.V. Stratalinker 2400 (254 nm). Time of crosslinking was determined empirically, and good results were obtained with 7 - 30 minutes (7 to 28 Joule/cm^2) of crosslinking. After the crosslinking was done, aliquots of each tube were removed for EMSAs. $CaCl_2$ and $MgCl_2$ were added to 8 and 5 mM, respectively. DNase I (Sigma) and micrococcal nuclease were added to final concentrations of 0.07 mg/ml and 0.02 units/ml, respectively. The DNA was digested for 30 minutes at 37°C. The reaction was stopped by adding 1 volume of 2 X SDS sample buffer. The samples were boiled for 2 minutes and then electrophoretically separated on a SDS 10% or 12% polyacrylamide gel with ^{14}C-labelled protein standards as markers. After electrophoresis the gels were dried, and an autoradiogram was prepared to visualize the proteins.

DNase I Protection Assays

A DNA fragment which had been labelled at a unique site with ^{32}P, either at the 5′ or the 3′ end of the fragment was used for these experiments. The binding reaction of the EMSA was scaled up for DNase I protection assays, so that 100,000 to 200,000 counts per minute of radioactivity were present in the reaction. After the DNA and protein had been allowed to bind, $MgCl_2$ was added to

create a final concentration of 5 mM. DNase I (RQ1® Promega) was added and digestion was allowed to proceed for 1 minute at 25°C. The reaction was stopped with 1/10 volume of 15 mM EDTA pH 7.2, 0.2% SDS. One-tenth volume of 1 mg/ml calf thymus DNA was added to the tube. After a phenol/chloroform extraction, the DNA was ethanol precipitated and then resuspended in formamide loading buffer. Two controls were used, a control of unbound DNA, and a control of the labelled DNA which had been chemically sequenced (see below).

Preparative EMSAs were also used for DNase I protection assays. In this case, the binding reaction was scaled up to a total 1 X 10^6 counts per minute. The binding and DNase I digestion were done as described, however the reaction was stopped with only EDTA. The sample was electrophoresed through a preparative EMSA. An autoradiograph was made of the wet gel in order to determine where the bands were. The bands were then cut out of the gel. DNA was eluted from each band in the presence of 400 μl of 20 mM HEPES pH 7.5, 0.5 M sodium chloride, 1 mM EDTA overnight at 37°C. The buffer was removed from each gel slice, which was washed once with 100 μl of buffer. The 400 and 100 μl fractions were pooled and passed through siliconized glass to remove gel particles. A phenol/chloroform extraction was done, carrier calf thymus DNA was added to 0.15 mg/ml and the labelled DNA was precipitated with ethanol. The final product was dissolved in equal volumes of water and formamide loading buffer, and just before electrophoresis boiled for 2 minutes to denature the sample. Ten-thousand counts were applied per lane and the partially digested DNA was separated using a sequencing gel. The electrophoresis apparatus used was a Bio-Rad Sequi-Gen® Nucleic Acid Sequencing Cell and the power supply was a model 3000/300 Bio-Rad

Power Pack.

Chemical Sequencing

Chemical sequencing of DNA based on Maxam and Gilbert (1980) was done to provide a control lane for DNase I protection assays. The G + A, and C + T reactions were used. Approximately 100,000 counts per minute of end-labelled DNA was used. Each reaction had 1 μg of carrier DNA added. For the G + A reaction, the DNA was diluted to 30 μl with water and then 1 μl of 1 M formic acid was added. The reaction was incubated 25 minutes at 37°C. For the C + T reaction, the radioactive DNA was diluted to 10 μl with water and 15 μl of hydrazine was added. The reaction was incubated at room temperature for 10 minutes. The above reaction times were varied empirically so that not too much or too little modification of the labelled DNA occurred. After the incubation, 1 ml of n-butanol was added, the sample was mixed well and microcentrifugation for 1 minute caused precipitation of the DNA. The butanol was removed.

Cleavage of the modified DNA was the same for both the G + A and C + T reactions. Freshly prepared 1 M piperidine (150 μl) was added to the reaction which would then be placed on ice until it was heated to 90°C for 30 minutes. The reaction was cooled to room temperature. N-butanol (1.2 ml) was added, the sample was mixed well and microcentrifuged for 1 minute to pellet the DNA. The DNA was dissolved in 150 μl of water and moved to a fresh tube. The DNA was reprecipitated with 1 milliliter of n-butanol. The final product was dissolved in equal volumes of water and formamide loading buffer. In DNase I protection assays 10,000 counts were applied as a control lane to a sequencing gel.

Protein Purification

Heparin-Agarose Column Chromatography

Heparin-agarose column chromatography was done according to Plumb *et al.* (1989). Crude nuclear extract was diluted to less than 5 mg/ml in buffer D with a final potassium chloride concentration of 0.2 molar. PMSF, iodoacetamide, aprotinin and leupeptin were added at the concentrations defined in the list of buffers. The protein was applied to the heparin agarose column at a ratio of 6 mg of protein per ml of column. Three or more column volumes of buffer D with 0.2 M potassium chloride containing 1 mM PMSF were used to wash the column. The column was then eluted with buffer D containing 0.4 M potassium chloride with PMSF. Flow rates of 2 ml/minute were used for a column of 30 ml in a column with a 2.5 centimeter diameter. Fractions of 10 ml were collected. Protein peaks were monitored by measuring absorbance at 280 nm. GATA-1 activity was monitored with the EMSA. GATA-1 activity elutes at 0.4 M potassium chloride. Protein recovery of approximately 10% was seen. The active fraction from the heparin-agarose column could be stored at -80°C for one to two weeks. Storage resulted in a loss of complexed GATA-1 activity, so that more of the GATA-1 activity was seen as GATA-1 alone rather than as higher GATA-1 containing protein complexes.

S-Sepharose Column Chromatography

Protein from the heparin agarose column (0.4 M KCl) was diluted by ½ with buffer E. NP-40 was added depending on the step to be used after the S-sepharose column. For instance NP-40 is inappropriate for application to the phenyl-agarose

column, if it was to be used subsequent to the S-sepharose column. Recovery of GATA-1 and GATA-1-complex activity appeared to be better with NP-40 present. Approximately 2.5 milligram of protein was applied per milliliter of S-sepharose. The column was washed with 3 or more column volumes each of buffer E with 0.2 M NaCl, buffer E with 0.3 M NaCl, buffer E with 0.5 M NaCl. Fractions of about 5 ml were collected. The flow rate for the column was 1 ml per minute for a 7.5 ml column with a diameter of 1.5 cm. Protein peaks were monitored by measuring absorbance at 280 nm. GATA-1 activity was monitored with the EMSA. Storing these samples by freezing to -80°C resulted in significant loss of GATA-binding activity. Approximately 35% of the protein applied to the S-sepharose column was collected in the 0.5 M NaCl fraction and 4% in the 0.3 M fraction.

Phenyl-Agarose Column Chromatography

Protein from the heparin-agarose or S-sepharose column was applied to the phenyl-agarose column. The protein was applied at 10 mg/ml of column matrix. The protein from the heparin-agarose column was diluted to ½ with buffer D (without glycerol) with the addition of dry ammonium sulphate to a final concentration of 0.7 molar for application to the column. The column chromatography was done with a Pharmacia FPLC. The column was developed with a gradient from 0.7 M ammonium sulphate to 0 M ammonium sulphate in buffer D (without glycerol). Fractions of 1 ml were collected. Flow rate for the column was 1 ml per minute for a 1 ml column of 0.5 cm diameter. Protein peaks were monitored by measuring absorbance at 280 nm. GATA-1 activity was monitored with the EMSA. After the protein samples were collected from the

column, PMSF, NP-40 to 0.1% and glycerol to 20% were added. These samples could be stored at -80°C. It was not determined how long these fractions could be stored at -80° without significant loss of activity.

Double-Stranded DNA Cellulose Column Chromatography

The double-stranded DNA cellulose column chromatography was done according to Evans and Felsenfeld (1989). The matrix used was a 2:1 mixture of cellulose:double-stranded DNA cellulose. Partially purified protein was diluted to less than 0.05 M of the salt of the previous column with buffer G (without ammonium sulphate). The protein was applied to the column and eluted in a step gradient from 0.075 to 0.25 M ammonium sulphate in buffer G (with PMSF) in increments of 25 mM. For a 7 ml column of 1.5 cm diameter, 50 mg of protein was applied and eluted with 10 ml at each salt concentration. Flow rates were 1 ml/min. Fractions consisted of the entire eluate at each salt concentration. GATA-binding activity from this column was stable when frozen to -80°C, presumably because of the presence of NP-40 in the buffer. It was not determined how long fractions from this column were stable at -80°C.

Single-Stranded DNA Cellulose Column Chromatography

The single-stranded DNA cellulose column chromatography was done according to Evans and Felsenfeld (1989). The matrix used was a 2:1 mixture of cellulose:single-stranded DNA cellulose. Partially purified protein was diluted to less than 0.1 molar salt with buffer H (without potassium chloride). The protein was applied to the column and eluted in a step gradient from 0.125 to 0.3 M potassium

chloride in buffer H (with PMSF) in 25 mM increments. For a 7 ml column of 1.5 cm diameter, 50 mg of protein was applied and eluted with 20 ml at each salt concentration at flow rates of 1 ml/min. Fractions consisted of the entire eluate at each concentration. GATA-binding activity from this column was stable when frozen to -80°C, presumably because of the presence of NP-40 in the buffer. It was not determined how long fractions from this column were stable at -80°C.

GATA-Affinity Column Chromatography

Partially purified protein was diluted with buffer D (without potassium chloride) so that the salt concentration of the previous column was below 0.2 M. NP-40 and PMSF were added, after which the protein was applied to the column. The column was developed with at least three column volumes of buffer D with 0.2 M KCl, NP-40 and PMSF and eluted in 3 column volumes with buffer D with 0.4 M KCl. The column volume was 1 or 2 ml and the amount of protein applied was determined empirically. The entire eluate at 0.4 M KCl was collected. Flow rate for the column was 1 ml per minute. GATA-1 activity was monitored with the EMSA.

Ammonium Sulphate Precipitation of Proteins

Ammonium sulphate precipitation of proteins was used for some experiments. After precipitation the proteins were resuspended in a smaller volume in order concentrate the proteins. Ammonium sulphate was added and dissolved to 0.35 grams per ml of protein sample. The precipitated protein was collected by microcentrifuging at 4°C for 10 minutes. It was then resuspended in a smaller

volume of appropriate buffer. If the protein was resuspended in a buffer containing NP-40, recovery of activity was usually good. Precipitation of the proteins worked best if NP-40 was not present. If NP-40 was present, but glycerol levels were below 10%, the precipitation still worked. At higher glycerol concentrations, GATA binding activity was present in a precipitate that appeared at the top of the sample after centrifugation.

TCA Precipitation of Protein

Casamino acids (20% w/v) were added if protein samples were dilute. Protein samples were made 20% with respect to cold trichloroacetic acid (TCA). The samples were centrifuged 30 minutes at 10,000 rpm in an SS34 rotor or for 10 minutes in a microcentrifuge. The protein pellet was washed once with 20% TCA and then twice with acetone. The pellets were dried and resuspended in SDS loading buffer.

Staining and Silver Staining Polyacrylamide Gels

The acrylamide gel was stained with Coomassie Blue G-250 (45% methanol, 1.5 M acetic acid, 0.04% Coomassie Blue G-250). After staining the gel was destained rapidly in a 12.5% methanol, 2.2 M acetic acid solution. When the gel had been partially destained, further destaining was achieved in a solution of 5% methanol, 1.3 M acid acid. For silver staining, the gel was soaked in 50% methanol overnight with at least one change of solution. This was followed by 30 minute washes in the following buffers: water, 1 mM dithiothreitol, 0.1% (w/v) silver nitrate. The gel was washed rapidly with water and then 3% sodium citrate

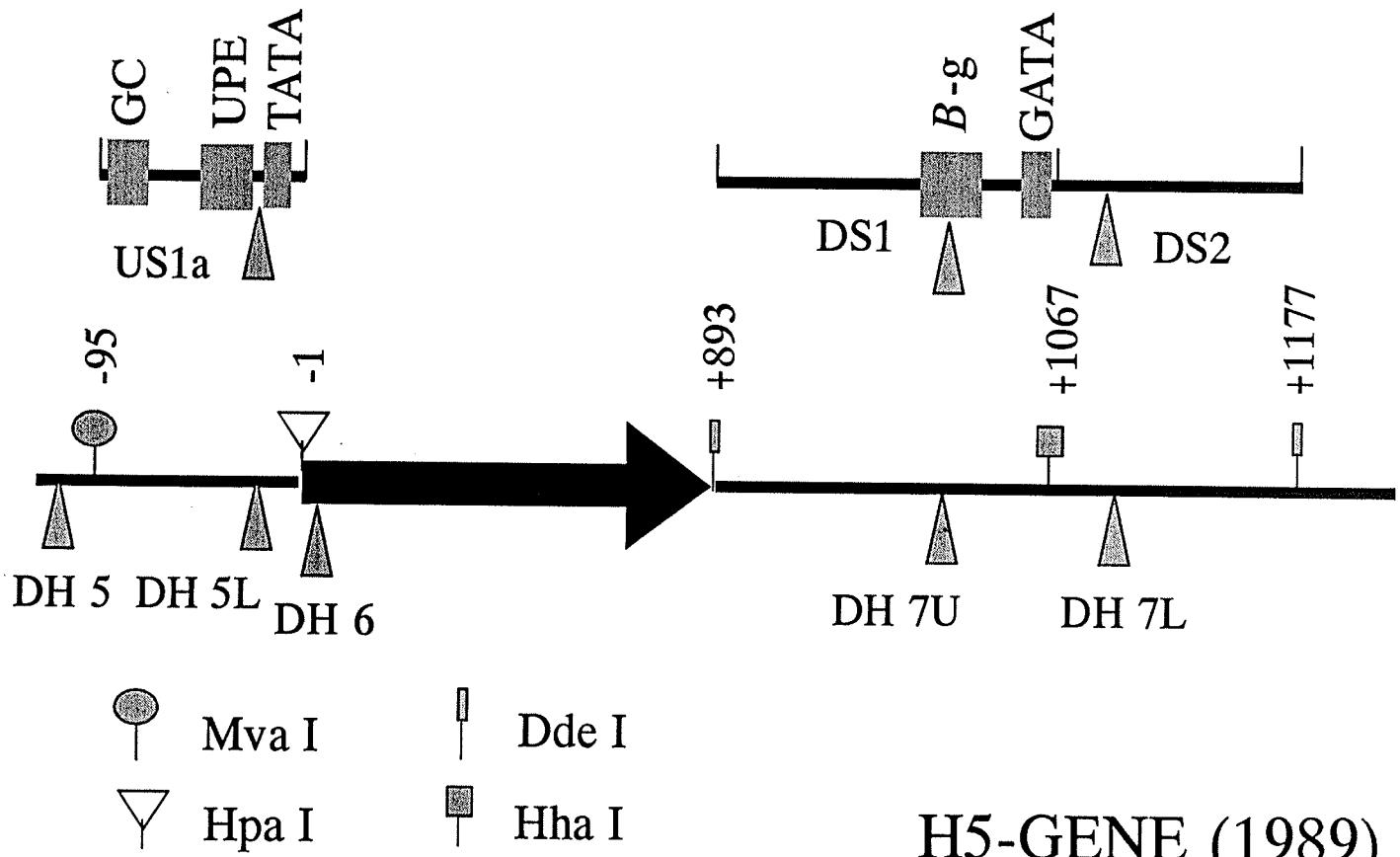
containing 0.025% formaldehyde and then stained with 3% sodium citrate. The colour reaction was stopped by adding 1/10 volume 2.3 M citric acid.

Results

An important diagnostic tool used for this research was the electrophoretic mobility shift assay (EMSA). In the EMSA, protein is allowed to interact with radioactively labelled DNA sequences which are potential binding sites for sequence-specific DNA-binding proteins. After the protein-DNA complexes have formed, they are electrophoresed through a non-denaturing gel, in order to separate protein-bound DNA from free DNA. Although charge and conformation of the DNA-protein complexes affects migration during electrophoresis, the protein-DNA complexes are also separated on the basis of mass. Therefore when an autoradiogram of the gel is prepared, the positions of the labelled DNA are detected. Unbound DNA migrates furthest through the gel and protein-bound DNA is retarded or shifted relative to it. DNA bound by small proteins is shifted less than DNA bound by larger proteins or protein complexes.

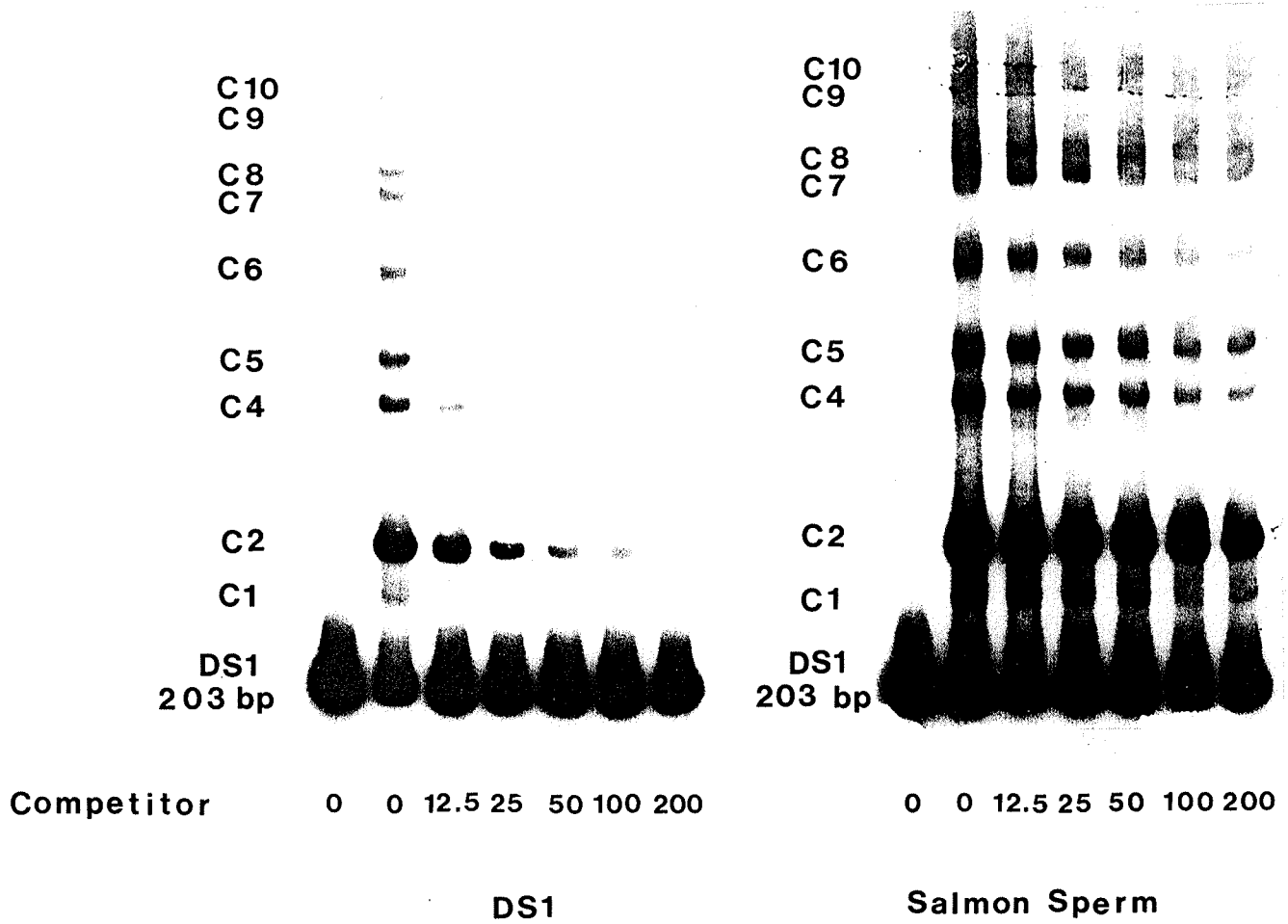
In this project the 5′ promoter and 3′ enhancer of the chicken histone H5 gene were investigated. At the start of this project, the histone H5 gene had been partially characterized. This gene is characterized by a number of DNase hypersensitive sites (Figure 7; Rousseau *et al.*, 1989). Deletion mutants of the gene had been used in transient transfection assays to determine the downstream enhancer and regions of the promoter which were important for transcription, and by comparison of the sequences of the histone H5 gene, potential binding sites for few transcription factors were known. Within the promoter region there is a GC box (-83 to -75) which is a binding site for Sp1 and an UPE located at -51 to -33 just 5′ of the TATA box region (Rousseau *et al.*, 1989). DH 5_L maps between the latter two regions (Figure 7). Factors binding to the promoter between -95 to -1

Figure 7. Model of histone H5 gene in 1989. An enhancer is located 3′ of the protein coding region. The DNase I hypersensitive sites in the region of the promoter and in the 3′ enhancer are indicated. The elements are described in the text.


GC
UPE
TATA
US1a
B-g
GATA
DS1
DS2
-95
-1
+893
+1067
+1177
DH 5
DH 5L
DH 6
DH 7U
DH 7L
Mva I
Hpa I
Dde I
Hha I

H5-GENE (1989)

Figure 8. Specific interactions between the histone H5 enhancer region and factors from adult chicken immature erythrocyte nuclear extracts. A 203 bp Xba I-EcoR I (DS1) end-labelled DNA fragment was incubated in the presence or absence of nuclear proteins extracted from adult chicken immature erythrocytes and used for EMSAs. The left lane in each panel did not contain any protein and the other lanes each contained 10 μg of proteins and 1 ng of labelled DNA. The second lane from the left does not contain any competitor DNA, and the next lanes contain increasing amounts of competitor DNA as indicated below each lane. Competitor DNA amounts are in molar excess of DS1 (left panel) and excess by weight for salmon sperm DNA (right panel). DS1 is the free DNA fragment, and C1 - C10 are the protein-DNA complexes.


C10
C9
C8
C7
C6
C5
C4
C2
C1
DS1
203 bp
C10
C9
C8
C7
C6
C5
C4
C2
C1
DS1
203 bp
Competitor
0 0 12.5 25 50 100 200
0 0 12.5 25 50 100 200
DS1
Salmon Sperm

were investigated. The 3′ enhancer is located in the region +851 to +1185 (Rousseau *et al.*, 1989; Trainor *et al.*, 1987) and factors binding to this region were studied. The 3′ enhancer has a GATA binding site and a 34 base pair region with sequence similarity to the site bound by NF-E4 in chicken β^A-globin (Trainor and Engel, 1987; Emerson *et al.*, 1987; Gallarda *et al.*, 1989). DH sites 7_U and 7_L map at 1003 and 1095 base pairs downstream of the start of transcription respectively. For the purpose of this study the 3′ enhancer was subdivided into DS1 and DS2 (Figure 6). DS1 was further divided into smaller fragments as indicated in Figure 6. These smaller fragments are described in Results with the data relating to them. Oligonucleotides which cover protein binding sites within the histone H5 gene elements were also synthesized and used in the study. These smaller fragments are described in Materials and Methods and are explained in Results with the data relating to them.

Most of the data concerning interactions of regions of the histone H5 gene with proteins were examined using extracts from nuclei of adult chicken immature erythrocytes. The immature reticulocytes make up 95% of the red blood cells at this time and 90% of the blood cells are late- or mid-polychromatic (Sung *et al.*, 1977; Figure 3). Therefore, the adult immature erythrocyte represents a stage in development of the tissue in which erythroid genes are actively expressed. This suggests that interactions between nuclear proteins and DNA sequences of the erythroid-specific gene histone H5 are expected to affect transcription (or chromatin structure as it relates to transcription) in a positive manner. Interactions of H5 gene elements with extracts from erythrocytes of chicken embryos at three stages of development (stages 28, 38 and 40), with extracts from chicken mature

erythrocyte nuclei and from uninduced and induced mouse erythroleukemia (MEL) cell nuclei were also studied. Nuclear extracts of chicken embryos were used in studies to determine if factors interacting with GATA-1 in adult chicken immature erythrocytes were also present in the embryo. This was of interest because at stage 28 the embryonic ß-globin genes epsilon and rho start to become repressed and the adult ß-globin genes start to be expressed (Emerson *et al.*, 1989). At stages 38 and 40, the adult ß-globin genes $ß^A$ and $ß^H$ are expressed (Emerson *et al.*, 1989). Therefore differences in GATA-1 multisubunit complexes might reflect the changing pattern of expressed ß-globin genes. Protein interactions on the histone H5 gene of the adult chicken mature erythrocyte were of interest because at this stage gene expression is decreased (Gomez *et al.*, 1992; Jackson *et al.*, 1989).

The other system which received brief attention was the uninduced and induced MEL cells. It was of interest to know if GATA-1 multisubunit complexes were present in MEL cells. The uninduced and induced cells also represent a population of cells prior to and during expression of mature erythrocyte genes, respectively. Therefore, changes in DNA-binding activities of proteins between induction states could represent alterations which affect expression of mature erythrocyte genes. These cells were studied for the presence of mouse GATA-1 multisubunit complexes.

Proteins Binding to DS1

The EMSA was used to identify the nuclear proteins that bind to the histone H5 3′ enhancer. The 203 bp fragment containing DS1 (Figure 6; Trainor *et al.*, 1987; Rousseau *et al.*, 1989) was incubated with a nuclear extract from adult

chicken immature erythrocytes. Figure 8 shows that ten or more shifted bands were generated, indicating the presence of several binding activities. Complex C3 does not appear in the gel-shift pattern shown in Figure 8, but in later data will be shown to be due to non-specific protein binding to DS1, particularly DS1a. All of the other complexes showed specific binding to DS1. A 100-fold molar excess of unlabelled DS1 competed effectively for all of these complexes, whereas the non-specific competitor salmon sperm DNA did not compete effectively, even at high levels. Demczuk *et al.* (1990) noted that addition of DNA probe to the binding reaction before the nuclear protein extract promotes the formation of non-specific protein-DNA complexes. The order of addition of protein to the binding mixture did not alter the number of complexes observed on the histone H5 DNA sequences.

The Majority of the Complexes Forming on DS1 are Associated with a DNA Segment Containing the GATA-Binding Site

An experiment similar to the stairway assay was used to determine where the complexes were forming with DS1 (van Wijnen *et al.*, 1992). In the stairway assay, a DNA sequence is subdivided into fragments, which are then used in EMSAs. EMSAs for the DNA fragments are compared to determine the region of DNA which binds a particular complex. The DS1 fragment was subdivided into a number of smaller fragments and EMSAs were performed on the various segments of DS1. By comparing the intensity of the complexes formed with the DNA fragment with different nuclear extracts, as well as their position in the gel pattern, the complexes formed with the DNA fragments were assigned. Figure 9A shows that, with two different nuclear extracts, complexes C1, C2, a weak C3 and C4-

Figure 9. Multiple protein-DNA complexes are formed with the GATA region of DS1. The lower part of this Figure shows the DNA fragments that were used in the EMSA. The arrows show the position of the Dde I (+893) and Xma III (+1064) restriction endonuclease cleavage sites. The fragment enclosed by these sites forms DS1. Restriction endonuclease cleavage sites (Hind III, Xba I, Kpn I and EcoR I) in the multiple cloning region of pGEM-3 and restriction endonuclease sites (Mva I and Hha I) in DS1 are shown. The restriction endonuclease cleavage site that was labelled is indicated with an asterisk. Note that the 41 bp Hha I-EcoR I fragment (C) was contaminated with a longer fragment. The 23 bp oligonucleotide H5-GATA was synthesized as described in Materials and Methods. The DNA fragments used for each of the EMSAs are indicated at the bottom of each panel. In each case the left lane is control DNA with no added protein. The middle and right lanes show the protein-DNA complexes formed with two different nuclear extracts (10 μg of protein/ng of DNA) of adult chicken immature erythrocytes. H5-GATA is the free oligonucleotide, and C1 - C10 are the protein-DNA complexes. EMSAs were performed with six different nuclear extracts and each of the DNA fragments shown. By comparing the intensities of various complexes formed with each DNA fragment and nuclear extract, the complexes which formed with each of the DNA fragments were assigned.

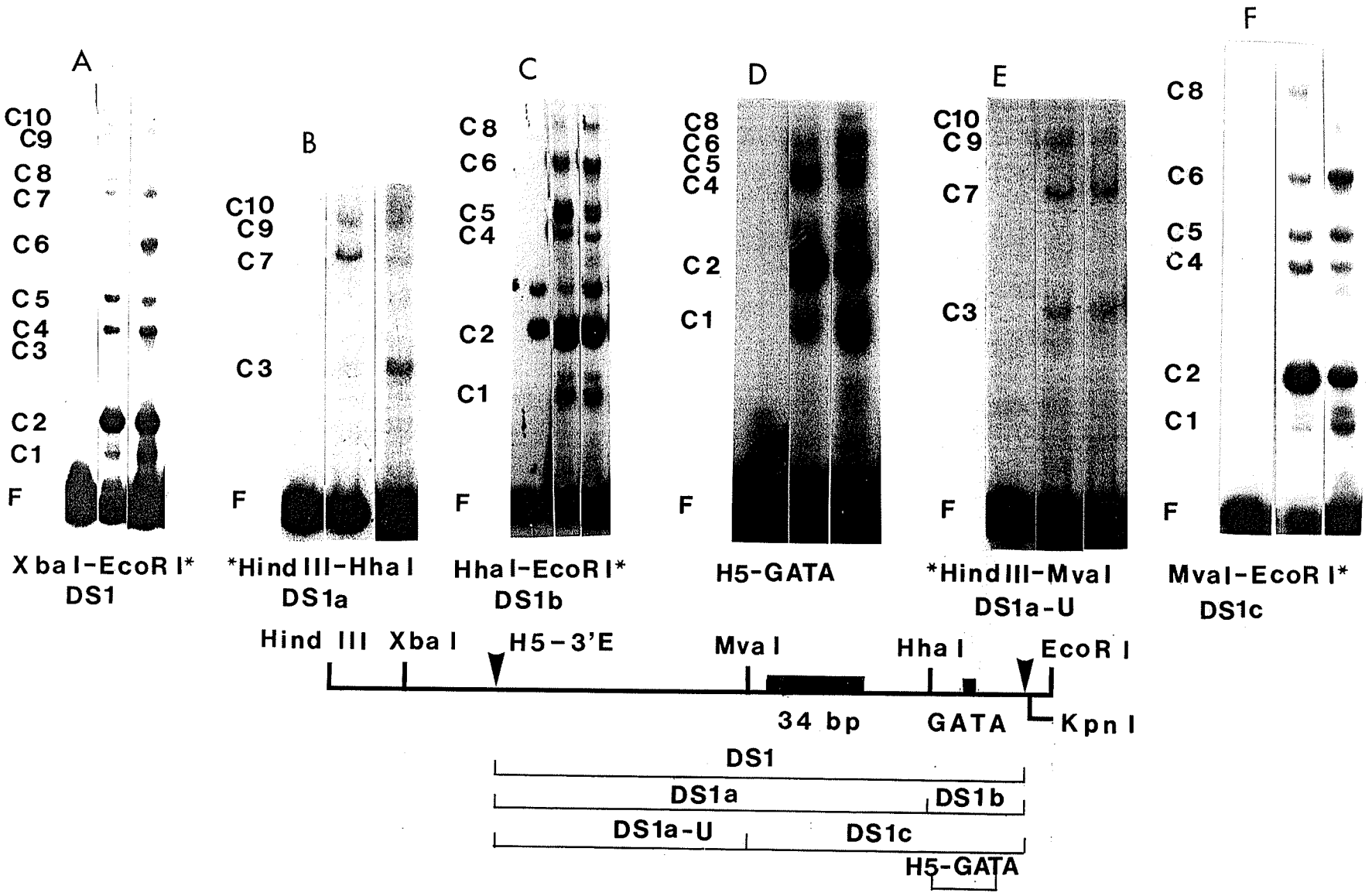
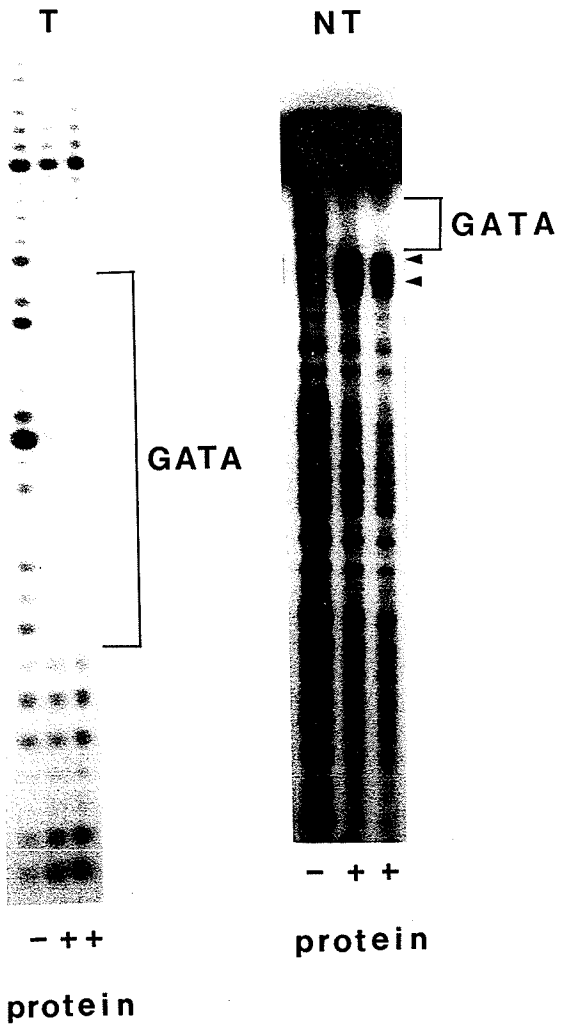

A
C10
C9
C8
C7
C6
C5
C4
C3
C2
C1
F
Xba I-EcoR I*
DS1
B
C10
C9
C7
C3
F
*Hind III-Hha I
DS1a
C
C8
C6
C5
C4
C2
C1
F
Hha I-EcoR I*
DS1b
D
C8
C6
C5
C4
C2
C1
F
H5-GATA
E
C10
C9
C7
C3
F
*Hind III-Mva I
DS1a-U
F
C8
C6
C5
C4
C2
C1
F
Mva I-EcoR I*
DS1c
Hind III
Xba I
H5-3'E
Mva I
Hha I
EcoR I
34 bp
GATA
Kpn I
DS1
DS1a
DS1b
DS1a-U
DS1c
H5-GATA

Figure 10. DNase I protection analysis of the histone H5 DS1. Two 3' end-labelled DNA fragments, a 203 bp Xba I-EcoR I [labelled template (T) strand] and a 240 bp Hind III-Kpn I [labelled non-template (NT) strand], containing DS1 were incubated with (+) or without (-) proteins (P) (10 μg of protein/ng of DNA) extracted from adult chicken immature erythrocyte nuclei. The DNase I-resistant region generated by the GATA protein (Rousseau *et al.*, 1989) is shown. The arrowheads indicate the DH sites of the GATA-DNA complex. The DNA sequence was determined by comparison of the control lane (-protein) to other gels in which a chemical sequencing lane was present.

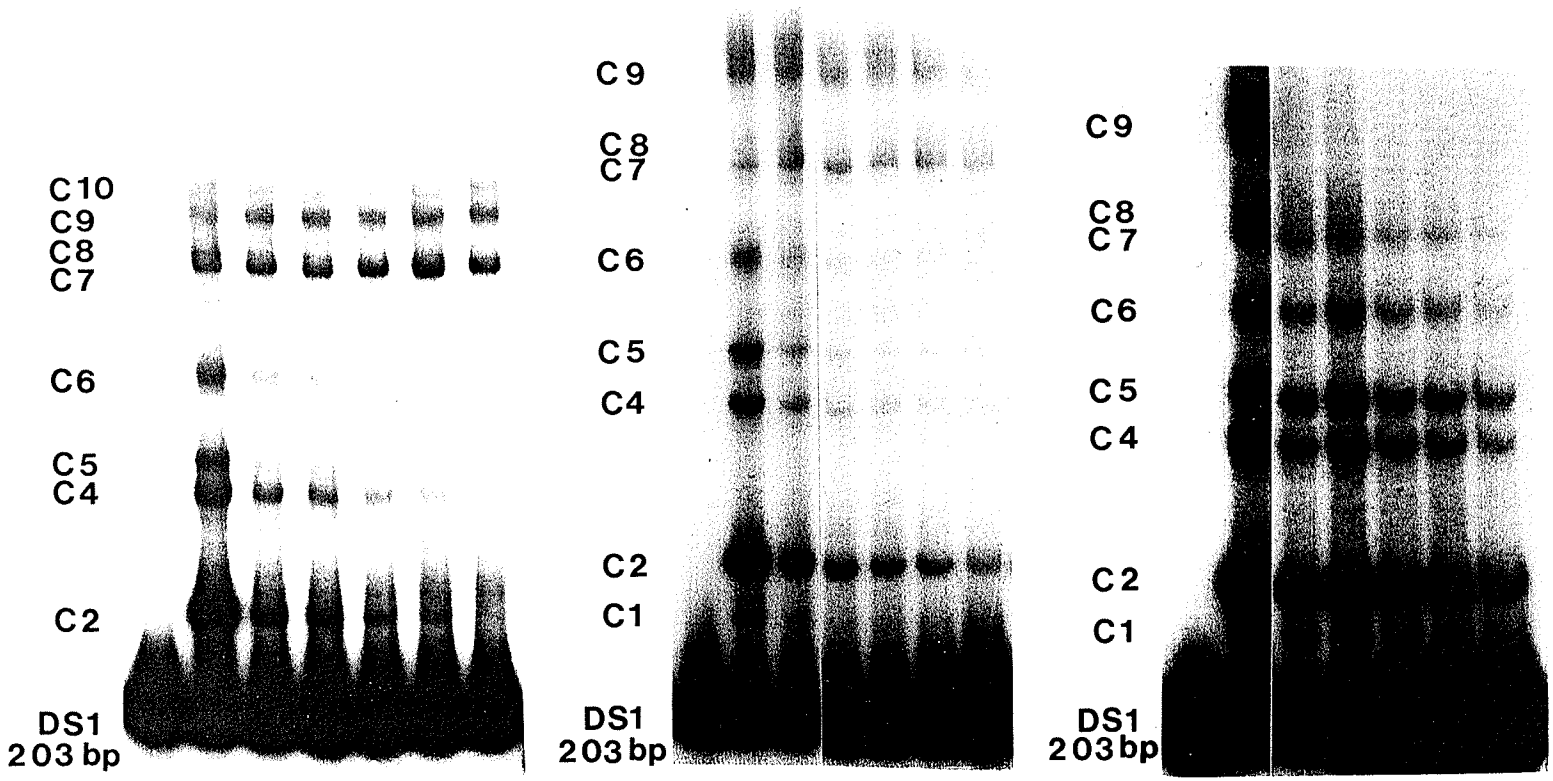

T
NT
GATA
GATA
- + +
protein
- + +
protein

C10 were formed on DS1. Complexes C3, C7, C9 and C10 were associated with DS1a (207 bp HindIII-HhaI) (Figure 9B) and DS1a-U (146 bp HindIII-MvaI, Figure 9E). These data are shown in more detail in subsequent sections of the results. A 23 bp oligonucleotide (H5-GATA) containing the sequence AGATAA was also used in the EMSA. Figure 9D shows that complexes C1, C2, C4-C6, and C8 formed with this oligonucleotide and with DS1b. Thus the GATA DNA-binding sequence (AGATAA) and as little as 8 bp on either side of it were sufficient to generate complexes C1, C2, C4-C6 and C8. These complexes were independent of C3, C7, C9 and C10 since the two groups of complexes (C1, C2, C4-C6 and C8) and (C3, C7, C9 and C10) were seen with the two separate regions of the DS1 DNA fragment. The GATA binding site of DS1 was also protected from DNase I nuclease digestion when the DNA was incubated with adult chicken immature erythrocyte nuclear protein. This indicates that proteins, which bind the GATA site of DS1, are present in the nuclear extract (Figure 10; Rousseau *et al.*, 1989).

Multiple Factor Complexes are Formed with the GATA-Binding Site

The results presented in Figure 9 demonstrated that several complexes were formed with an oligonucleotide containing the GATA-binding site AGATAA. We investigated whether DNA fragments with the GATA sequence could compete for complexes C1, C2, C4-C6 and C8 forming with the 203-bp DS1 DNA fragment. Figure 11 shows that the 23 bp H5-GATA oligonucleotide selectively and effectively competed for complexes C1, C2, C4-C6 and C8. Complexes C7, C9 and C10 were relatively unaffected. A 328-bp XbaI-EcoRI DNA fragment that harbours the ß-globin 3′ enhancer and contains two binding sites for GATA

Figure 11. DNA fragments containing the GATA sequence compete for complexes C1, C2, C4-C6 and C8. The left lane in each panel did not contain any protein and the other lanes contained 10 μg of nuclear extracted protein from chicken immature erythrocytes and 1 ng of DNA (3' end-labelled 203-bp Xba I-EcoR I DNA fragment containing DS1). The second lane from the left does not contain any competitor DNA, and the next lanes contain increasing amounts of competitor DNA as indicated in molar excess. Competitor DNA was the synthesized 23 bp oligonucleotide H5-GATA, a 328 bp Xba I-EcoR I restriction fragment containing the ß-globin-3' enhancer and a 528 bp Xba I-EcoR I restriction fragment containing the H5-5' promoter. DS1 is the free DNA fragment, and C1 - C10 are the protein-DNA complexes.


C10
C9
C8
C7
C6
C5
C4
C2
DS1
203 bp
C9
C8
C7
C6
C5
C4
C2
C1
DS1
203 bp
C9
C8
C7
C6
C5
C4
C2
C1
DS1
203 bp
Competitor
0 0 12.5 25 50 100 200
0 0 12.5 25 50 100 200
0 0 12.5 25 50 100 200
H5-GATA
ß-GLOBIN-3'E
H5-Promoter

(AGATAA and TGATAG) also selectively and effectively competed for the same complexes as the H5-GATA oligonucleotide. It should be noted that the 328-bp ß-globin enhancer fragment contained an NF-E4 binding site (Gallarda et al., 1989) the DNA segment with sequence similarity to the 34 bp region of the histone H5 enhancer (Trainor and Engel, 1987). However, this DNA fragment did not effectively compete with complexes C7, C9 and C10. A 528-bp XbaI-EcoRI DNA fragment containing the histone H5 5′ promoter region (H5-5′ P) was also used in the competition experiments. This DNA fragment does not contain a GATA consensus site. Figure 11 shows that the higher molecular mass complexes C8, C9 and C10, but not C1-C7, were competed at low levels of H5-5′ P DNA. The oligonucleotide containing the Sp1 binding site from the Stratagene Hotfoot kit (Materials and Methods) was not effective at preventing formation of any of the GATA-1 multisubunit complexes (data not shown).

To investigate further whether the nucleotides flanking the GATA-binding site influenced the generation of multiple complexes, another 23 bp GATA-containing oligonucleotide with the sequence of one of the GATA-binding sites in the ß-globin 3′ enhancer (DNase I protected region IV) was synthesized. Figure 12 shows the results of EMSAs with labelled oligonucleotides H5-GATA and ß-globin-GATA. The same complexes were formed with both oligonucleotides, demonstrating that the DNA sequence surrounding the consensus GATA sequence was not important in forming the complexes C1, C2, C4-C6 and C8.

The DNA-binding location of each of the complexes C1, C2, C4-C6 and C8 was demonstrated by DNase I protection assays. The non-template strand of the 240 bp HindIII-KpnI DS1 DNA fragment was end-labelled and incubated with a

Figure 12. Complexes C1, C2, C4 - C6 and C8 form with oligonucleotides containing the GATA-binding site. EMSAs were done with nuclear extract and each of the oligonucleotides H5-GATA and the β^A-globin enhancer GATA binding sequence as described in Materials and Methods. In each of the panels the left lane does not contain protein and the right lane has nuclear extracted protein from chicken immature erythrocytes (10 μg of protein/ng of DNA) added to the assay. The 23 bp oligonucleotides with the GATA sequence were synthesized as described in Materials and Methods.

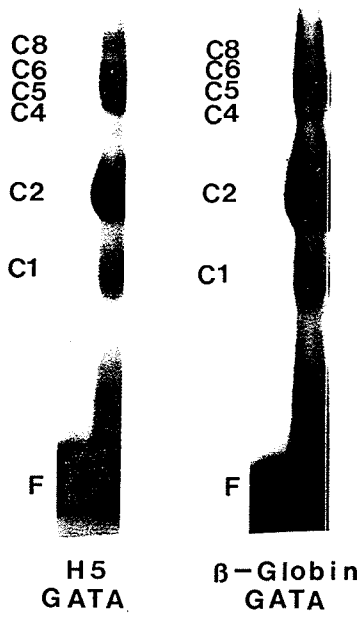
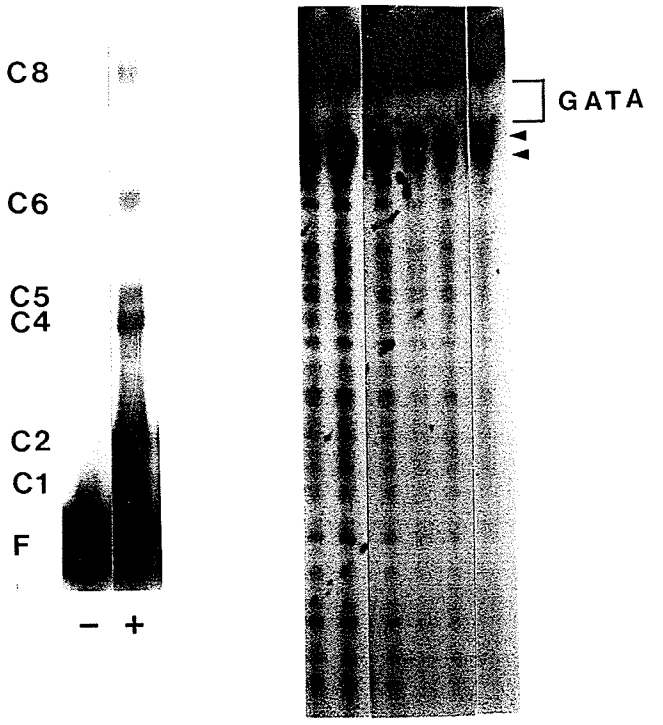

C8
C6
C5
C4
C2
C1
F
H5
GATA
C8
C6
C5
C4
C2
C1
F
β−Globin
GATA

Figure 13. DNase I protection analysis of the complexes of histone H5 DS1. The 3′ end-labelled DNA fragment, a 240 base pair Hind III-Kpn I (labelled non-template strand) containing DS1 was incubated with (+) or without (-) (10 μg protein per ng DNA) proteins extracted from adult chicken immature erythrocyte nuclei. After partial DNase I digestion of the protein-bound DNA, the complexes were separated by preparative EMSA as described in Materials and Methods. The DNA of each complex was purified and the DNase I protected region was revealed on a denaturing 6% polyacrylamide (DNA sequencing gel). The complex from which the DNA was isolated is indicated at the lower edge of each lane of the DNase I protection assay. The arrowheads indicate sites which are hypersensitive to DNase I digestion. The DNA sequence had been determined from control chemical sequencing reactions as detailed in Materials and Methods.

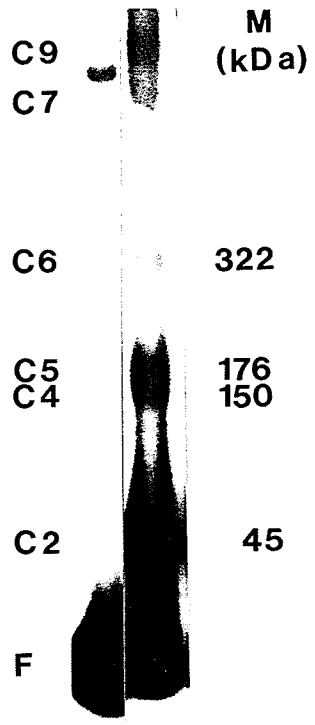

C8
C6
C5
C4
C2
C1
F
– +
GATA
1 2 4 5 6 8 Complex
EMSA
DNase 1
Protection
Assey

nuclear extract from adult chicken immature erythrocytes, after which the DNA was partially digested with DNase I. After isolation of each complex by electrophoresis on a preparative gel, the DNA of each complex was purified and electrophoretically separated on a sequencing gel to reveal sequences protected from DNase I digestion due to the presence of bound protein. The characteristic GATA protected site (Figure 10) was present in DNA purified from each of the complexes (Figure 13). These results provide evidence that the only protein in the complexes that is binding to DNA is the GATA protein, suggesting that these complexes are generated by the interaction of heterologous proteins with GATA-1, since GATA-1 does not dimerize (Evans and Felsenfeld, 1989).

The GATA-1 DNA-Binding Protein is Common to at Least Six Complexes

The approximate molecular mass of the DNA-bound proteins in the complexes was determined by the method described by Bading (1988). This method is subject to a number of assumptions, including that the protein does not bend DNA and the charge of the protein is negligible compared with the charge of the DNA fragment. The limit of mass determination is 450 kDa. Thus this method provides a rough estimate of the molecular mass of the DNA-binding protein. The end-labelled 240 bp HindIII-KpnI H5-3'E DNA was incubated with a nuclear extract isolated from chicken immature erythrocytes, and the complexes were resolved on a 5% polyacrylamide gel (Figure 14). The migration of the complex relative to that of the free DNA was used to calculate the molecular mass of the bound protein(s) (Materials and Methods, Bading, 1988). The protein in complex C2 had a molecular mass of approximately 45 kDa, which is consistent with the bound protein being

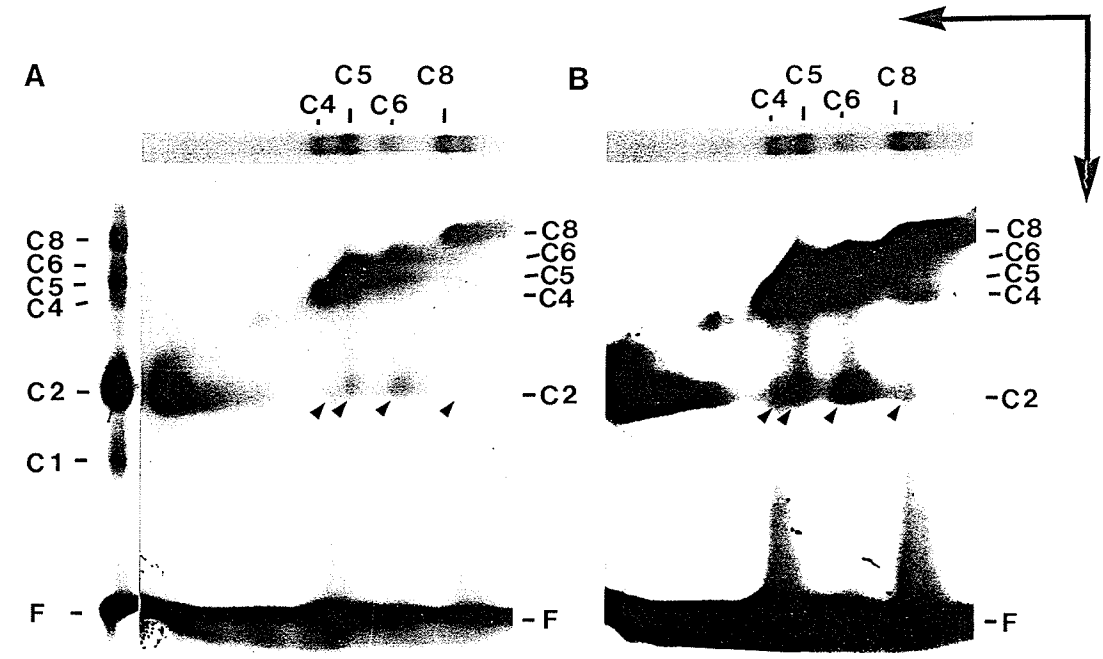
Figure 14. Molecular weight determination of DS1 complexes. The EMSA and molecular weight determinations were according to Bading (1988) in a non-denaturing 5% polyacrylamide gel. The left lane contains DS1 without added protein. The right lane contains labelled DS1 bound by proteins of adult chicken immature erythrocyte nuclei. The complexes are indicated on the left and the calculated molecular masses of the protein complexes are indicated on the right.

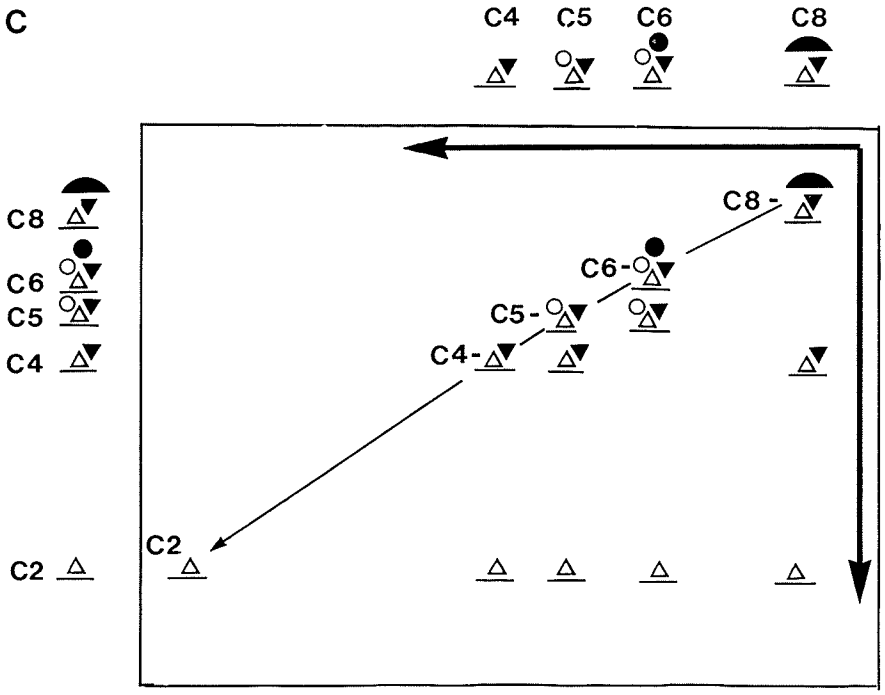

C9
C7
C6
C5
C4
C2
F
M
(kDa)
322
176
150
45

GATA-1. The increase in apparent molecular mass in the EMSA over the reported 39 kDa for GATA-1 resolved in SDS-polyacrylamide gel electrophoresis may be due to DNA-bending by GATA-1 (Schwartzbauer *et al.*, 1992). The binding site for GATA-1 is close to the 3' end of the gene, therefore the DS1 DNA fragment was not shifted as much as it might have been, since shifting in EMSAs by a DNA bending protein is greatest when the factor binding site is near the middle of the DNA sequence. The molecular mass of the DNA-binding protein of complex C1 was approximately 31 kDa. The protein(s) of complexes C4, C5 and C6 had molecular masses of approximately 150 kDa, 176 kDa and 322 kDa respectively. The molecular masses of the DNA-binding proteins of complexes C7, C8, C9 and C10 were greater than 450 kDa. Complexes with bound proteins of approximately 56 kDa (GATA-2) or 55 kDa (GATA-3) were not detected.

The novel diagonal EMSA described by Schaufele *et al.* (1990) was used to determine the protein components of the GATA-related complexes. The labelled 23 bp H5-GATA oligonucleotide was incubated with a nuclear extract isolated from chicken immature erythrocytes, and the sample was electrophoresed for such a time that complex C2 was running off the non-denaturing gel. A lane from the first gel was placed horizontally onto a second non-denaturing gel and electrophoresed. During electrophoresis in the second gel, a portion of each of the complexes dissociated as it entered the second dimension. Therefore complexes remaining on the diagonal were intact, whereas those below the diagonal had dissociated. It should be noted that only those dissociated complexes containing a DNA-binding protein will be detected. From the pattern of dissociation it was possible to determine which complexes were part of a larger complex. Figure 15 shows the

Figure 15. Diagonal EMSA of GATA-1 complexes. The labelled 23 bp oligonucleotide H5-GATA was incubated with nuclear extracted protein of chicken immature erythrocytes. The diagonal EMSA was done as described in Materials and Methods. As a reference, gel slices from the first and second gels are placed on the top and left side of the diagonal gel shift pattern respectively. (a) and (b) are autoradiograms of the same two-dimensional gel pattern, except that the autoradiogram in (b) was exposed for a longer time than in (a). The arrowheads show the locations of the C2 complexes that had dissociated from the complexes C4 - C6 and C8. (c) is a schematic representation of the gel pattern shown in (a) and (b). The symbols represent individual proteins present in each complex. The symbol ▵ represents GATA-1, and '_' shows the DNA fragment. Other symbols are explained in the text.

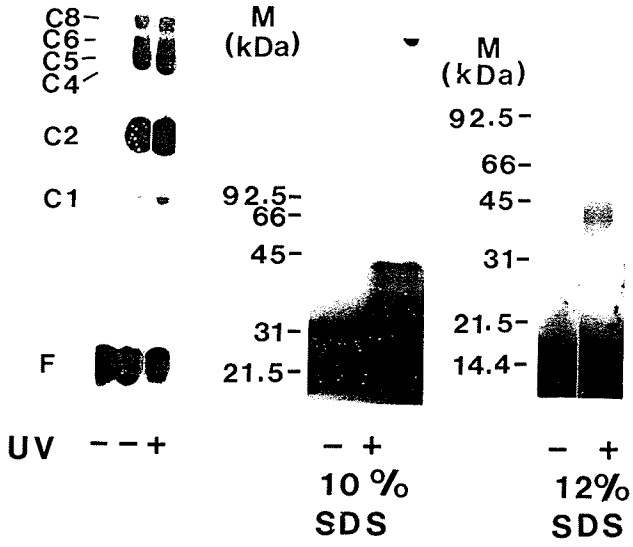

A
C5
C8
C4
C6
C8
C6
C5
C4
C2
C1
F
-C8
-C6
-C5
-C4
-C2
-F
B
C5
C8
C4
C6
-C8
-C6
-C5
-C4
-C2
-F


C
C4
C5
C6
C8
C8
C6
C5
C4
C2
C8-
C6-
C5-
C4-
C2

results of this experiment as well as a schematic representation (Figure 15C) of the gel shown in Figures 15(A) and 15(B). The diagonal EMSA was also done with labelled DS1. The gel-pattern results indicate that the pattern of dissociation of the complexes was as follows: C8 -> C4 and C2; C6 -> C5 and C2; C5 -> C4 and C2; C4 -> C2. Complex C2, the GATA-1-DNA complex, was a common component of complexes C4-C6 and C8 (indicated by arrowheads in Figure 15A and 15B). None of the complexes dissociated to C1, the complex containing the GATA-1 proteolytic product. These results suggest that C4 was also common to the larger complexes C5, C6 and C8. The interpretation of the results shown in Figure 15C was that there were at least four proteins in addition to GATA-1 involved in forming the complexes. GATA-1 (symbol △) was the only DNA-binding protein. Protein(s) A of 105 kDa (symbol▼) interacts with GATA-1, forming complex C4. Protein(s) B of 26 kDa (symbol○) interacted with protein A to generate complex C5. Protein(s) C, which is 146 kDa (symbol●) bound to protein A and/or B to form complex C6, whereas a large protein(s) D (symbol◓) associated with protein A generating complex C8.

The chicken GATA-binding proteins, GATA-1, -2 and -3, have molecular masses of about 39.5, 56, and 55 kDa on SDS-containing gels (Yamamoto *et al.*, 1990). Of these three GATA-binding proteins, GATA-1 is the most abundant in immature erythrocytes (Yamamoto *et al.*, 1990). It was conceivable that some of these complexes arose from the interaction of these different GATA-binding proteins with the GATA-binding site. To elucidate which GATA-binding protein was interacting with the histone H5 enhancer region, a uniformly labelled H5-GATA oligonucleotide was incubated with a nuclear extract isolated from chicken

Figure 16. Ultraviolet light crosslinking of the GATA-1 protein-DNA complexes reveals that the *trans*-acting factor GATA-1 is bound to the histone H5 enhancer region. The internally labelled 23 bp oligonucleotide (H5-GATA) containing the GATA sequence of the histone H5 enhancer region was incubated with the nuclear-extracted protein of chicken immature erythrocytes. The incubation mixtures were (+) or were not (-) irradiated with ultraviolet light and analyzed by EMSA (shown in the left panel). H5-GATA is the free DNA fragment, and C1, C2, C4 - C6 and C8 are the protein-DNA complexes. The protein-DNA complexes were digested with DNase I and micrococcal nuclease, and the affinity-labelled proteins were analyzed by electrophoresis on denaturing 10% or 12% polyacrylamide/SDS gels. Molecular-mass markers (M) are from Amersham. Molecular masses of the proteins were calculated from a $\log_{\text{molecular mass}}$ versus migration plot.

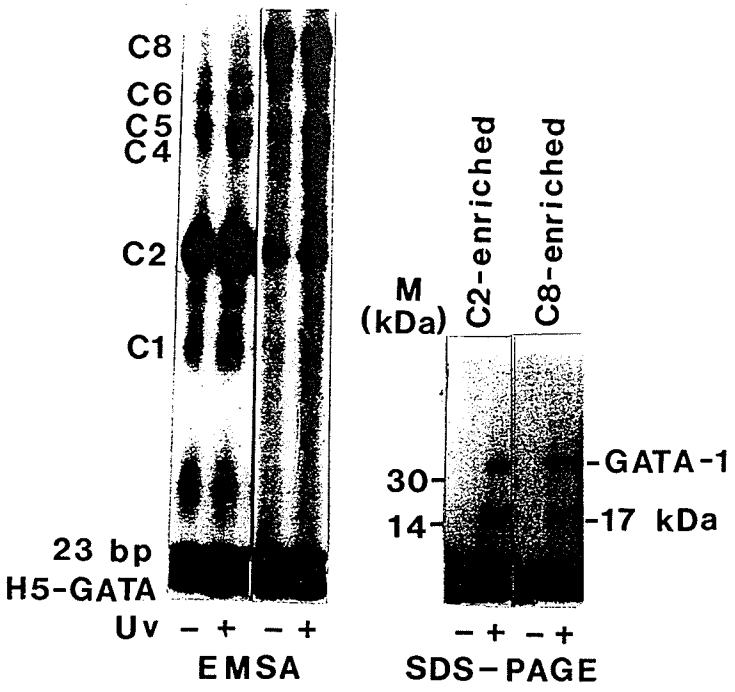

C8-
C6-
C5-
C4
C2
C1
F
UV - - +
M (kDa)
92.5-
66-
45-
31-
21.5-
- +
10 % SDS
M (kDa)
92.5-
66-
45-
31-
21.5-
14.4-
- +
12% SDS

immature erythrocytes, and then the sample was irradiated with ultraviolet light (254 nm). The function of irradiation was to specifically crosslink radioactively labelled nucleotides of a protein-recognition site of DNA with a DNA-binding protein (Ausubel *et al.*, 1987). Purine and pyrimidine free radicals are produced by irradiation of DNA with ultraviolet light. The free radicals interact and form a covalent bond with a protein if it is in close association with the DNA. Pyrimidines are more sensitive to ultraviolet light irradiation than purines and form photoadducts with cysteine, serine, methionine, lysine, arginine, histidine, tryptophan, phenylalanine and tyrosine (Ausubel et al., 1987). Bromodeoxyuridine increases ultraviolet sensitivity. After ultraviolet crosslinking, the sample was incubated with DNase I and micrococcal nuclease to digest non-crosslinked DNA, and the affinity-labelled proteins were identified on SDS-containing gels. Figure 16 shows that proteins of approximately 29 and 43 kDa were labelled. The results were the same for an H5-GATA oligonucleotide containing a labelled template or non-template strand. In a similar experiment using purified GATA-1, Perkins *et al.* (1989) observed that a protein of about 40 kDa was labelled. The 29 kDa labelled protein was likely a proteolytic product of GATA-1 that is sensitive to degradation (Evans and Felsenfeld, 1989; Perkins *et al.*, 1989; Evans and Felsenfeld, 1991). Proteins of molecular mass greater than the 43 kDa GATA-1 protein were not observed, even after longer exposure of the autoradiogram. This result suggested that GATA-1 was the major DNA-binding protein of the GATA-related complexes.

GATA-1 is the Only Protein Binding to DNA

It was of interest to know if GATA-1 was present in each of the complexes.

Figure 17. Ultraviolet light crosslinking of GATA-1 and GATA-1 multisubunit protein-DNA complexes reveals that GATA-1 is present within the GATA-1 multisubunit complexes C2 and C8. The internally labelled H5-GATA oligonucleotide was incubated with adult chicken immature erythrocyte nuclear extracts enriched in C2 and C8 activity. The incubation mixtures were (+) and were not (-) irradiated with ultraviolet light and analyzed by EMSA (left panel). H5-GATA is the free DNA and C1, C2, C4-C6 and C8 are the protein-DNA complexes. The protein-DNA complexes were digested with DNase I and micrococcal nuclease, and the affinity-labelled proteins were analyzed on a 12% polyacrylamide/SDS gel. Molecular mass markers (M) are as indicated. Molecular masses of the proteins were calculated from a $\log_{\text{molecular mass}}$ versus migration plot.


C8
C6
C5
C4
C2
C1
23 bp
H5-GATA
Uv - + - +
EMSA
M
(kDa)
C2-enriched
C8-enriched
30
14
GATA-1
17 kDa
- + - +
SDS-PAGE

The ultraviolet light crosslinking experiment was repeated using an extract enriched in either complex C2 or complex C8 although each of these preparations also had a small amount of the other complexes (Figure 17). These fractions were prepared by dialysis against buffer D and RSB + 0.3 M NaCl respectively as described below. Proteins of 37 and 17 kDa were cross-linked by ultraviolet light with each of the nuclear extract preparations. In 4 replications of this experiment, preparations enriched in either C2 (GATA-1 only) or higher complex activity cross-linked a protein of 37-40 kDa. This included EMSAs that were high in either of the two predominant complex activities, C5 and C8. This indicates that GATA-1 is the sole DNA binding protein of each of the complexes.

Sequence Specificity of GATA-1 Multisubunit Complexes

The GATA-binding sequences found in a number of erythroid genes were synthesized and used in competition studies (Table 1, Figure 18A). Competitions were done at 2.5 molar excess over the labelled H5-GATA sequence which was used for the EMSA. The rho and αG2 oligonucleotides were also used in competitions at 1.25 and 0.625 molar excess. Each of the sequences decreased the binding of GATA-1 and GATA-1 complexes to the H5-GATA sequence (compare to control lane). The strength of competition was alphaG2 > rho, H5-GATA > ß-globin enhancer, ß-globin TATA/GATA. Plumb *et al.* (1989) also observed that the ß-globin (IV) region was less efficient than the αG2 sequence at binding GATA-1. In EMSAs the ß-globin TATA/GATA sequence formed not only the complexes C2, C4, C5, C6 and C8 which is characteristic of GATA-1 binding sites, but also a complex of slightly slower mobility than the complex C2 containing the individual

Table 1. Comparison of sequences used for competition assay. The canonical sequence is based on sequences to which GATA-1 binds and which are found within erythroid genes (Evans and Felsenfeld, 1989). The PCR amplified sequence is an average of the sequences which were amplified by successive rounds of the PCR of oligonucleotides bound by GATA-1 in the EMSA (Engel *et al.*, 1992). The superscript represents the position of the nucleotide relative to the core GAT sequence as defined by Engel *et al.* (1992) and the subscript indicates the relative frequency of a particular nucleotide occurring at a given position. These numbered positions are indicated at the lower edge of the Table. The next sequences are GATA-1 binding sites from selected genes. H5-GATA is the sequence found in the histone H5 gene 3′ enhancer (Rousseau *et al.*, 1989). The ß-globin enhancer (IV) sequence is the first of two GATA-1 binding sites in the 3′ enhancer of the chicken ß-globin gene at DNase I protected region IV (Emerson *et al.*, 1987). The ß-globin promoter TATA/GATA sequence is found in the promoter of chicken ß-globin gene and is a binding site for both TFIID and GATA-1 (Fong and Emerson, 1992). *α*G2 (mouse) is the GATA binding sequence in the promoter of the mouse *α*-globin gene (located at -189 to -174) (Plumb *et al.*, 1989).

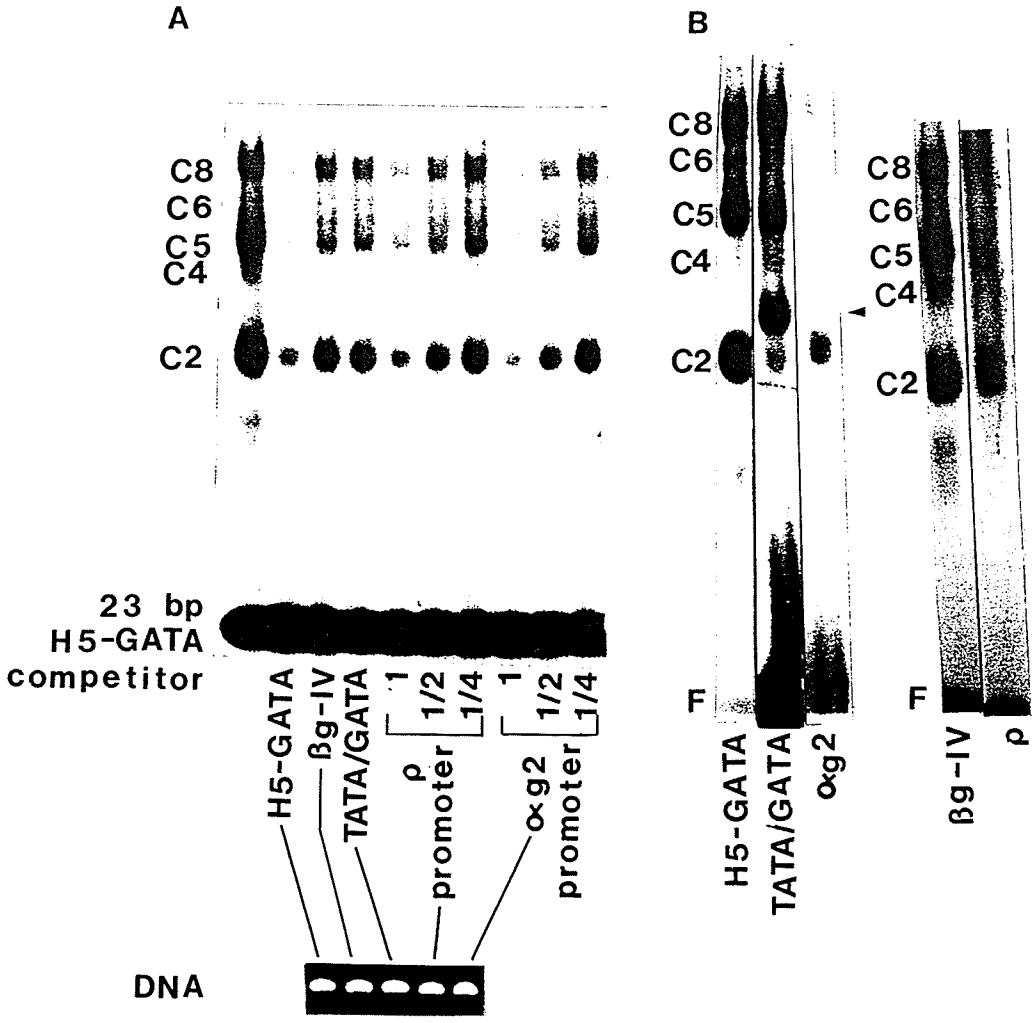
GATA Sequences

Canonical				A/T	GAT	A	A/G		
PCR Amplified‡			$A/C_{0.81}$	$A_{0.63}$	GAT	A	$A_{0.86}$	$C_{0.59}$	$A_{0.59}$
H5-GATA	T	G	G	A	GAT	A	A	C	A
ß-globin enhancer	T	G	C	A	GAT	A	A	A	C
TATA/GATA ß-globin promoter	A	G	G	C	GAT	A	A	A	A
rho-globin promoter	G	C	A	A	GAT	A	A	G	G
*α*G2 (Mouse)	A	A	C	T	GAT	A	A	G	G
position	-4	-3	-2	-1		+1	+2	+3	+4

Figure 18. Interaction of GATA-1 and GATA-1 containing protein complexes with GATA-binding sites found in a number of erythroid genes.

A. Competition by GATA-binding sequences found in a number of erythroid genes. H5-GATA oligonucleotide was labelled and allowed to bind to proteins of adult chicken immature erythrocyte nuclei in the absence and presence of competitor DNA. The left lane does not have competitor DNA added and is a control lane. The oligonucleotides used for competitions are as indicated at the lower edge of each lane and are described in Materials and Methods and Table 1. Each of the competitors were used at 2.5 molar excess over the labelled H5-GATA oligonucleotide (designated 1 for rho and alphaG2) The rho and alpha G2 sequences were also used at 1.25 (designated 1/2) and 0.625 (designated 1/4) molar excess of labelled H5-GATA oligonucleotide. 50 ng of each of the competitor oligonucleotides were also electrophoresed in a 12% polyacrylamide gel (1 X TBE buffer) and the gel was stained with ethidium bromide as shown in the lower portion of the Figure. The oligonucleotides are as indicated.

B. Interactions of labelled GATA-1 binding oligonucleotides with chicken erythrocyte nuclear extracts. EMSAs were performed with the labelled GATA binding site containing oligonucleotides. The ß-globin enhancer (0.5 ng) and rho (0.25 ng) sequences were incubated with 18 μg crude nuclear extract. H5-GATA (0.025 ng) was incubated with 1 μg of crude nuclear extract. The ß-globin TATA/GATA (0.125 ng) sequence was incubated with 10 μg of crude nuclear extract. The alphaG2 (0.25 ng) was incubated with 18 μg of crude nuclear extract. The GATA-1 complexes are labelled. The complex, which is presumed to be a TFIID/adaptor complex, is marked with an arrow. Note that only the H5-GATA and ß-globin TATA/GATA sequences were labelled to a high specific activity. The low specific activity of the other fragments made it difficult to optimize EMSA conditions.


A
C8
C6
C5
C4
C2
23 bp
H5-GATA
competitor
H5-GATA
βg-IV
TATA/GATA
ρ
promoter
1
1/2
1/4
αg2
promoter
1
1/2
1/4
DNA
B
C8
C6
C5
C4
C2
F
H5-GATA
TATA/GATA
αg2
C8
C6
C5
C4
C2
F
βg-IV
ρ

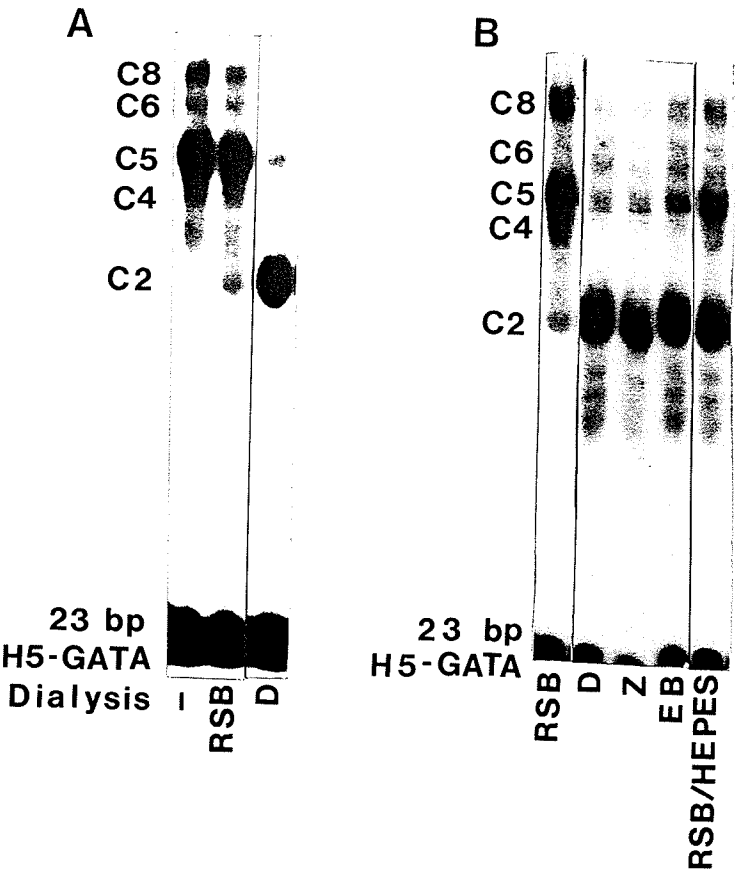
GATA-1 (Figure 18B). Fong and Emerson (1992) showed that a TFIID/adaptor complex bound at the TATA/GATA sequence migrated slightly slower than GATA-1 under their EMSA conditions. It is likely that the complex seen on these shifts is also the TFIID/adaptor complex. Each of the other oligonucleotides used in the study also formed the characteristic GATA-1 complexes C2, C4, C5, C6 and C8, although EMSA conditions were not optimal for most of these sequences (Figures 18B, 12).

Dissociation of GATA-1 Containing Protein Complexes through Dialysis

Other research groups have never reported a physical interaction by GATA-1 with heterologous proteins to form multisubunit protein complexes. It was noted that other investigators use different conditions for preparation of GATA-1. Therefore, the effect of buffers on GATA-1 containing protein complex stability was examined. It was found the GATA-1 containing protein complexes readily dissociate under certain dialysis conditions. Figure 19 shows that complexes C5 and C8 predominated in the EMSA when the H5-GATA-oligonucleotide was incubated with undialysed nuclear extract or extract dialysed against RSB + 0.3 M NaCl. However, complex C2 predominated in the EMSA when H5-GATA was incubated with the nuclear extract that had been dialysed against buffer D. RSB has a pH of 7.5 and buffer D is at pH 7.9. The difference in pH did not affect the dissociation of GATA-1 containing protein complexes to GATA-1 (data not shown). The EMSA patterns generated with DNA fragments DS1a, DS2, US1a and the Sp1 oligonucleotides were the same with nuclear extracts that were or were not dialysed against RSB + 0.3 M NaCl or buffer D (data not shown). Other commonly

Figure 19. Dissociation by dialysis of GATA-1 containing protein complexes formed with an oligonucleotide containing the GATA binding site.

A. An end-labelled oligonucleotide (0.025 ng) containing the H5-GATA motif was incubated in the presence (2 μg) of nuclear proteins extracted from adult chicken immature erythrocytes. A) Nuclear extract was either not dialysed (-), dialysed against RSB + 0.3 M NaCl, 20% glycerol and 1 mM dithiothreitol (RSB) or against buffer D (D). B) Nuclear extract was dialysed against RSB + 0.3 M NaCl, 20% glycerol, 1 mM dithiothreitol (RSB); buffer D (D); buffer Z (Z); equilibration buffer (E.B.) or RSB + 0.3 M NaCl, 20% glycerol, 1 mM dithiothreitol with HEPES instead of Tris buffer (RSB/HEPES) prior to incubation with labelled H5-GATA for use in the EMSA. The buffer D, buffer Z and equilibration buffer are detailed in Materials and Methods. C2, C4, C5, C6 and C8 are the protein-DNA complexes.


A
C8
C6
C5
C4
C2
23 bp
H5-GATA
Dialysis
-
RSB
D
B
C8
C6
C5
C4
C2
23 bp
H5-GATA
RSB
D
Z
EB
RSB/HEPES

used buffers include buffer Z and equilibration buffer (Evans and Felsenfeld, 1989; Emerson and Felsenfeld, 1984). Dialysis against these buffers also results in dissociation of GATA-1 containing protein complexes (Figure 19B). Dialysis against RSB + 0.3 M NaCl, with HEPES instead of Tris as the buffer resulted in partial dissociation of complexes (Figure 19B).

Many researchers (marked ^ in the references) prepare nuclear extracts based on the Dignam (1983b) method. In this method and in a number of other methods (marked * in the references) the proteins are dialysed against buffer D. From the data in Figure 19, the GATA-1 containing protein complexes seen in EMSAs of these preparations are expected to be dissociated. For other extraction methods (marked by ! in references) HEPES was present in at least one of the buffers. Each of buffer D, equilibration buffer and buffer Z included HEPES. In Figure 19B (RSB lane versus RSB/HEPES lane), HEPES in the dialysis buffer caused at least partial dissociation of GATA-1 containing protein complexes. The presence of HEPES together with other differences in buffer conditions were in all likelihood responsible for dissociation of GATA-1 containing protein complexes in the work of the researchers. We did not determine which of the other differences between the various buffers in addition to HEPES was responsible for dissociation of the GATA-1 containing protein complex which formed C5 in the EMSA. For some of the methods which used HEPES buffer, whole cell extracts were used for EMSAs (marked # in the references). We noted that contamination of nuclear extracts with cytoplasm caused partial dissociation of GATA-1 containing protein complexes (data not shown). Therefore the data suggest that other researchers would not have noted GATA-1 multisubunit complexes in the EMSA due to methods and

buffers used for preparation of nuclear extracts.

Developmental and Species Specificity of GATA-1 Complexes

Chicken erythrocytes of stage 28 (days 5-6), stage 38 (day 12) and stage 40 (day 14) embryos were isolated and nuclear extracts of the red blood cells were prepared without dialysis. When these fractions were used in EMSAs, the pattern of GATA-1 containing complexes was similar to that formed by extracts of adult immature erythrocytes. The GATA-1 complexes C2, C4, C5 and C8 were formed by each of the extracts (Figure 20), although the stage 28 extract formed a higher relative amount of C2 than did extract of the stages 38 and 40. This result may be significant, however, the stage 28 extract showed more proteolytic degradation than stages 38 and 40 as evidenced by the higher relative amount of complex C1 in the EMSA. It may be that proteolytic degradation was responsible for the lower relative amounts of C5 and C8 and the higher relative amounts of complex C2 in the EMSA of the stage 38 extract. In any event the data indicate that the GATA-1 multisubunit complexes are present in the developing embryos during the time that the embryonic rho and epsilon-globin genes are expressed. The complexes are also present during the time the A and H ß-globin genes are expressed. This does not indicate a different role for the protein complexes in the embryo when compared to the adult chicken.

Nuclear extracts of uninduced and induced MEL cells were used for EMSAs. A complex, which migrated slightly slower than chicken GATA-1, was seen (Figure 21). The murine GATA-1 is 50 kDa (Whitelaw *et al.*, 1990), and therefore would be expected to migrate slightly slower than the chicken GATA-1. Therefore

Figure 20. Interactions between nuclear proteins extracted from stage 28, stage 38 and stage 40 chicken embryo erythrocyte nuclei. Protein (0.375 μg) extracted from the nuclei of each of the stages was incubated with 0.025 ng of labelled H5-GATA oligonucleotide for use in the EMSA. Complexes are indicated on the left.

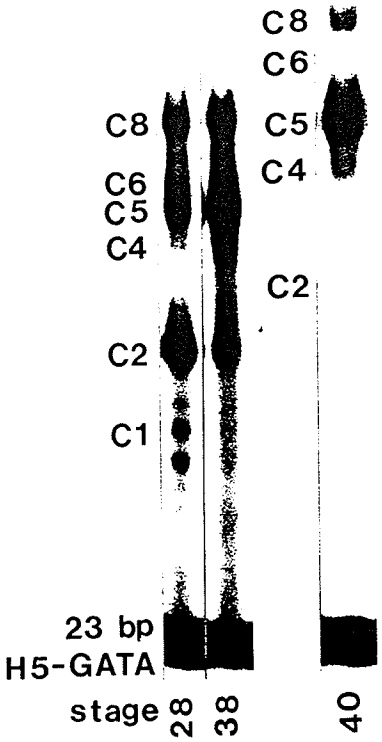
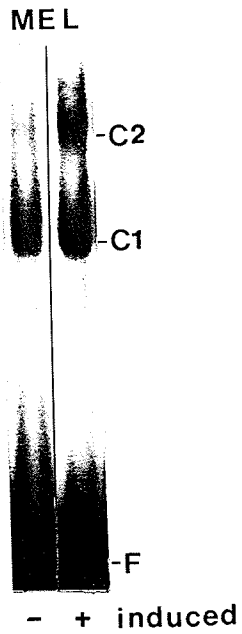

C8
C6
C5
C4
C2
C1
23 bp
H5-GATA
stage 28
38
40

Figure 21. Interactions between labelled H5-GATA oligonucleotide and nuclear proteins extracted from uninduced (-) and induced (+) mouse erythroleukemia cells. The complexes are indicated on the right. Nuclear protein (30 μg of uninduced, 10 μg of induced) was incubated with 0.2 ng of DNA in the presence of 1 μg of poly dI.dC.


MEL
-C2
-C1
-F
- + induced

complex C1 of the MEL extracts was likely the murine GATA-1. A higher complex(es) was seen in addition to the hypothesized GATA-1 containing complex C1. In this assay, we were not able to determine complexes which corresponded to the complexes formed on H5-GATA using nuclear proteins extracted from chicken erythrocytes. The MEL complex C2 may well correspond to one of the complexes produced by the chicken erythrocyte extracts. If the complexes are homologous, the finding could be indicative of similarity of GATA-1 function between different species, despite the dissimilarities between GATA-1 proteins among species. It should be noted that these results are from a preliminary experiment.

Partial Purification of GATA-1 Containing Protein Complexes

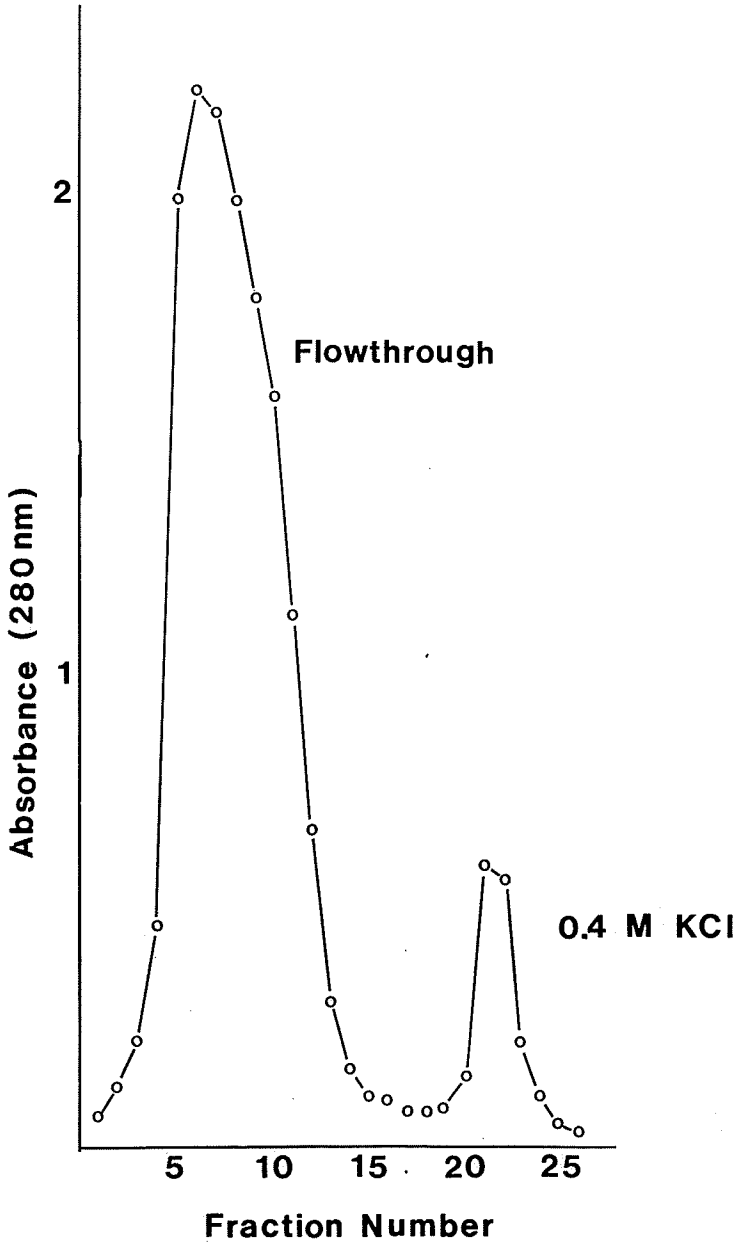
The aim of this part of the project was to purify the proteins which form complex C5 in the EMSA, so that subsequently the proteins of the complex could be separated into its component parts and then reconstituted. Complex C5 was chosen for purification because from the data (Figure 15) it is expected to only contain a small number of proteins in addition to GATA-1, and C5 is more abundant than complex C4, the next simplest GATA-1 containing protein complex. Complex C4 contains the protein(s) of 105 kDa in addition to GATA-1, and complex C5 contains these proteins and additionally a protein(s) of 26 kDa. Chromatographic columns that were used included heparin-agarose, S-sepharose, Q-sepharose, phenyl-agarose, double-stranded DNA cellulose, single-stranded DNA cellulose and a GATA affinity column based on the GATA-binding sequence of the chicken ß-globin gene enhancer.

Both the heparin-agarose and S-sepharose columns have a high negative charge and thus enriched proteins with positive charge. GATA-1 is a basic protein (+12) and thus both GATA-1 and GATA-1 containing protein complexes have affinity for these columns. The phenyl-agarose column selected for proteins on the basis of hydrophobicity. This property was useful for 2 reasons. It separated the complex C5 from the other complexes and from the decreased requirement for poly dI.dC in the EMSA, it was apparent the column reduced the amount of non-specific DNA binding proteins in the fractions with GATA-1 and GATA-1 containing protein complexes. The double-stranded and single-stranded DNA cellulose columns selected for proteins with an affinity for DNA in general. DNA-binding proteins have varying affinities for either double- or single- stranded DNA, and therefore each of these columns was useful. The GATA affinity column which consisted of concatamers of GATA-binding sequences was designed to select for protein based on a specific affinity for the GATA binding site. The Q-sepharose column was also used. This column selects for proteins with negative charge. Most of the GATA-1 containing protein complexes had little affinity for Q-sepharose, and were displaced by 0.1 M NaCl, although a small portion of the C5 complex was displaced at 0.4 M NaCl.

Heparin-Agarose Chromatography

Typically, crude nuclear extract was chromatographed first on a heparin-agarose column. After the protein was applied to the heparin-agarose column, the column was washed with buffer D containing 0.2 M KCl, and GATA-1 containing protein complexes eluted in the next wash at 0.4 M KCl (Plumb *et al.*, 1989). An

Figure 22. Protein elution profile of heparin-agarose column. Crude nuclear extract (240 mg) in 60 ml of buffer D with a final concentration of 0.2 M KCl was applied to a 60 ml heparin-agarose column. The column was eluted sequentially with buffer D containing 0.2 (150 ml) and 0.4 (80 ml) M KCl. Fractions of 10 ml were collected. Note that the flowthrough and 0.2 M KCl fractions occur together, since the protein was applied in this concentration of KCl. The fraction numbers are indicated at the lower edge of the plot and the scale for the absorbance at 280 nm is indicated at the left edge of the plot.


Absorbance (280 nm)
2
1
Flowthrough
0.4 M KCl
5
10
15
20
25
Fraction Number

elution profile of the heparin-agarose column is shown in Figure 22. Note that the flowthrough and 0.2 M KCl together form the protein peak labelled flowthrough. Figure 25B demonstrated that the column does enrich for GATA-1 and GATA-1 containing protein complexes, since 0.5 μg of protein from the starting crude nuclear extract formed a smaller amount of GATA-1 and GATA-1 containing protein complexes with labelled H5-GATA than did the same amount of protein from the heparin-agarose column (also see Figure 28A, compare nuclear extract to heparin-agarose). Approximately 10% of the starting protein was present in the 0.4 M KCl heparin-agarose fraction. In order to determine enrichment of protein, a filter binding assay could be employed. In the filter binding assay, protein and radioactively labelled DNA are allowed to interact. DNA bound by protein is retained on the filter and the radioactivity can be measured, in order to determine the amount of DNA binding activity relative to protein present in the assay. This assay would be useful for determining the enrichment of the GATA-1 containing protein complexes, although it would need to be done in conjunction with EMSAs which would show the particular GATA-1 containing protein complex which was enriched in the fraction.

S-Sepharose Chromatography

Chromatography of the S-sepharose column was done with the 0.4 M KCl heparin-agarose fraction. After application of the protein to the S-sepharose column, proteins were eluted with a step gradient. An elution profile of the S-sepharose column is shown in Figure 23. Note that the flowthrough and 0.2 M NaCl together formed the protein peak labelled flowthrough. Figure 24 shows that

Figure 23. Protein elution profile of S-sepharose column. Crude nuclear extract which had been partially purified by passage over a heparin-agarose column (0.4 M KCl fraction, 40 mg) was diluted to 1/2 of the initial concentration with buffer E without NaCl and applied in 150 ml of buffer E to a 7.5 ml S-sepharose column. The column was eluted sequentially with buffer E containing 0.2 (30 ml), 0.3 (20 ml) and 0.5 (20 ml) M NaCl. Fractions of 5 ml were collected. Note that the flowthrough and 0.2 M NaCl fractions occur together, since the protein was applied in this concentration of KCl. The fraction numbers are indicated at the lower edge of the plot and the scale for the absorbance at 280 nm is indicated at the left edge of the plot.

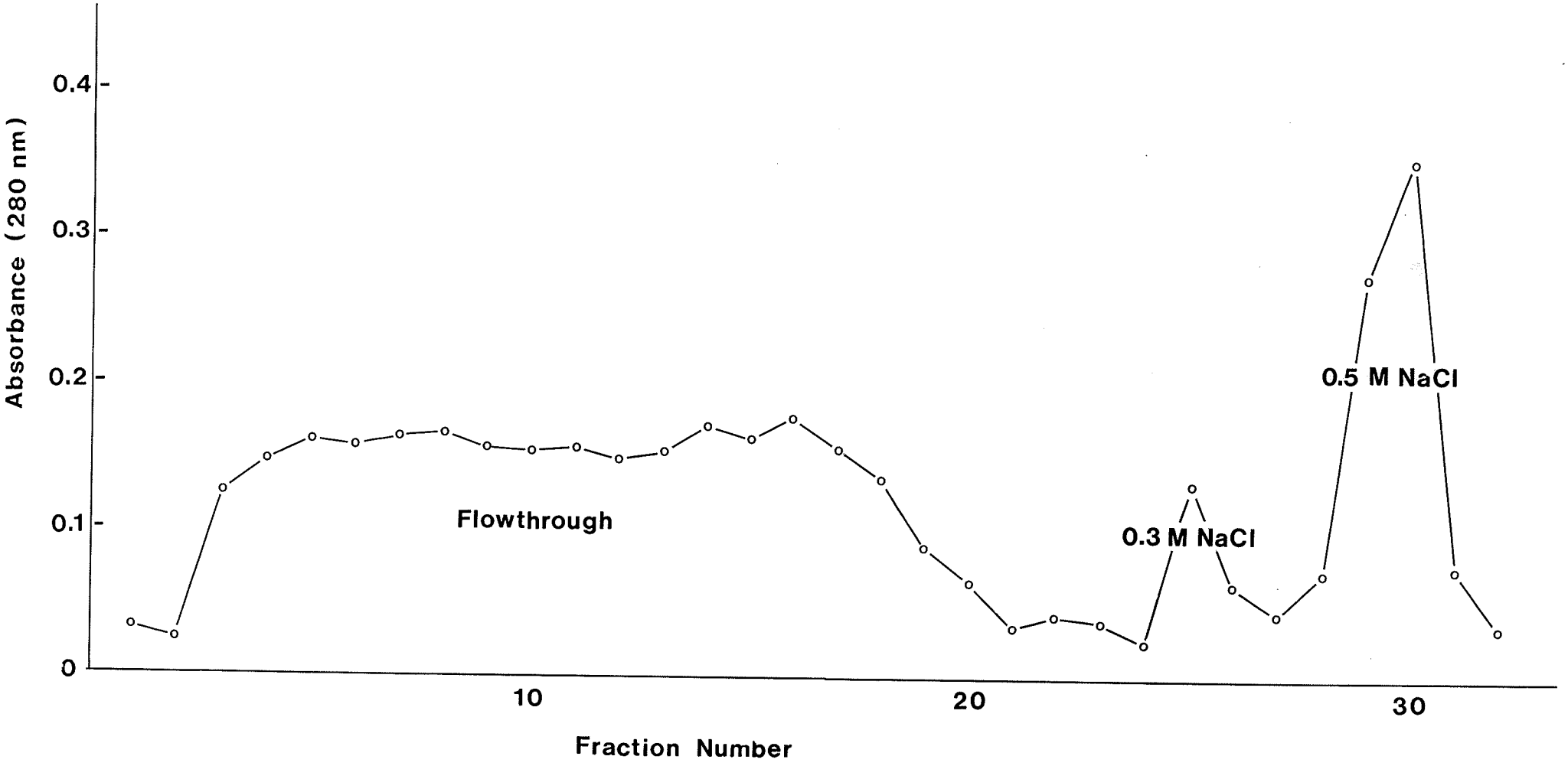

Absorbance (280 nm)
0.4
0.3
0.2
0.1
0
Flowthrough
0.3 M NaCl
0.5 M NaCl
10
20
30
Fraction Number

Figure 24. EMSAs of proteins eluted from the S-Sepharose column. The 0.4 M KCl fraction of a heparin-agarose column chromatography was applied to an S-Sepharose column and eluted with increasing concentrations of NaCl (0.2, 0.3, 0.5 M). EMSAs of the proteins of the heparin-agarose column fraction (hep-ag), flowthrough (F.T.), 0.2, 0.3 and 0.5 M NaCl fractions are as indicated at the lower edge of the lanes. Five μl of each fraction (3 μg, heparin-agarose; 0.3 μg, 0.3 M NaCl, S-Sepharose; 2.6 μg, 0.5 M NaCl, S-Sepharose) were used for the EMSAs. Note that there is no 0.4 M NaCl fraction.

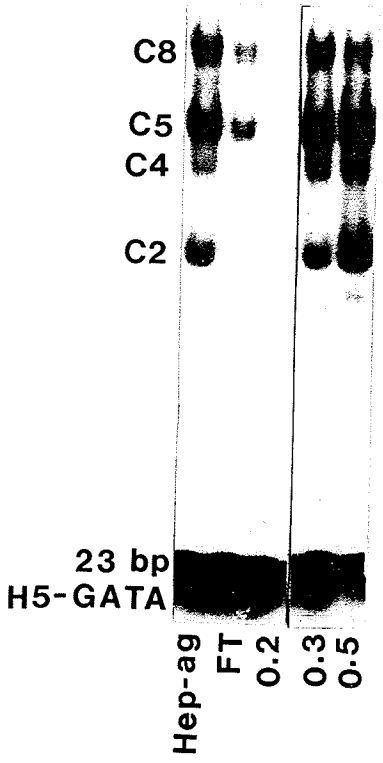
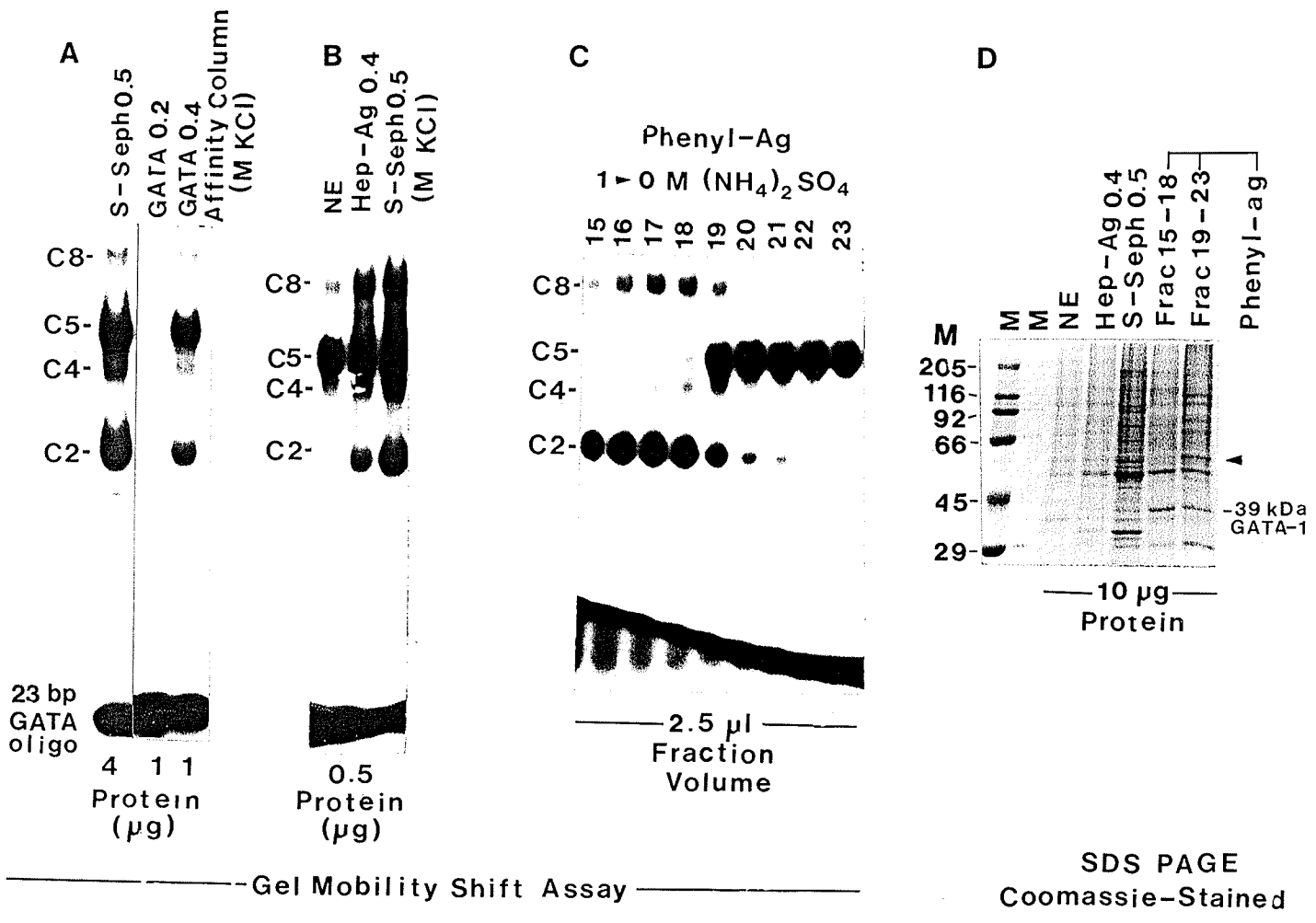

C8
C5
C4
C2
23 bp
H5-GATA
Hep-ag
FT
0.2
0.3
0.5

Figure 25. Partial purification of GATA-1 containing protein complexes. Panel A) GATA-affinity column chromatography of GATA-1 containing protein complexes. The 0.5 M NaCl fraction from an S-Sepharose column (derived from 0.4 M KCl fraction of a heparin-agarose column) was applied to a GATA-affinity column. The column was washed with 0.2 M KCl and GATA-1 activity eluted at 0.4 M KCl. Protein from the S-Sepharose (S-seph) fraction was used in EMSAs with labelled H5-GATA oligonucleotide as was protein from the 0.2 M KCl wash and 0.4 M KCl elution from the GATA affinity column as indicated in the Figure. Panel B) Crude nuclear extract (NE) of adult chicken immature erythrocyte was chromatographed over a Heparin-agarose (Hep-ag) column. The 0.4 M KCl fraction contained GATA-1 activity and this fraction was chromatographed over an S-Sepharose column (S-seph). Protein (0.5 μg) of the crude nuclear extract (N.E.), 0.4 M KCl heparin-agarose fraction and 0.5 M NaCl fraction of the S-Sepharose profile were used in EMSAs with labelled H5-GATA oligonucleotide. Panel C) The 0.5 M NaCl fraction from the S-Sepharose column was chromatographed over the phenyl-agarose column. Samples (2.5 μl) from each of the active phenyl-agarose (phenyl-ag) fractions were used in EMSAs with labelled H5-GATA oligonucleotide as indicated. Panel D) 10 μg of protein from the crude nuclear extract, 0.4 M KCl heparin-agarose fraction, 0.5 M NaCl of the S-Sepharose fraction and fraction 18 (C2 enriched) and fraction 23 (C5 enriched) of the phenol-agarose column were TCA precipitated and separated on a 12% polyacrylamide/SDS gel. The molecular masses of the marker proteins are indicated on the left. The position of the 39 kDa GATA-1 is indicated. An arrowhead marks the position of a 54 kDa protein which was enriched in GATA-1 containing protein complex containing fractions.


A
S-Seph0.5
GATA 0.2
GATA 0.4
Affinity Column (M KCl)
C8-
C5-
C4-
C2-
23 bp GATA oligo
4 1 1
Protein (μg)
B
NE
Hep-Ag 0.4
S-Seph0.5 (M KCl)
C8-
C5-
C4-
C2-
0.5
Protein (μg)
C
Phenyl-Ag
1 → 0 M $(NH_4)_2SO_4$
15 16 17 18 19 20 21 22 23
C8-
C5-
C4-
C2-
2.5 μl
Fraction Volume
D
M
M
NE
Hep-Ag 0.4
S-Seph 0.5
Frac 15-18
Frac 19-23
Phenyl-ag
205-
116-
92-
66-
45-
29-
39 kDa GATA-1
10 μg
Protein
Gel Mobility Shift Assay
SDS PAGE
Coomassie-Stained

the 0.2 M NaCl step did not have any GATA-1 containing protein complexes interacting with the H5-GATA oligonucleotide in the EMSA. GATA-1 containing protein complexes eluted at 0.3 M NaCl and 0.5 M NaCl. The EMSA shows that the 0.3 M NaCl (0.3 μg protein) and 0.5 M NaCl (2.6 μg protein) fractions were enriched in GATA-1 and GATA-1 containing protein complexes over the heparin-agarose column (3 μg protein). Less C2 complex was seen in the EMSA using the 0.3 M NaCl fraction, a higher amount of C8 complex and similar or less C4/C5 complex when compared to the 0.5 M NaCl fraction. The 0.5 M NaCl fraction formed a larger amount of C2 complex in the EMSA. SDS-polyacrylamide gel electrophoresis shows that the protein composition was complex in the 0.5 M NaCl fraction (Figure 25D, S-sepharose). The 0.5 M NaCl fraction was used for most further fractionations, although it was sometimes pooled with the 0.3 M NaCl fraction. The 0.3 M NaCl fraction alone was not used successfully for further fractionations. An SDS-polyacrylamide gel electrophoresis of the 0.3 M NaCl fraction is not available. Future fractionations using the 0.3 M NaCl S-sepharose fraction might be useful for purification of the GATA-1 containing protein complexes.

GATA-Affinity Chromatography

The GATA-affinity column was also used for protein purification (Figure 25). Partially purified protein (0.5 M NaCl fraction of S-sepharose) was applied to the column and eluted at 0.2 M KCl and 0.4 M KCl as described in Materials and Methods. GATA-1 and GATA-1 containing protein complexes eluted at 0.4 M KCl. A 0.3 M KCl step was not included because it was found that GATA-1 and GATA-1

containing protein complexes eluted in this fraction. Protein complexity did not decrease significantly through use of this column, although a 54 kDa protein was seen in the GATA-1 containing protein complex enriched fractions (data not shown). Had a 0.3 M NaCl wash step been included, perhaps a lot of non-specific proteins would have been eliminated. Recent research by Ko and Engel (1993) and data presented in this thesis (see above) showed the sequence used for the column was not optimal for GATA-binding. If a sequence with a higher affinity for GATA-1-binding activity were used, a 0.3 M NaCl wash might be employed and result in purification of the GATA-1 containing protein complex.

Phenyl-Agarose Chromatography

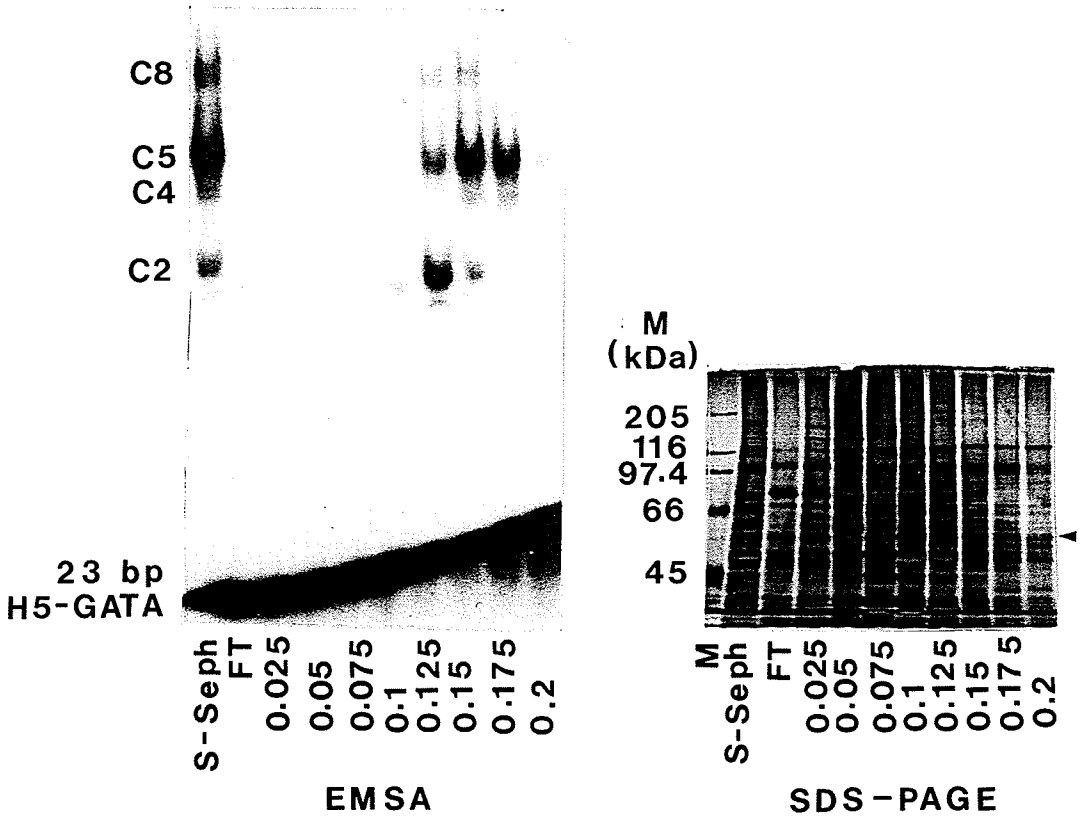
Partially purified protein from either the heparin-agarose or S-sepharose column (heparin-agarose first, Figure 25B) was applied to a phenyl-agarose column in 1 M ammonium sulphate (Figure 25C). The protein was also applied in 0.7 M ammonium sulphate and this worked well. The advantage of the 0.7 M ammonium sulphate is that less proteins interact with column, which leaves more matrix available for the GATA-1 containing protein complexes to bind. The column was developed with a decreasing gradient of ammonium sulphate. A mixture of complexes C2 and C8 were seen in the EMSA of proteins which eluted in fractions 15 to 18, and the concentration of ammonium sulphate in this pooled fraction was 0.38 M. For a number of different chromatographic separations over the phenyl-agarose column complexes C2 and C8 formed in the EMSA using fractions that eluted from 0.45 to 0.32 M ammonium sulphate in about 5% of the starting protein (Heparin agarose and S-sepharose column fractionation first, Figure 25). When

fractions 19 to 23 were used for the EMSA, complex C5 was more abundant than using other fractions which eluted from the phenyl-agarose column. The pooled fractions 19 to 23 had a concentration of 0.15 M ammonium sulphate. Complex C5 was characteristically formed in the EMSA using fractions which eluted at 0.23 to 0.083 M ammonium sulphate in about 5% of the starting protein (Heparin agarose and S-sepharose first). Complexes C4 and C6 were minor components in the EMSAs. Complex C4 was seen in the EMSA in fractions which were eluted between at the fractions which formed complex C2 and complex C5 in the EMSA (Figure 25C). Typically this was about 0.3 M ammonium sulphate. Thus, the column was useful for separating proteins forming complex C5 in the EMSA from proteins forming other GATA-1 multisubunit complexes in the EMSA (see also Figure 28, heparin-agarose versus phenyl-agarose). The column also decreased the requirement for non-specific competitor DNA poly dI.dC in the EMSA (data not shown), indicating a significant decrease in non-specific DNA-binding proteins. When the proteins of the fractions which contained GATA-1 and GATA-1 containing protein complexes were size-fractionated by SDS-polyacrylamide gel electrophoresis, the pattern showed a protein of 39 kDa, which corresponds in size to GATA-1. Another protein seen consistently with fractions which formed complex C5 in the EMSA was a 54 kDa protein (marked by an arrowhead, Figure 25). SDS gels with a higher percentage of polyacrylamide would be useful for separating proteins of smaller size such as the 26 kDa protein which the data suggest is present in complexes C5 and C6.

Double-Stranded DNA Cellulose Chromatography

Chromatography of partially purified extracts over a double-stranded DNA cellulose was employed in order to attempt to purify GATA-1 containing protein complexes. A 0.4 MKCl fraction from a heparin-agarose column enriched in GATA-1 containing protein complexes of adult immature erythrocyte crude nuclear extract, was chromatographed over an S-Sepharose column. The 0.5 M NaCl fraction was applied to a double-stranded DNA cellulose column. The double-stranded DNA cellulose column was eluted with a step gradient of ammonium sulphate in increments of 0.025 M as described in Materials and Methods (Figure 26). More than 50% of the protein eluted in the flowthrough and in the 0.05 M ammonium sulphate fraction. The 0.125 M ammonium sulphate fraction formed complexes C2 and some C5 in the EMSA. Complex C5 was seen maximally in the EMSA in fractions which eluted at 0.15 and 0.175 M ammonium sulphate in about 10% of the protein. The EMSA demonstrates that the amount of complex formed by the 0.125 (0.17 μg protein), 0.15 (0.19 μg protein) and 0.175 (0.14 μg protein) M ammonium sulphate fractions was only slightly less than that of the S-sepharose fraction (25 μg protein), demonstrating an enrichment of GATA-1 and GATA-1 containing protein complexes which is estimated to have been up to 50 fold. The SDS-polyacrylamide gel electrophoresis revealed that the protein fractions from this column were complex. At the concentration of polyacrylamide used for the gel (10%), proteins of 39 kDa (GATA-1) run close to the dye front and therefore were not separated and detected. Due to the protein complexity, it could not be determined if the 54 kDa protein was enriched (arrowhead, Figure 26).

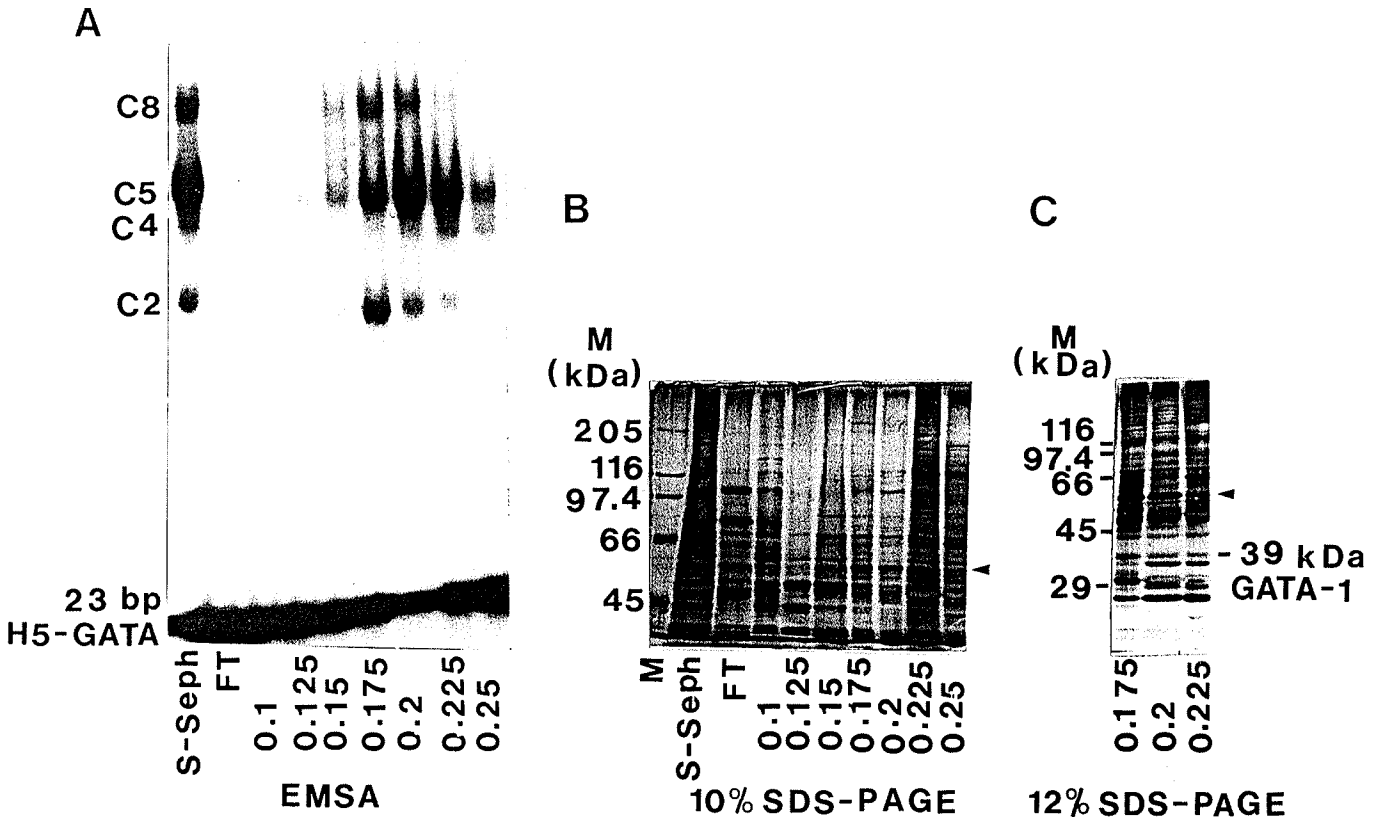
Figure 26. Double-stranded DNA-cellulose column chromatography of GATA-1 and GATA-1 containing protein complexes. The 0.4 M KCl fraction of heparin-agarose column fractionated crude nuclear extract of adult chicken immature erythrocytes was fractionated over an S-Sepharose column. The pooled 0.3 and 0.5 M NaCl fraction from this column was applied to a double-stranded DNA-cellulose column as described in Materials and Methods and eluted with increments of 0.025 M ammonium sulphate. Panel A shows EMSAs of labelled H5-GATA oligonucleotide using 5 μl (0.14 to 0.2 μg of 0.05 to 0.175 M ammonium sulphate, 0.035 μg of 0.025 and 0.2 M ammonium sulphate, and 0.28 μg of the flowthrough fraction) of each of the fractions of the double-stranded DNA column profile and 2.5 μl (25 μg) of the S-Sepharose (S-seph) starting fraction. Panel B shows a silver stained 10% polyacrylamide gel of 10 μg of each of the fractions. Molecular masses of the standards are indicated on the left. The arrowhead indicates 54 kDa. Since the gel was 10% polyacrylamide, GATA-1 ran too close to the dye front to be noted.


C8
C5
C4
C2
23 bp
H5-GATA
S-Seph
FT
0.025
0.05
0.075
0.1
0.125
0.15
0.175
0.2
EMSA
M
(kDa)
205
116
97.4
66
45
M
S-Seph
FT
0.025
0.05
0.075
0.1
0.125
0.15
0.175
0.2
SDS-PAGE

Single-Stranded DNA Cellulose Chromatography

Chromatography of partially purified extracts over a single-stranded DNA cellulose column was employed in order to attempt to purify GATA-1 containing protein complexes. The GATA-1 containing protein complexes of adult immature erythrocyte crude nuclear extract, which had been chromatographed over a heparin-agarose column, was chromatographed over an S-Sepharose column. The pooled 0.3 and 0.5 M NaCl fractions were applied to a single-stranded DNA cellulose column. The single-stranded DNA cellulose column was eluted with a step gradient of KCl. About 50% of the protein eluted in the flowthrough and 0.1 M KCl. EMSAs were done with the starting S-sepharose fraction (25 μg) and each of the eluted fractions. Some of each of the GATA-1 containing protein complexes eluted at 0.175 M KCl (0.15 μg protein; Figure 27). A large proportion of the complex C5 was seen in the EMSA using the fraction which eluted at 0.2 M KCl and somewhat less was seen in the 0.225 M KCl fraction (0.15 μg, 0.05 μg protein, respectively). A comparison of the amount of protein present in the EMSA and the amount of activity of each of the fractions compared to the S-sepharose starting material for the column demonstrated that the fractions with GATA-1 and GATA-1 containing protein complexes were enriched. The GATA-1 containing protein complex enriched fractions were enriched in a 54 kDa protein (0.175 to 0.225 M KCl in Figure 27).

Figure 27. Single-stranded DNA-cellulose column chromatography of GATA-1 and GATA-1 containing protein complexes. The 0.4 M KCl fraction of heparin-agarose column fractionated crude nuclear extract of adult chicken immature erythrocytes was fractionated over an S-Sepharose column. The pooled 0.3 and 0.5 M NaCl fractions from this column were applied to a single-stranded DNA-cellulose column as described in Materials and Methods and eluted with increments of 0.025 M KCl. Samples of protein from the column fractions were used for EMSAs and size fractionated on SDS polyacrylamide gels. Panel A shows EMSAs using 5 μl (0.45 to 0.15 μg) of each of the fractions from the single-stranded DNA cellulose column and 2.5 μl (25 μg) from the S-Sepharose fraction. Panel B shows a silver stained gel (10% polyacrylamide) of 10 μg of each of the fractions. Panel C shows a silver stained gel (12% polyacrylamide) of 10 μg of the fractions 0.175, 0.2 and 0.225 M KCl. The arrowhead indicates 54 kDa. The position of GATA-1 (39 kDa) is indicated.

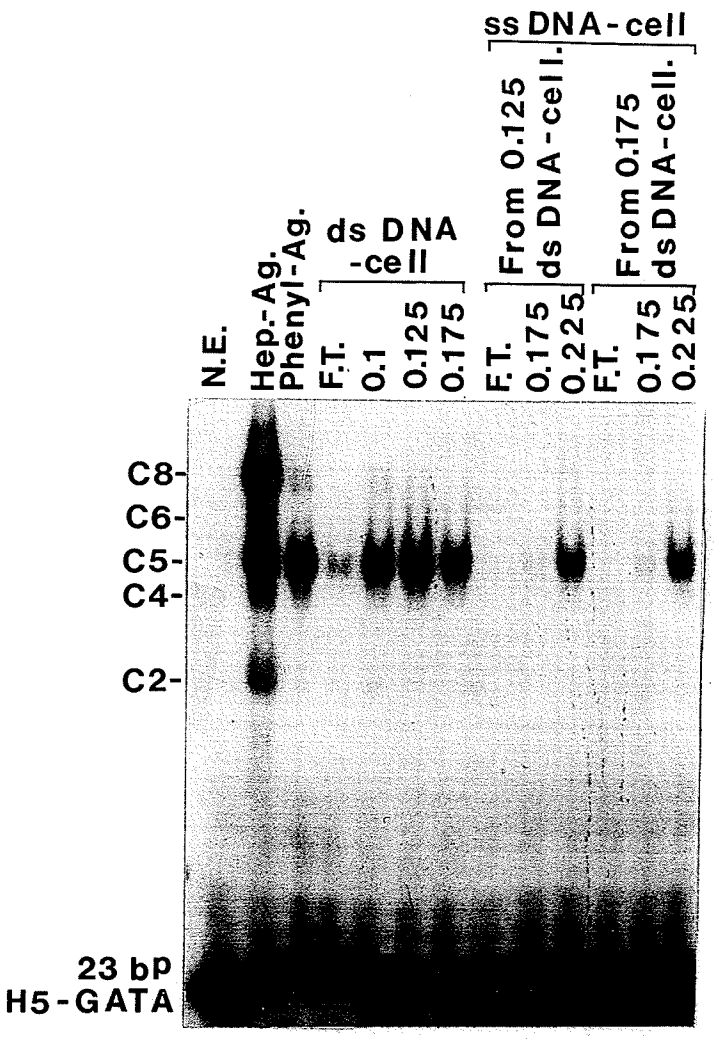

A
C8
C5
C4
C2
23 bp
H5-GATA
S-Seph
FT
0.1
0.125
0.15
0.175
0.2
0.225
0.25
EMSA
B
M
(kDa)
205
116
97.4
66
45
M
S-Seph
FT
0.1
0.125
0.15
0.175
0.2
0.225
0.25
10% SDS-PAGE
C
M
(kDa)
116
97.4
66
45
29
39 kDa
GATA-1
0.175
0.2
0.225
12% SDS-PAGE

Fractionation of Nuclear Extract over Heparin-Agarose, S-Sepharose, Phenyl-Agarose, Double-Stranded DNA Cellulose and Single-Stranded DNA Cellulose Columns

In another series of column chromatographs, crude nuclear extract was fractionated over a heparin-agarose column (0.4 M KCl fraction), a phenyl-agarose column (0.23 to 0.083 ammonium sulphate, GATA-1 containing protein complex C5 enriched), followed by double-stranded and single-stranded DNA cellulose column chromatography as described in Materials and Methods. The EMSAs for this series of column chromatographs is shown in Figure 28. The starting crude nuclear extract had very little GATA binding activity in 0.5 μg. The heparin-agarose fraction had significantly more GATA binding activity in 0.5 μg protein. The phenyl-agarose column separated the proteins forming complex C5 in the EMSA from those forming the other GATA-1 complexes in the EMSA.

Each of the double-stranded DNA cellulose fractions 0.1, 0.125 and 0.175 M ammonium sulphate had GATA-1 containing protein complex activity, contrary to the data shown in Figure 26, where the 0.1 M ammonium sulphate fraction had little GATA-1 and GATA-1 multisubunit complex activity forming in the EMSA. It should be noted that in Figure 26 the series of columns included a fractionation over an S-sepharose column, whereas this series of column chromatographs did not. A number of EMSA profiles of column fractionations have demonstrated that GATA-1 containing protein complexes elute over more than one salt concentration. Examples are the S-sepharose column fractionation, where GATA-1 containing protein complexes eluted at 0.3 and 0.5 M NaCl, both the double and single-stranded DNA cellulose columns, and as will be shown in Figure 29, GATA-1

Figure 28. EMSA profile of GATA binding factors eluting sequentially from a number of columns. GATA-1 and GATA-1 containing protein complexes present in crude nuclear extract (N.E.) were eluted sequentially from a heparin-agarose (hep-ag) column, phenyl-agarose column (phenyl-ag), double-stranded DNA cellulose (ds DNA cell) column and single-stranded DNA cellulose (ss DNA cell) column. Fractionations were as described in Material and Methods. GATA-1 multisubunit complex activity (C5) from the phenyl-agarose (25 mg) was applied to a 4 ml double-stranded DNA cellulose column, which was then eluted with 12 ml at each of the ammonium sulphate concentrations. The double-stranded DNA cellulose fractions were diluted 2 X with buffer and applied to a 2 ml single-stranded DNA cellulose column. The column was then eluted with 6 ml at each of the KCl concentrations. Protein (0.5 μg) of the crude nuclear extract, heparin-agarose and phenyl-agarose fractions were used in the EMSA. Five μl of each of the double-stranded DNA cellulose and single-stranded DNA cellulose fractions were used for the EMSA.


ss DNA-cell
From 0.125 ds DNA-cell.
From 0.175 ds DNA-cell.
ds DNA -cell
N.E.
Hep.-Ag.
Phenyl-Ag.
F.T.
0.1
0.125
0.175
F.T.
0.175
0.225
F.T.
0.175
0.225
C8-
C6-
C5-
C4-
C2-
23 bP
H5-GATA

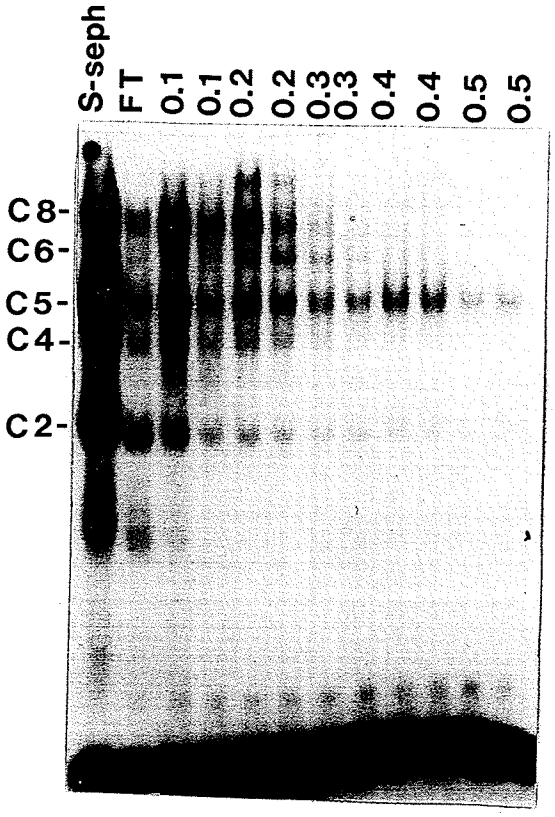
containing protein complexes eluted over a range of NaCl concentrations from the Q-sepharose column. These results are complicated by dissociation of GATA-1 containing protein complexes to their constituent components in some of the buffers used, although we tried to keep this to a minimum by working quickly with fractions which were not stable in a given buffer. Protein purification was not achieved. The 54 kDa protein was enriched after the final columns (data not shown).

Q-Sepharose Column Chromatography

Crude nuclear extract of adult chicken immature erythrocytes was also applied to and eluted from a Q-sepharose column. It should be noted that since the S-sepharose fraction was diluted 5 X prior to applying to the Q-sepharose column, the EMSA for the flowthrough fraction does not reflect the amount of activity present in the fraction, since equal volumes of each fraction were used for the EMSA. Therefore the flowthrough fraction had significant GATA-1 and GATA-1 containing protein complexes. Fractions were eluted batchwise from the column in increments of 0.1 M NaCl.

Much of the GATA binding activity eluted at 0.1 M NaCl including most of the monomeric GATA-1 (complex C2, Figure 29), a smaller amount of all of the GATA-1 containing complexes eluted at 0.2 M NaCl. There appeared to be an enrichment in the proteins forming C6 and a complex larger in size than C8 in the 0.2 molar NaCl fraction. A small amount of proteins forming complex C5 in the EMSA eluted at 0.4 M NaCl. This demonstrates that some of the GATA-1 containing protein complexes, especially the complex forming C5 in the EMSA

Figure 29. EMSA profile of GATA binding factors eluting from a Q-sepharose column. The 0.5 M NaCl fraction of a S-sepharose column was diluted 5 X with buffer (20 mM Tris pH 8.0, 20% glycerol, 0.2 mM EDTA) and applied to a Q-sepharose resin. The resin was eluted batchwise with 2 X 2.5 buffer volumes of the column at each of the NaCl concentrations 0.1, 0.2, 0.3, 0.4 and 0.5 M NaCl. The EMSAs with 5 μl of each of these fractions and the starting S-sepharose fraction are shown.

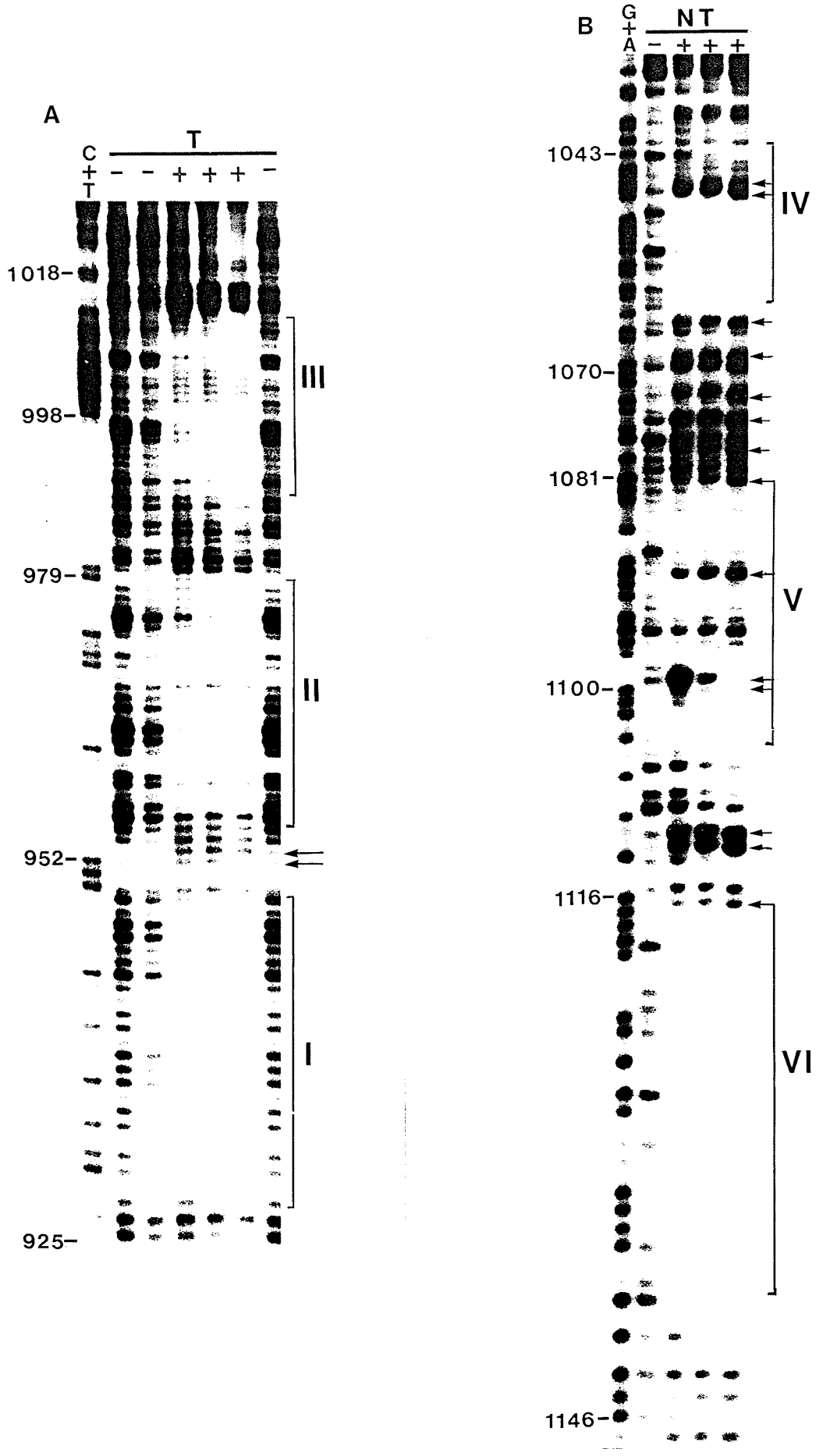

S-seph
FT
0.1
0.1
0.2
0.2
0.3
0.3
0.4
0.4
0.5
0.5
C8-
C6-
C5-
C4-
C2-

consisted of a population of complexes with varying affinities for the column. Variations in the amount of post-translational modifications to the proteins present in the complexes could account for the varying affinities for the columns. It should be noted that GATA-1 was found not to be post-translationally modified, therefore the proteins interacting with GATA-1 may be expected to have the modifications (Evans *et al.*, 1988). This column was not used for further fractionations.

DNase I Protection Assays of DS1 and DS2

The chicken H5 3′ enhancer was localized in the region +851 to +1064 (Figure 6, Trainor *et al.*, 1987). Proteins other than GATA-1 formed protein-DNA complexes with DS1 and these proteins bound in the region +893 to +980. The binding sites of the adult chicken immature erythrocyte nuclear proteins were identified in DNase I protection assays of DS1 and DS2. Figure 30A and 30B show the DNase I protection sites generated by proteins binding to DNA sequences present in DNA fragments DS1a and DS1 + DS2, respectively. Three DNase I protected regions were observed with DNA fragment DS1a with the level of protection being I>II>III. DNase I protected regions I and II contained potential Sp1 binding sites (TGGGC^{+939}GGGT and TAGGA^{+966}GGGC), and these DNase I protected regions covered approximately 22 and 24 bp, respectively, which was similar to protection generated by human Sp1 (Jones *et al.*, 1985; Jones and Tjian, 1985). Both of these Sp1 binding sites were present in the region spanned by DNA fragment DS1a-U. Interestingly, DNase I hypersensitive sites generated *in vitro* were localized at sites approximately equidistant between the two Sp1 binding sites (boxes I and II), suggesting that the binding of the Sp1-like proteins distorted the

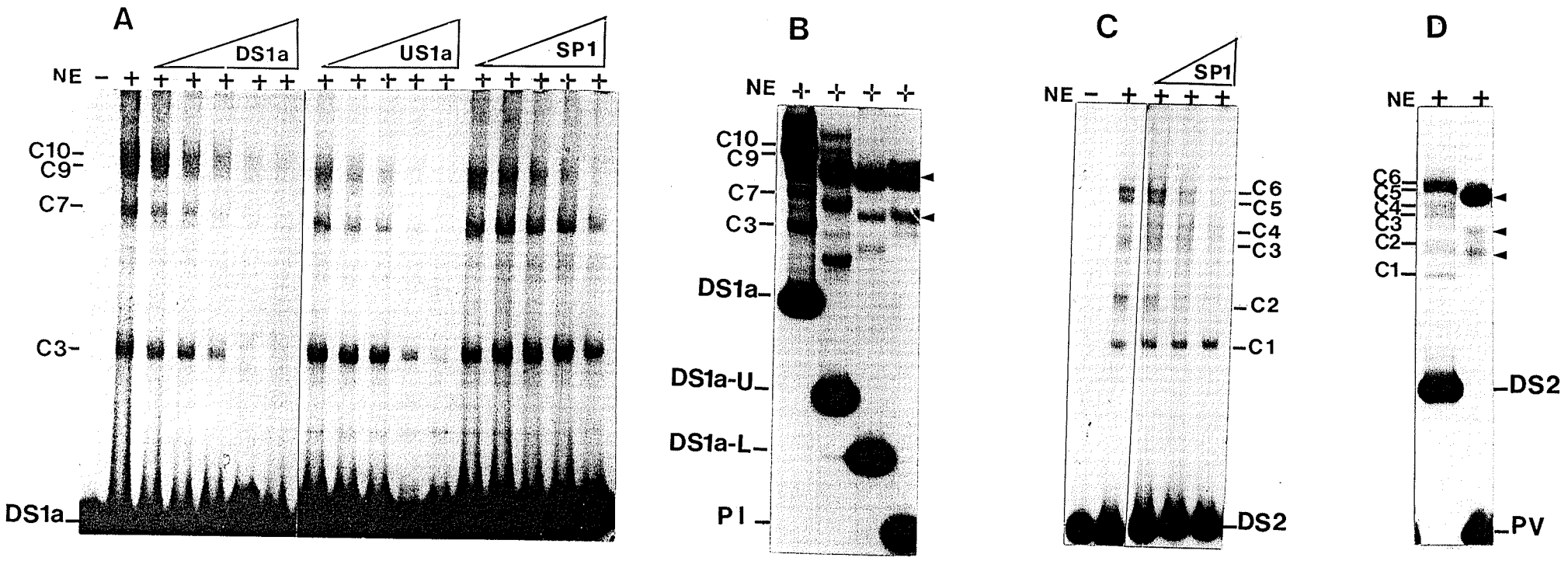
Figure 30. DNase I protection analysis of proteins interacting with the histone H5 enhancer region. DNA fragments DS1a (A) and DS1 + DS2 (B) were end-labelled on the template (T) or non-template (NT) strand, and incubated with (+) or without (-) chicken immature erythrocyte nuclear extract (12 or 14 μg of protein per ng of DNA for DS1a or DS1 + DS2 respectively). Identical DNase I protected regions were obtained when the complementary strand was end-labelled (not shown). G + A or C + T chemical sequencing reactions are indicated. The amount of DNase I added to 50 μl was as follows (left to right): 0.1, 0.05, 0.2, 0.1, 0.01 and 0.1 units (Panel A) and 0.1, 0.2, 0.1 and 0.01 units (Panel B). The arrowheads show the DNase I hypersensitive sites of the complex. The Roman numerals indicate the protected regions. This work was done by Mr. Jian-Min Sun.


A
C+T
T
– – + + + –
1018–
998–
979–
952–
925–
III
II
I
B
G+A
NT
– + + +
1043–
1070–
1081–
1100–
1116–
1146–
IV
V
VI

intervening DNA structure. The third protected region (DNase I protected region III), which was in DNA fragment DS1a-L, also harboured a potential Sp1 binding site (CTCCT^{+1002}CCTT). The protected region III also had the sequence GAGAGGGG which was the sequence of the ß-globin developmental stage selector element that bound the 65 kDa protein NF-E4 (Gallarda *et al.*, 1989). Three protected regions (DNase I protected regions IV, V and VI) were observed between +1041 and +1177. Protected region IV had the AGATAA motif at +1053, which is the binding site of GATA-1 described above. A potential Sp1 binding site (GGGGT^{+1093}GGGG) was found in the 24 bp protected region V. This sequence is also recognized by the CACCC factor (Jackson *et al.*, 1989). DNase I protected region VI harboured a TTGGCACAGCCCCAA sequence (+1122 to +1135) which was similar to the NF1 binding sequence (YTGGCANNNTGCCAR; Rupp *et al.*, 1990; Kruse *et al.*, 1991).

EMSAs with the DNA fragment DS1a (+893 to +1041; see Figures 6, 8 and 31A) used to detect chicken erythroid nuclear proteins that bind to this region shows that the DNA fragment DS1a formed four complexes, C3, C7, C9 and C10. The molecular masses of the DNA-bound proteins were approximately 200 kDa for complex C3, while the molecular masses of the proteins of complexes C7, C9 and C10 were greater than 450 kDa, which is the limit of determining size (Bading, 1988). Competition experiments were used to determine the specificity of binding of these complexes to DS1a (Figure 31A). Complexes C7, C9 and C10 showed specific binding to the DS1a fragment as they had for DS1 (Figure 8). A 100-fold molar excess of the unlabelled DNA fragment DS1a competed effectively for all four complexes (Figure 31A). The non-specific competitor salmon sperm DNA did not

Figure 31. Interactions between the histone H5 enhancer region and proteins from adult chicken immature erythrocyte nuclear extracts. The DNA fragment used in the EMSA is indicated at the bottom of each panel. Panel A: End-labelled DNA fragment DS1a (207 bp, 0.3 ng) was incubated with 0.36 μg poly (dI·dC)/μl in the presence (+, 5 μg) or absence (-) of nuclear proteins extracted (NE) from adult chicken immature erythrocytes. The amount in molar excess of competitor DNA fragment (DS1a, US1a or Sp1-oligonucleotide from the Stratagene Hotfoot kit), as indicated at the top of the panel, was 12.5, 25, 50, 100, and 200. Panel B; End-labelled DNA fragments DS1a, DS1a-U (91 bp), DS1a-L (63 bp), or oligonucleotide PI were incubated in the presence (+) or absence (-) of immature erythroid nuclear proteins. The oligonucleotide was incubated with 0.05 μg poly (dI·dC)/μl. The arrow heads show the complexes formed with the oligonucleotide PI. Panel C: End-labelled DNA fragment DS2 (124 bp, 1 ng) was incubated with 0.36 μg poly (dI·dC)/μl in the presence (+, 12 μg) or absence (-) of erythroid nuclear proteins. The amount in molar excess of competitor Sp1-oligonucleotide (histone H5 promoter) added was 12.5, 50 and 200. C1 to C6 are the protein-DNA complexes. Panel D: End-labelled DNA fragment DS2 and oligonucleotide PV were incubated with 0.36 or 0.05 μg poly (dI·dC)/μl, respectively, in the presence (+, 12 μg) of erythroid nuclear proteins. The arrow heads point to the complexes formed with the oligonucleotide PV. This work was done with Mr. Jian-Min Sun.


A
NE
DS1a
US1a
SP1
C10
C9
C7
C3
DS1a
B
NE
C10
C9
C7
C3
DS1a
DS1a-U
DS1a-L
PI
C
NE
SP1
C6
C5
C4
C3
C2
C1
DS2
D
NE
C6
C5
C4
C3
C2
C1
DS2
PV

effectively compete for C7, C9 and C10, but did compete for complex C3 (not shown), suggesting that complex C3 is non-specific. Competition studies were also done with a DNA fragment spanning the promoter region of the histone H5 gene (fragment US1a, Figure 6). The US1a DNA fragment competed for the four complexes forming with DS1a with the following efficiencies: C9, C10 > C7> C3 (Figure 31A).

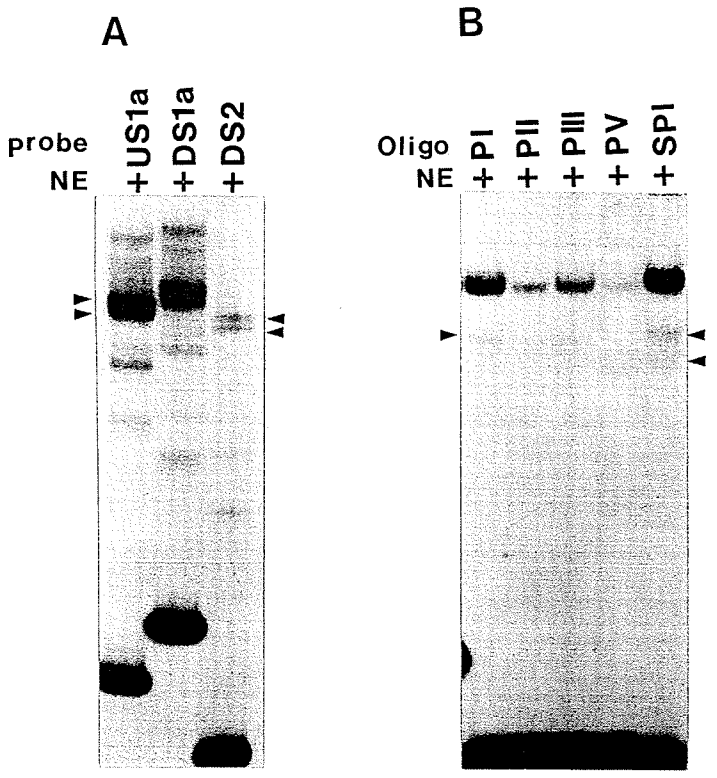
The results of the DNase I protection assays indicated that several potential Sp1 binding sites were localized in the region +893 to +1041. Since the histone H5 promoter region has a strong Sp1 binding site (GC-box), it was possible that the US1a DNA fragment was competing with the DS1a DNA fragment for the binding of a Sp1-like protein. This possibility was tested by using an oligonucleotide containing a high affinity Sp1 binding site (see Materials and Methods). The Sp1 oligonucleotide competed effectively for complexes C9, C10 and to a lesser extent, C7 that formed with the DNA fragment DS1a (Figure 30A). The Sp1-oligonucleotide did not compete effectively with complex C3. These observations suggested that a Sp1-like protein was involved in the generation of complexes C9 and C10.

To further delineate the location of the complexes, EMSAs were done with the DNA fragments DS1a-U (+893 to +980) and DS1a-L (+981 to +1041, Figure 6). Two strong and several weaker complexes were formed with fragment DS1a-U (Figures 31B and 8). Other gel patterns showed that the major complex formed with DS1a-U was two closely migrating complexes with mobilities slightly greater than complexes C9 and C10 that formed with DS1a. One major complex with electrophoretic mobility similar to the major complex generated with DS1a-U formed

with the fragment DS1a-L (Figure 8 and 31B). The intensity of this complex was similar to the intensity of the major complexes formed with DS1a or DS1a-U. To determine whether DS1a-U and DS1a-L contained Sp1-binding sites, competition experiments with an oligonucleotide containing the high affinity Sp1 binding site of the histone H5 promoter (-92 to -67) were done. This Sp1-oligonucleotide effectively competed for the major complexes forming with DS1a-U and DS1a-L.

Oligonucleotides containing the potential Sp1 binding sites present in DS1a were synthesized (see Materials and Methods). The oligonucleotides, which are referred to as PI, PII, and PIII, correspond to DNase I protected regions I, II and III respectively (Figure 30A). Two complexes, one strong and the other weak, formed with oligonucleotides PI, PII and PIII (Figure 32B). The observation that several complexes form with the Sp1-oligonucleotide is consistent with the observations of others, indicating factors other than Sp1, bind to these sequences (Jackson *et al.*, 1989; Ma *et al.*, 1991; Imataka *et al.*, 1992; Alemany *et al.*, 1992). These factors include CACCC factors, Sp1-like factors and factors which are not well characterized. The mobilities of the major complexes formed with these oligonucleotides were identical. However, the relative intensities of the major complexes were different with PI > PIII > PII, reflecting their relative binding affinities for Sp1. The mobilities of the two complexes formed with oligonucleotide PI were identical to those generated with DS1a-L and slightly greater than the migration of the major complexes formed with DS1a and DS1a-U (Figure 31B). Thus, although the lengths of the DNA fragments DS1a (156 bp), DS1a-U (91 bp), DS1a-L (63 bp) and the PI oligonucleotide (26 bp) were quite different, the migration of the major complexes formed with these fragments were similar. These

Figure 32. Analysis of protein binding to oligonucleotides containing Sp1 sites. Panel A; End-labelled DNA fragments US1a (182 bp, 0.5 ng), DS1a (207 bp, 0.5 ng) and DS2 (124 bp, 0.5 ng) were incubated with 0.36 μg poly (dI·dC)/μl in the presence (+, 5 μg) of nuclear proteins extracted (NE) from adult chicken immature erythrocytes. The arrow heads point to the major complexes formed with these DNA fragments. Panel B; End-labelled oligonucleotides PI, PII, PIII, PV and Sp1 were incubated with 0.36 μg poly (dI·dC)/μl in the presence (+, 5 μg) of nuclear proteins extracted (NE) from adult chicken immature erythrocytes. The arrow heads point to the minor complexes formed with these DNA fragments. This work was done by Mr. Jian-Min Sun.


A
probe
NE
+US1a
+DS1a
+DS2
B
Oligo
NE
+PI
+PII
+PIII
+PV
+SPI

observations demonstrated that an Sp1-like protein was involved in the generation of the major complexes formed with DNA fragments DS1a, DS1a-U and DS1a-L.

EMSAs with the DNA fragment DS2 (Figure 6) were done to ascertain whether erythroid nuclear proteins bind to the region +1068 to +1177. Figure 31C shows that 6 protein-DNA complexes were formed with this DNA fragment, with complexes C5 and C6 being the most intense (see also Figure 32A). The approximate molecular masses of the proteins present in complexes C1 and C2 were 227 and 347 kDa, respectively. The proteins of complexes C3 to C6 were greater than 450 kDa. Unlabelled DNA fragment DS2 competed effectively for the six complexes, while salmon sperm DNA competed effectively for only C1 (not shown), suggesting that complex C1 is non-specific. DNase I protection analysis (Figure 30B) indicated a potential Sp1 binding site, GGGGT^{+1093}GGGG, in the DS2 fragment. Figure 31C shows that an oligonucleotide containing the high affinity Sp1 binding site of the histone H5 promoter (-92 to -67) effectively competed for complexes C2 to C6 (DS2 DNA fragment). An oligonucleotide (PV) containing the potential Sp1 binding site (DNase I protected region V in Figure 30B) in DS2 was synthesized. One major complex, which had a mobility similar to C5 and C6, and several less intense complexes were generated with the oligonucleotide PV (Figure 31D). This is consistent with an Sp1-like protein(s) being involved in the generation of the major complexes formed with DNA fragment DS2.

Comparison of the intensities of the major complexes formed with DNA fragments DS2, DS1a and US1a indicated that the Sp1 binding site of DS2 had the lowest relative affinity for Sp1. This was confirmed by comparing the relative intensities of the major complexes formed with oligonucleotides PI, PII, PIII, PV and

Sp1 which has the high affinity Sp1 binding site of the histone H5 promoter. The relative intensities of the major complexes were Sp1 > PI > PIII > PII > PV (Figure 32B).

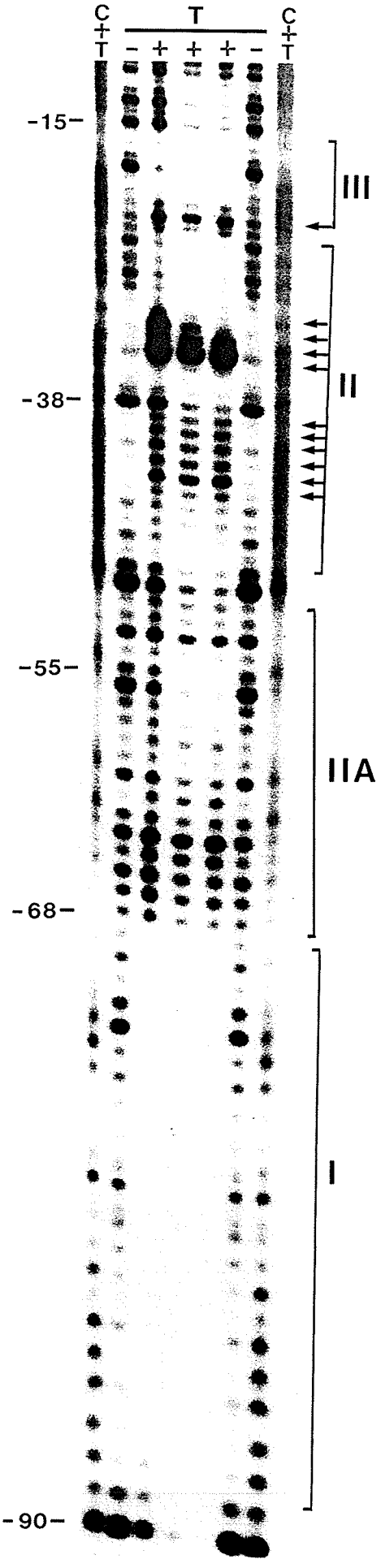
Identification of Proteins that Bind to DNA Fragments Spanning the Histone H5 5' Promoter

DNase I Protection Assays of US1a

Figure 33 shows the DNase I protected regions generated by proteins binding to DNA sequences present in the US1a DNA fragment, which spans the promoter region (-95 to -1, Figure 6). Four protected regions were observed. DNase I protected region I was a 21 bp region that harboured the strong Sp1 binding site (GGGGC^{-79}GGGG), DNase I protected region II contained the UPE, DNase I protected region IIA was from -68 to -52 and DNase I protected region III spanned the TATA box at -22 (TTAAAT). DNase I protected region IIa includes the region -68 to -63, since this region was also protected in Figure 36 and in Rousseau *et al.* (1993). DNase I protected region IIa was not characterized further.

Rousseau *et al.* (1993) noted a similarity between the sequence of DNase I protected region IIA and a portion of conserved region CTGCGGGGACACATNT in the histone H2A and H2B genes of *C elegans*. The GGGACAC sequence occurs between the *his*-II and *his*-12 (H2A and H2B) genes, *his*-10 and *his*-9 (H4 and H3) and in the upstream of the *his-3* (H2a) gene in *C. elegans* (Roberts *et al.* 1989). Imperfect copies of the sequence are also present in the *his*-1 (H4) and elsewhere in these *C. elegans* genes, many of which are conserved (Roberts *et al.*, 1989). Rousseau *et al.* (1993) refers to this sequence as a direct repeat. The region was

Figure 33. DNase I protection analysis of proteins binding to the histone H5 promoter region. DNA fragment US1a was end-labelled on the template (T) strand and incubated in the absence (-) or presence (+) of the nuclear extract (20 μg protein per ng DNA) isolated from chicken immature erythroid nuclei. Identical DNase I protected regions were obtained with the end-labelled non-template strand (not shown). C + T indicates the chemical sequencing reaction. The amount of DNase I added to 50 μl was as follows (left to right): 0.1, 0.2, 0.1, 0.01, and 0.1 units. The arrowheads are the DNase I hypersensitive sites of the complex. The Roman numerals indicate the protected regions. This work was done by Mr. Jian-Min Sun.

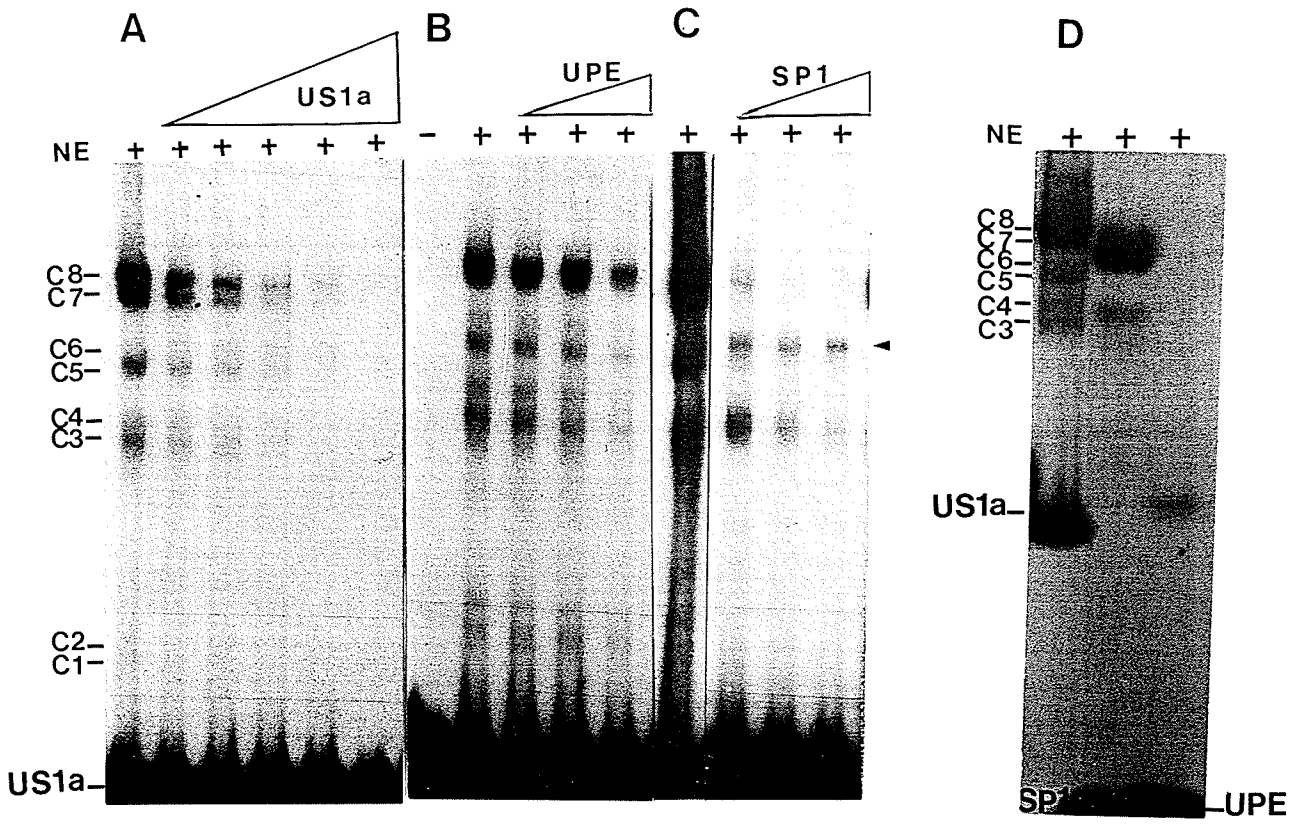

C+T
T
C+T
– + + + –
-15-
-38-
-55-
-68-
-90-
III
II
IIA
I

not required for promoter activity in transient transfection assays in HD3 cells (Rousseau *et al.*, 1989). Characterization of binding factors for this region of the histone H5 gene would allow the function of this *cis* element to be determined.

EMSAs were done with DNA fragment US1a, and a nuclear extract from adult chicken immature erythrocytes. Figure 34 shows that several protein-DNA complexes were formed with the US1a DNA fragment. Of the eight complexes detected, complexes C7 and C8 were the most intense followed by complexes C3, C4 and C5. Complexes C1, C2 and C6 were weak. The approximate molecular masses of the proteins present in the various complexes were as follows: C1, 90 kDa; C2, 117 kDa; C3, 334 kDa; C4, 380 kDa. All the protein-DNA complexes showed specific binding to the US1a fragment. A 100-fold molar excess of the unlabelled DNA fragment US1a competed effectively for all the complexes (Figure 34A), whereas the non-specific competitor salmon sperm DNA did not effectively compete (not shown). DNA fragment DS1 (enhancer region) also competed effectively for all the complexes, with the exception of complex C6 (not shown).

The two positive elements of the histone H5 gene promoter are a GC-box (the strong Sp1 binding site) and an upstream promoter element (UPE) (Rousseau *et al.*, 1989). Oligonucleotides containing these elements were used in competition and EMSA experiments. A 25-fold molar excess of the Sp1-oligonucleotide competed effectively for complexes C5, C7 and C8, but not complex C6 (Figure 34C). Similar results were obtained when this study was done with another oligonucleotide containing a strong Sp1 binding site from the Stratagene HotFoot DNase I Footprinting kit (see Materials and Methods). A 50-fold excess of the UPE-oligonucleotide effectively competed for complex C6 (Figure 34B). However, this

Figure 34. Interactions between the histone H5 promoter region and proteins from adult chicken immature erythrocyte nuclear extracts. Panel A: End-labelled DNA fragment US1a (182 bp, 0.5 ng) was incubated with 0.36 μg poly (dI·dC)/μl in the presence (+, 5 μg) or absence (-) of nuclear proteins extracted (NE) from adult chicken immature erythrocytes. The amount in molar excess of competitor DNA, which is indicated at the top of each panel, was 12.5, 25, 50, 100, and 200 for US1a and 12.5, 25 and 50 for UPE-oligonucleotide (panel B) and Sp1-oligonucleotide (histone H5 promoter) (panel C). The arrowhead points to complex C6. Panel D: End-labelled DNA fragment US1a (182 bp, 0.5 ng), oligonucleotide Sp1 or oligonucleotide UPE were incubated with 0.36 (or 0.05 for the oligonucleotides) μg poly (dI·dC)/μl in the presence (+, 5 μg) of nuclear proteins extracted (NE) from adult chicken immature erythrocytes. C1 to C8 are the protein-DNA complexes. This work was done by Mr. Jian-Min Sun.


A
B
C
D
US1a
UPE
SP1
NE
C8
C7
C6
C5
C4
C3
C2
C1
US1a
SP1
UPE

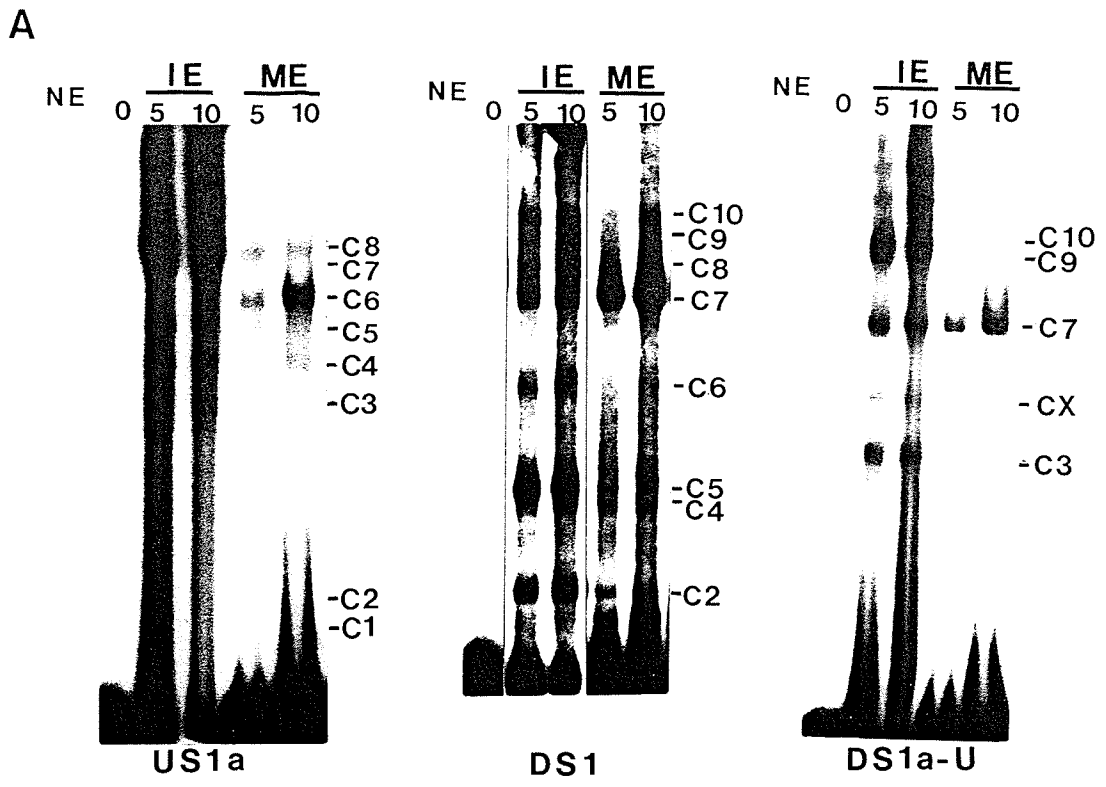
oligonucleotide was not effective at competing for the formation of complexes C7 and C8. An oligonucleotide containing the GATA sequence did not compete effectively (not shown). The Sp1 oligonucleotide generated a major complex and several minor complexes that had mobilities similar to complexes of similar intensities formed with the much longer US1a fragment (Figure 34D). This result demonstrated that the Sp1 was involved in the formation of the major complexes C7 and C8.

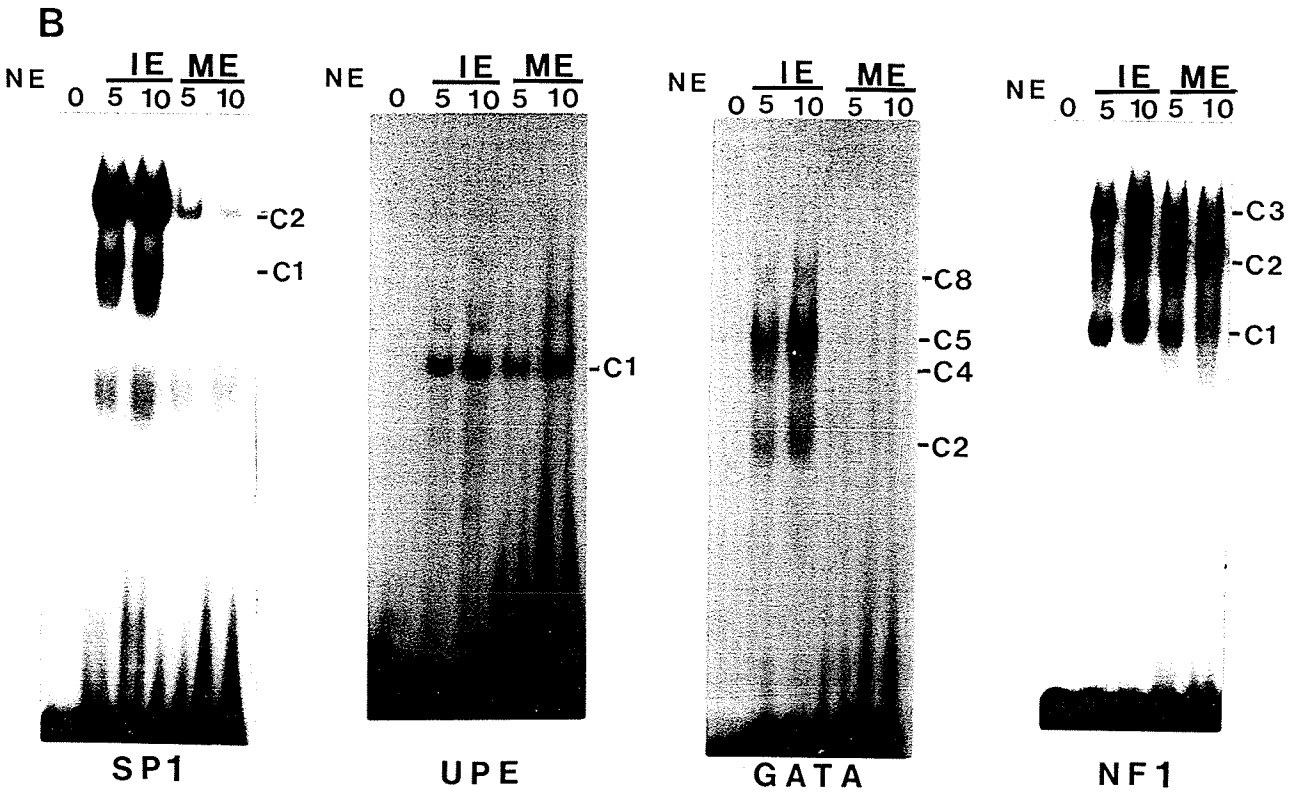
One of the minor complexes formed with the PV oligonucleotide would be generated by the binding of the CACCC factor (Jackson *et al.*, 1989). The DNase I protected region II over the histone H5 upstream promoter region was interesting in that it had protected regions interspersed with DNase I hypersensitive sites (Figure 33). To provide further evidence that a protein was associated with the histone H5 upstream promoter element (UPE), an oligonucleotide containing this element was used in EMSAs. Figure 34D shows that the UPE-oligonucleotide formed a complex with electrophoretic mobility slightly slower than the free US1a fragment.

Levels of Transcription Factors in Immature and Mature Erythrocyte Nuclei

The EMSA with DNA fragments US1a, DS1 and DS1a-U was used to detect adult chicken immature and mature erythroid nuclear proteins that interacted with the histone H5 promoter and enhancer regions. The abundance of the US1a complexes C7 and C8 was low in mature erythrocyte nuclear extracts (Figure 35A). The amounts of these complexes in different mature erythrocyte nuclear extracts was variable, but their levels were always considerably lower than those of immature erythrocyte nuclear extracts. Furthermore, the Sp1 complexes C9 and

Figure 35. Transcription factor levels of adult chicken immature and mature erythrocytes. The DNA fragment or oligonucleotide used in the EMSAs are indicated at the bottom of each panel. A. Histone H5 upstream DNA fragment US1a and downstream DNA fragments DS1, DS1a-U were incubated with nuclear extracts (NE) isolated from immature (IE) or mature (ME) erythrocytes. One ng of end-labelled DNA fragments was incubated with either 5 or 10 μg of protein as indicated. B. End-labelled oligonucleotides were incubated with 5 or 10 μg of immature or mature erythrocyte nuclear extracted protein as indicated. C1 to C10 are the protein-DNA complexes. This work was done by Mr. Jian-Min Sun.


A
NE
IE
ME
0 5 10 5 10
-C8
-C7
-C6
-C5
-C4
-C3
-C2
-C1
US1a
-C10
-C9
-C8
-C7
-C6
-C5
-C4
-C2
DS1
-C10
-C9
-C7
-CX
-C3
DS1a-U


B
NE
IE
ME
0 5 10 5 10
-C2
-C1
SP1
-C1
UPE
-C8
-C5
-C4
-C2
GATA
-C3
-C2
-C1
NF1

C10 that formed with DS1a-U were lower in abundance in mature erythrocytes nuclear extracts than in immature erythrocyte nuclear extracts (Figure 35A). This suggested that mature erythrocyte nuclear extracts had low levels of Sp1. This was confirmed in EMSAs with SP1-oligonucleotide and mature or immature erythrocyte nuclear extracts. Figure 35B shows that Sp1 binding activity was reduced in mature erythrocyte nuclear extracts which had approximately 10 to 20% of the Sp1 activity present in immature erythrocyte nuclear extracts. Transcription factors have been shown to leak out of the erythrocyte nuclei during their isolation (Jackson *et al.*, 1989; Emerson *et al.*, 1987). However, when nuclear extracts and cytoplasmic supernatants from mature erythrocytes were analyzed by EMSA, the majority of the Sp1 and GATA-1 activities were present in the nuclear extracts (not shown). The complex CX in this EMSA is a minor component, which was not examined more closely.

The major complex formed with mature nuclear extracts and the US1a fragment was the UPE complex C6. EMSAs with a UPE-oligonucleotide and nuclear extracts from immature and mature erythrocyte nuclear extracts demonstrated that the abundance of the UPE-binding proteins was similar in both cell types (Figure 35B).

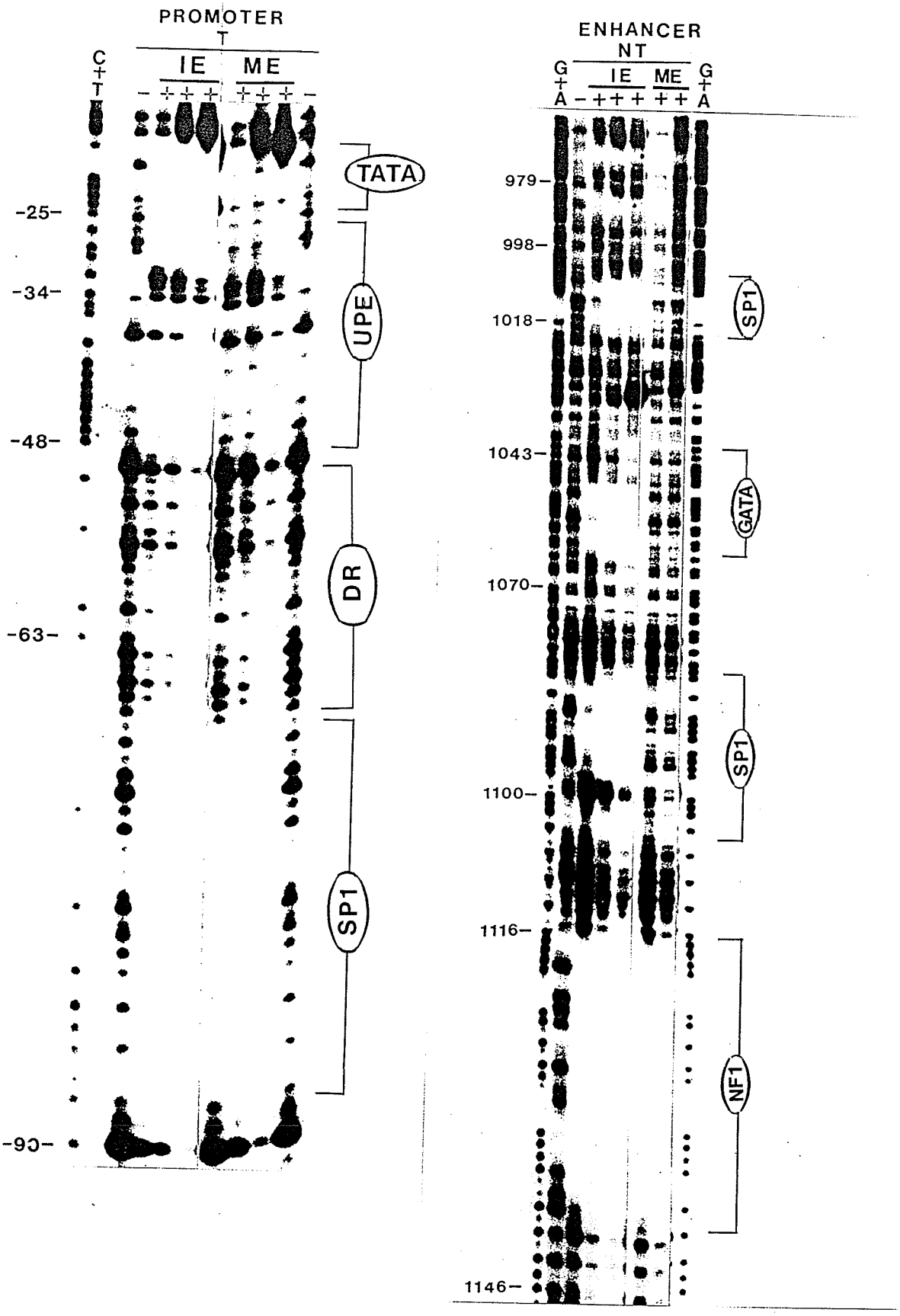
The relative levels of each of the GATA-1 complexes C2, C4, C5 and C8 formed with DS1 were lower in mature erythrocyte nuclear extracts than in immature erythrocyte nuclear extracts. EMSA with GATA-oligonucleotide and either mature or immature erythrocyte nuclear extracts demonstrated that GATA-1 and GATA-1 multisubunit complexes were low in mature erythrocyte nuclear extracts (approximately 10% of the immature erythrocyte nuclear GATA-1 activity)

(Figure 35B).

In contrast to GATA-1 and Sp1, the abundance of NF1 was similar in nuclear extracts from mature and immature erythrocytes (Figure 35B). Both nuclear extracts formed a similar spectrum of complexes with a NF1-oligonucleotide that contained the NF1 binding site in the histone H5 enhancer. These complexes were effectively competed with unlabelled NF1-oligonucleotide but not with the Sp1 oligonucleotide (not shown).

Figure 36 shows DNase I protection analysis of DNA fragments spanning either the histone H5 promoter or enhancer and nuclear extracts of immature and mature erythrocytes. The TATA box region, DR (direct region at -68 to -52) and UPE of the promoter were protected by both the mature and immature nuclear extracts. Similarly, the high affinity Sp1 binding site of the promoter was protected with either immature or mature erythrocytes. With different mature erythrocyte nuclear extracts this high affinity Sp1 binding site was generally well protected. However, the weaker Sp1 binding sites in the enhancer region were less protected with nuclear extracts from mature cells than with those from immature erythrocytes. The degree of protection at the Sp1 binding sites in the enhancer region varied with the different mature erythrocyte nuclear extracts, reflecting the different levels of Sp1 activity in these preparations. The level of protection at these enhancer Sp1 binding sites generated with mature erythrocyte nuclear extracts was always less than that protected by immature erythrocyte nuclear extracts. Similarly, the protection of the GATA binding site was always greater with immature erythrocyte nuclear extracts than with mature erythrocytes nuclear extracts. In contrast, the UPE and NF1 binding sites were similarly protected with

Figure 36. DNase I protection analysis of adult chicken immature and mature erythrocyte nuclear extracted proteins interacting with the histone H5 promoter and enhancer regions. DNA fragments (1 to 2 ng) US1a (promoter) or DS1 + DS2 (enhancer) were end-labelled on the template (T) or non-template (NT) strand, and incubated with (+) or without (-) 10 to 30 μg of nuclear extracted protein isolated from immature (IE) or mature (ME) erythrocytes. C + T or G + A chemical sequencing reactions are indicated. The amount of DNase I added to 50 μl was as follows (left to right): 0.2, 0.1, and 0.01 units. For enhancer panel (ME) the units of DNase I were 0.2 and 0.1 units. The factor binding sites are indicated. This work was done by Mr. Jian-Min Sun.


PROMOTER
T
C+T
IE
ME
- + + + + + + -
-25-
-34-
-48-
-63-
-90-
TATA
UPE
DR
SP1
ENHANCER
NT
G+A
IE
ME
- + + + + + G+A
979-
998-
1018-
1043-
1070-
1100-
1116-
1146-
SP1
GATA
SP1
NF1

nuclear extracts from immature and mature erythroid nuclear extracts. It should be noted that this assay measures only DNA-binding activity and not absolute levels of factor. For example factors could be present, but such that they no longer bind to DNA and are thus rendered undetectable in the EMSA.

Discussion

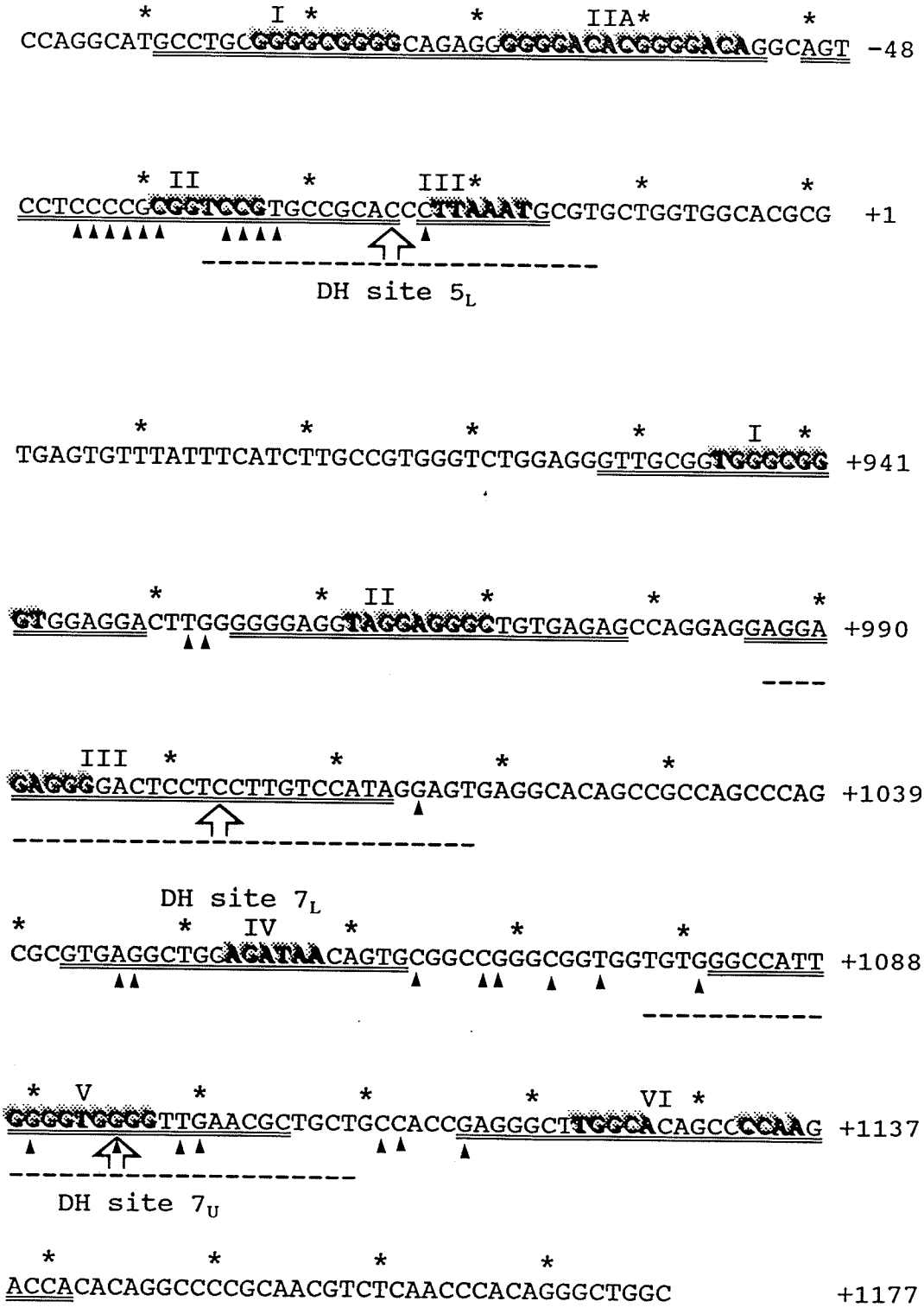
Factors Binding to the Histone H5 Promoter and 3' Enhancer

In vitro binding of adult chicken immature erythroid nuclear proteins to the promoter and the 3' enhancer regions of the chicken histone H5 gene were examined in this project (Figure 6). These data are summarized in Figure 37. The chicken histone H5 promoter region has two positive elements, a GC-box at -83 to -75 and a proximal element (UPE) at -54 to -38 (Rousseau *et al.*, 1989). In DNase I protection assays, adult chicken immature erythrocyte nuclear extracts protected four regions of the DNA fragment US1a (Figure 6) from digestion. These regions corresponded to the GC box, a high affinity Sp1 binding site (DNase I protected region I), a direct repeat (DNase I protected region IIA), UPE (DNase I protected region II) and TATA box (DNase I protected region III, Figure 30).

To gain a better understanding of the factors binding to the promoter region, the DNA fragment US1a (Figure 6) was used in EMSAs with nuclear extracts of adult chicken immature erythrocytes. This DNA fragment formed several protein-DNA complexes. Complexes C5, C7 and C8 which formed major and minor interactions with US1a as determined by relative intensities on the EMSA were accounted for by the binding of protein(s) to the GC box. The GC box was shown to be a strong Sp1 binding site as demonstrated by 1) EMSAs with US1a and with an oligonucleotide containing the Sp1 binding sequence and 2) competition with an oligonucleotide with Sp1 binding sequences.

The less abundant complex C6, which formed on the US1a DNA fragment, was due to the binding of the UPE-binding protein which formed DNase I protected region II. The histone H5 UPE has sequence similarity to the H4 gene subtype-

Figure 37. Summary of data from studies of proteins binding to the histone H5 promoter and enhancer regions and mapping of DNase I hypersensitive sites in histone H5 gene chromatin. The sequences are numbered from the histone H5 gene 5′ promoter and 3′ enhancer regions. Double underlines indicate the DNase I protected regions and the arrowheads show the positions of the DNase I hypersensitive sites in the complexes. The shadowed sequences show the recognition sequence of various transcription factors. Roman numerals correspond to those shown in Figure 30 and 33. The dashed line and arrow show the positions of the DNase I hypersensitive regions in the histone H5 gene chromatin of chicken immature erythrocytes (Renaud and Ruiz-Carrillo, 1986).


I
IIA
II
III
DH site 5L
I
II
III
DH site 7L
IV
V
VI
DH site 7U
-48
+1
+941
+990
+1039
+1088
+1137
+1177

specific element which binds to the 65 kDa transcription factor H4TF2 (binding sequence, GGTTCTCN$_4$CGGTCCG (Rousseau *et al.*, 1989; Faisst and Meyer, 1992). However, the chicken histone H5 gene UPE-binding factor has a molecular mass of 77 kDa (Mr. J.-M. Sun, unpublished data) suggesting that the UPE-binding protein is not the same as H4TF2. Other minor complexes may be due to the binding of TATA-box binding proteins, since the DNase I protected region III was over the TATA box. Other less prevalent complexes may be due to binding of protein to DNAse I protected region IIA.

The approximate mass of the protein(s) of the major complexes C7 and C8 was greater than 450 kDa which is much greater than the molecular mass of the Sp1-like protein (typically 95/105 kDa; Faisst and Meyer, 1992). In preliminary studies, the molecular mass of the chicken erythrocyte nuclear proteins that bind to the Sp1 oligonucleotide are 75/97/105 kDa (Mr. J.-M. Sun, unpublished data). It has been demonstrated by several groups that Sp1 forms multimers (Mastrangelo *et al.*, 1991; Pascal and Tjian, 1991; Su *et al.*, 1991). Mastrangelo *et al.* (1991) have shown that Sp1 is first organized as a tetramer which is subsequently assembled into multiple tetramers. It is possible that complexes C7 and C8 consist of several stacked Sp1 tetramers. The slowest migrating complex C8 of the US1a EMSA may be generated by the combined association of Sp1-like proteins and other proteins. For example, Sp1 has been shown to interact with *Drosophila* TFIID, the multisubunit complex which binds to the TATA box (Hoey *et al.*, 1993; Weinzierl *et al.*, 1993).

Comparison of Published DNase I Protected Regions of the Histone H5 Promoter

Rousseau *et al.* (1993) also performed DNase I protection assays of the promoter region of the histone H5 gene. Several of the protected regions which they observed were different from those noted in this project. Small differences between the protected region observed by Rousseau *et al.* (1993) and this project are probably not significant. In some cases the data of Rousseau *et al.* (1993) were for the opposite strand from that which we studied. Some DNA-binding proteins will interact with different strands of DNA to different degrees. The enzyme used DNase I also displays some sequence specificity with respect to sites of action. We also found that changes in the concentration of DNase I used for a reaction also had an affect on the extent of digestion of DNA near a DNase I protected site (for example compare the 3 lanes in which protein was present in Figure 30A). This could have explained small differences in the extent of protection of DNase I protected regions between Rousseau *et al.* (1993) and ourselves. The region used by Rousseau *et al.* (1993) extended from -313 to +67, and therefore extends 5′ and 3′ of US1a (-97 to +10) used in this project. Note that the numbering system of Rousseau *et al.*, 1993 is shifted 5 nucleotides in the 5′ direction when compared to our numbering. The reason for this is that the original clone (Krieg *et al.*, 1983; Ruiz-Carrillo *et al.*, 1983) included a cloning artifact in which the sequence GGCGGCAGC was absent in the mRNA leader. Gomez-Cuadrado *et al.* (1992) reported the revised start site of transcription. The work of this project was however, done with the original plasmid and therefore the earlier numbering system is used for both the work reported here and in comparisons between this work and the literature data. The fragment used by Rousseau *et al.*

(1993) included the UNE and their data showed an Sp1 binding site located at -164 5′ of US1a and at the 3′ end of the fragment the cIBR/cIBF binding site. Rousseau et al. (1993, 1989) did not determine the effect of deletion of the Sp1 binding site at -164. The cIBR/cIBF site was partially deleted in the DNA fragment US1a used in this project. The sources of nuclear extract used by Rousseau *et al.* (1993) were differentiated HD3 cells, adult chicken mature erythrocytes and T-lymphoblastoma (MSB-1) cells. Our project used adult chicken immature and mature erythrocyte nuclei.

The Sp1 binding site (DNase I protected region III of Rousseau *et al.*, 1993; DNase I protected region I in Figure 33 of this thesis) was determined to be the same by both Rousseau *et al.* (1993) and us except that we interpreted the protected region as extending 1 nucleotide further 3′. Rousseau *et al.* (1993) used competition with Sp1 binding oligonucleotides in the DNase I protection assay to determine that Sp1 bound to this site. In our project Sp1 binding oligonucleotides were used in EMSAs and in competition against EMSAs with US1a to also show that Sp1 bound this site. This binding site is key to the model we propose for histone H5 gene chromatin.

Protected region IV of Rousseau *et al.* (1993) which extends from -68 to -52 of US1a was also determined by us (Figure 33) using nuclear extracts of adult chicken immature erythrocytes. This region was best protected using nuclear extracts of chicken mature erythrocytes and T-lymphoblastoma (MSB-1) cells (Figure 5 in Rousseau *et al.*, 1993). The smaller amount of protection by non-differentiated HD3 cells suggested the factor was in low abundance or was modified in such a way as to be unavailable for binding in these cells. Neither the

transcription factor binding in this region or the effect of deletion of only this region on promoter function was determined, although the sequence is similar to a direct repeat found in several histone gene promoters (Rousseau et al., 1993).

Our determination of the region protected by the UPE factor was -50 to -25. A comparison of this data to that presented by Rousseau *et al.* (1993) shows that they interpreted this region as consisting of two DNase I protected regions (V, -50 to -41 and VI, -32 to -20) with a protein free region between -40 and -33 and the second protected region extending further 3' than ours. We however, interpreted the region -40 to -33 as being hypersensitive to DNase I digestion due to binding of the UPE factor and that UPE bound over the entire region (-50 to -25). Examination of the DNAse I protected region of Rousseau *et al.* (1993) shows a similar hypersensitivity at -40 to -33. Rousseau *et al.* (1993) interpreted their DNase I protected region VI (-32 to -20) as a binding site for CACCC factor, TEF-2 or Con. DNase I protection assays with competitor DNA containing binding sites for Sp1 or the CACCC sequence, prevented detection of DNase I protected region VI (Rousseau et al., 1993), suggesting that Sp1 or CACCC binding factor may have bound the site. However, we found that the UPE oligonucleotide formed a single major complex in EMSAs, which migrated further than the Sp1 and CACCC binding factor complexes (Figure 34D). It should be noted that our DNase I protection data was for the template strand and that of Rousseau *et al.* (1993) was for the non-template strand. A synthesized oligonucleotide (UPE, -69 to -39) which has sequence similarity to the cognate sequence of H4TF2 was able to bind human and the 65 kDa chicken H4TF2 in EMSAs (Rousseau et al., 1993). However, our data show that the protein bound by UPE is 77 kDa (Mr. J.-M. Sun, unpublished data).

In order to determine the factor(s) which binds this region *in vivo*, further studies need to be done.

Our data show that a factor was binding from -23 to -16. This included part of protected region VI (-32 to -20) of Rousseau *et al.* (1993) and sequences 3′ of it. Our interpretation of this protected region was that TFIID was bound at this site which includes the TATA box. A potential reason that Rousseau *et al.* (1993) did not report factor binding at this site is that poly dA·dT was used as the non-specific competitor. Poly dA·dT may have competed for TFIID binding at the TATA box, a region rich in adenines and thymines.

Factors Binding to the 3′ Enhancer of Histone H5 Gene

The enhancer located at the 3′ end of the chicken histone H5 gene is erythroid-specific. The enhancer activity has been shown to be present in the region +851 to +1185 by Rousseau *et al.* (1989) and +851 to +1064 by Trainor *et al* (1987). In DNase I protection assays, adult chicken immature erythrocyte nuclear extracts protected six regions of DS1 and DS2 (+883 to +1177, Figure 6) from digestion. These regions corresponded to a medium (DNase I protected region I) and three weak Sp1 binding sites (DNase I protected regions II, III and V), a GATA binding site (DNase I protected region IV) and an NF1 binding site (DNase I protected region VI, Figure 30). Both the DNA fragments which span the region +893 to +1064 (DS1) and which span the region +1065 to +1177 (DS2) were used to detect adult immature chicken erythrocyte nuclear DNA-binding proteins that interacted with the enhancer DNA element (Figure 6). The DS1 fragment formed a number of complexes, C1-C10. The DS1 fragment could be separated

into two sequences; DS1a, which interacted with C3 (a non-specific complex), C7, C9 and C10, and DS1b, which interacted with C1, C2, C4-C6 and C8. The proteins forming complexes with DS1b were shown to interact with the GATA binding site. This was determined through DNase I protection assays and competition assays with sequences containing GATA binding sites. Furthermore, only the GATA core sequence was required for the nuclear protein interactions as demonstrated by the competition assays and through EMSAs with both the H5-GATA sequence and the chicken ß-globin gene sequence (+1888 to +1910) containing the GATA-binding site. Each of the complexes generated DNase I protected region IV, a DNase I protected region which had previously been observed by Rousseau *et al.* (1989), who used a DNA fragment encompassing the region +980 to +1087.

In adult chicken immature erythrocytes there are several proteins that recognize the GATA DNA binding motif. The levels of the transcripts encoding these proteins belonging to the GATA family indicate that, in adult immature erythroid cells, GATA-1 is the most abundant, followed by GATA-2, then GATA-3 (Yamamoto *et al.*, 1990). The results of the ultraviolet light-cross-linking experiments presented evidence that GATA-1 was the only member of the GATA-binding proteins that was affinity-labelled when nuclear extracts of adult chicken immature erythrocytes were used, indicating that GATA-1 is the major protein binding to the GATA sequence of the histone H5 enhancer region. Furthermore, complexes containing proteins of 55 and 56 kDa, which correspond to the size of GATA-2 and GATA-3, were not detected in the EMSA. Therefore the GATA-binding complexes were formed by proteins associating with the GATA-1 *trans-*

acting factor. It is unlikely that these complexes were formed by GATA-1 dimers, since complexes of approximately 80 to 90 kDa were not seen. Moreover, Evans and Felsenfeld (1989) reported that chicken GATA-1 did not dimerize. Furthermore the data indicate that GATA-1 is the sole DNA binding protein of these complexes.

The estimated molecular masses of these GATA-related complexes and the results of the diagonal EMSAs suggest the following arrangement of proteins in the complexes. A protein(s) of approximately 105 kDa associates with GATA-1, forming complex C4. The larger complexes are generated by other proteins associating with GATA-1 and the 105 kDa protein(s) as follows; complex C5, a 26 kDa protein(s); complex C6, 26 kDa and 146 kDa proteins; complex C8, a protein(s) with molecular mass greater than 450 kDa. Note that the molecular masses of these proteins are rough estimates. Partial purification of complex C5 caused enrichment of a protein of 54 kDa. This suggests that the 105 kDa protein(s) present in each of the complexes is composed of a dimer of two 54 kDa proteins which together form a molecular mass of 108 kDa which is close to the estimated mass of 105 kDa noted for the complex. Complex C5 and/or complex C8 may be the relevant GATA-1 protein complexes *in vivo*, since these were the most abundant complexes in EMSAs when nuclear extracts were prepared without dialysis. Complex C4, C6 and C2 may have been present in the EMSAs as a result of dissociation of the larger complexes during preparation of the nuclear extract. The proteolytic form of GATA-1, which forms complex C1, obviously has the ability to bind its recognition sequence. However, the presence of the proteolytic form was not detected in the complexes C4-C6 and C8, suggesting that the GATA-1 binding proteins do not interact with finger II of GATA-1 which is the DNA-binding

domain.

Several of the non-GATA-1 complexes of DS1, C7, C9 and C10 which also bound DS1a could be accounted for by the DNase I protected regions I, II and III. These regions and the DNase I protected region V of DS2 contained Sp1 binding sites (Figure 30). The CACCC sequence of DNase I protected region V may bind Sp1 or the CACCC factor (Jackson *et al.*, 1989). EMSAs, DNase I protection assays and competition experiments with the DNA fragments DS1a, DS1a-U, Ds1a-L and DS2 which span the histone H5 3′ enhancer region (Figure 30), demonstrated that Sp1-like high molecular masses bound to these protected regions of the downstream enhancer with varying affinities. The sequences of these Sp1 binding sites suggest that their relative affinities for the Sp1 protein are medium (DNase I protected region I) and weak (DNase I protected regions II, III and V) (Kadonaga *et al.*, 1986; Letovsky and Dynan, 1989; Lobanenkov *et al.*, 1990; Yu *et al.*, 1991). The level of DNase I protection (e.g. DNase I protected region I was stronger than DNase I protected region II with DS1a) and the relative intensities of the complexes formed with oligonucleotides PI, PII, PIII and PV agrees with the proposed affinities (Figures 30, 31).

In addition to being a weak Sp1-binding site, DNase I protected region III has the sequence GAGAGGGA, which is similar to the sequence recognised by the ß-globin enhancer chicken adult erythroid-specific factor NF-E4 (Emerson *et al.*, 1987; Gallarda *et al.*, 1989). Both the chicken ß-globin enhancer NF-E4 binding sequence and the histone H5 DNase I protected region III sequence are part of a 34 bp sequence which is similar in the two genes. The differences between the 34 bp sequence in the histone H5 3′ enhancer and the ß-globin 3′ enhancer include a

seven nucleotide mismatch and a single base-pair deletion (Trainor *et al.*, 1987). It should be noted that NF-E4 protects an 18 bp region of the ß-globin enhancer (Emerson *et al.*, 1987), and that the sequence of this 18 bp segment differs by five mismatches in the histone H5 enhancer. The DNA fragment (DS1a-L) containing this sequence does not form protein-DNA complexes well, suggesting that the abundance of NF-E4 is low or that NF-E4 has a weak affinity for this sequence in the histone H5 enhancer.

Protected region VI has an inverted repeat TTGGCACAGCCCCAA which is similar to the inverted repeat found in the chicken $ß^H$ globin gene promoter region and the consensus recognition sequence for the chicken TGGCA protein/NF1 (Rupp *et al.*, 1990; Kruse *et al.*, 1991; Rupp *et al.*, 1988). The TGGCA sequence is also found in the chicken ß-globin enhancer (DNase I protected region activity I, Emerson *et al.*, 1987). Deletion of this region did not affect ß-globin enhancer activity (Reitman and Felsenfeld, 1988). Sun *et al.* (1993) demonstrated that 58, 59 and 64 kDa NF1 forms were enriched in nuclear matrix extracts of mature chicken adult erythrocytes. This suggests that certain members of the NF1 family of proteins assist in the specific association of the histone H5 gene chromatin with the nuclear matrix. Therefore, NF1 which activates transcription (Jones *et al.*, 1987) may aid in the targeting of specific genes to sites on the nuclear matrix that have bound the transcription and RNA processing machinery (Sun *et al.*, 1993). Sun *et al.* (1993) suggest that a NF1 homo- or heterodimer binds to the NF1-oligonucleotide and that similar to what has been noted for Sp1 and GATA-1, higher molecular weight complexes may be formed by proteins associating with NF1 dimers by protein-protein interactions (Goyal *et al.*, 1990; Martin, 1991).

A Comparison of Published DNase I Protected Regions of the Histone H5 3' Enhancer

The regions protected in DNase I protection assays of the 3' enhancer were generally determined to be in the same regions by both Rousseau *et al.* (1993) and ourselves. DNase I protected region I which we determined to be from +928 to +949 was determined by Rousseau *et al.* (1993) to extend from +931 to +942. When the DNase I protected region was compared between mature erythrocytes and differentiated HD3 cells, it was noted that the region protected by extracts from differentiated HD3 cells extended 3' of +942 (Rousseau et al., 1993). They demonstrated through competition with Sp1-binding oligonucleotides in DNase I protection assays, using nuclear extracts from mature chicken erythrocytes, that factors other than Sp1 were able to bind this site. However, the data presented in this thesis demonstrate that the DNase I protected region I bound Sp1-like factors in EMSAs using extracts derived from adult chicken immature and mature erythroid nuclei. Taken together, these results suggest that erythrocytes actively expressing histone H5 as is the case for differentiated HD3 cells and immature adult chicken erythrocytes, have Sp1 bound at protected region I. The sequence may also allow binding of other factors such as CACCC factors. In mature erythrocytes, therefore where Sp1 levels are decreased and H5 is not expressed, a CACCC factor is also able to bind to the region, although the physiological relevance of this binding is unknown.

DNase I protected region II of the histone H5 3' enhancer was determined by this project to be from +955 to +978. Rousseau *et al.* (1993) determined this region to be from +953 to +969. Both groups agreed on the factor binding to this

region. Rousseau *et al.* (1993) used oligonucleotides containing an Sp1 binding site and a CACCC factor binding site in competitions with DNase I protection assay. Their data show the Sp1 and CACCC factors were able to bind this site. They also suggest that NF-E4 may bind this site. Our data show that in EMSAs using synthesized oligonucleotide containing the DNase I protected region II, complexes were formed that corresponded in size to Sp1 complexes and potentially a CACCC factor. Competitions with oligonucleotides containing an Sp1 binding site also prevented formation of this complex on DS1a in the EMSA.

DNase I protected region III was determined by us to extend from +986 to +1013 and by Rousseau *et al.* (1993) to be from +992 to +1014. Both the data of Rousseau *et al.* (1993) and our data suggest Sp1/CACCC factor/NF-E4 binding.

DNase I protected region IV is the binding site of GATA-1. We determined the region to be at +1043 to +1063 and Rousseau *et al.* (1993) determined it to be located at +1051 to +1062. We included DNase I hypersensitive sites at +1046 and +1047 and the 4 bases pairs 3′ of it which shows protection in our determination of the extent of the DNase I footprinted region. Our data are for the non-template strand. A difference in hypersensitivity depending on the strand which was labelled for the DNAse I protection assay was also seen in Figure 10.

We determined DNase I protected region V to extend from +1082 to +1105 whereas Rousseau *et al.* (1993) determined it to be at +1085 to +1099. Similar to DNase I protected region I, they determined that this site was not bound by Sp1 when extracts of mature chicken erythrocyte nuclei were used in DNase I protection assays, however, the region protected by differentiated HD3 cells more closely resemble the pattern we noted. We demonstrated that Sp1 and/or CACCC

factor were able to bind this region in EMSAs using an oligonucleotide which contained this binding site. We suggest that occupancy of this site by Sp1 occurs in erythrocytes actively expressing histone H5, but not in mature erythrocytes, where Sp1 levels are decreased. In mature erythrocytes the data of Rousseau *et al.* (1993) suggest that another factor binds to this region.

We determined site VI as extending from +1116 to +1141 whereas Rousseau *et al.* (1993) determined it to be from +1116 to +1143. This small discrepancy may have been due to opposite strands being examined. The data from both groups showed NF1 binding to the sequence. Rousseau *et al.* (1993), determined additional DNase I protected regions at +1160 to +1188 bound NF1, +1205 to +1237 and +1246 to +1264 (AP-1 consensus sequence) all of which were 3' of DS1 and DS2, the sequences used in this study. The latter two were not studied further by Rousseau *et al.* (1993). Preliminary evidence suggests that the NF1 site at +1160 to +1188 is not bound *in vivo*, since DNase I protection assays of histone H5 chromatin do not show a factor bound at the site (J.-M. Sun unpublished data). Chromatin beyond these sites is resistant to DNase I digestion and RNA polymerase II also terminates transcription in this region (Affolter and Ruiz-Carrillo, 1986). Therefore the last binding site marks the boundary of accessible chromatin (Renaud and Ruiz-Carrillo, 1986).

Enhancer Function of Histone H5 DNase I Hypersensitive Sites I and III

Rousseau *et al.* (1993) determined that DH sites I (-2296 to -1817) and III (-1321 to -1070) function as enhancers. They examined these two enhancers for potential factor binding sites. In the DH I region, nuclear extracts from chicken

erythroid cells protected 6 regions from DNase I digestion (Rousseau et al., 1993). Four of the sites were potential GATA-1 binding sites, two of which were shown to be capable of binding GATA-1 in EMSAs. The other DNase I protected sites were not studied in detail. One of the potential GATA-1 binding sites was also a potential NF1 binding site. DH site III has a GATA-1 binding site and a potential G-string protein binding site (Rousseau et al., 1993). DH sites I and III synergistically activate transcription when the 2 sequences are both present on a plasmid transfected into HD3 cells (Rousseau et al., 1993). The 3′ enhancer of histone H5 did not activate transcription synergistically with the 5′ enhancers in the same assay in non-differentiated and differentiated HD3 cells.

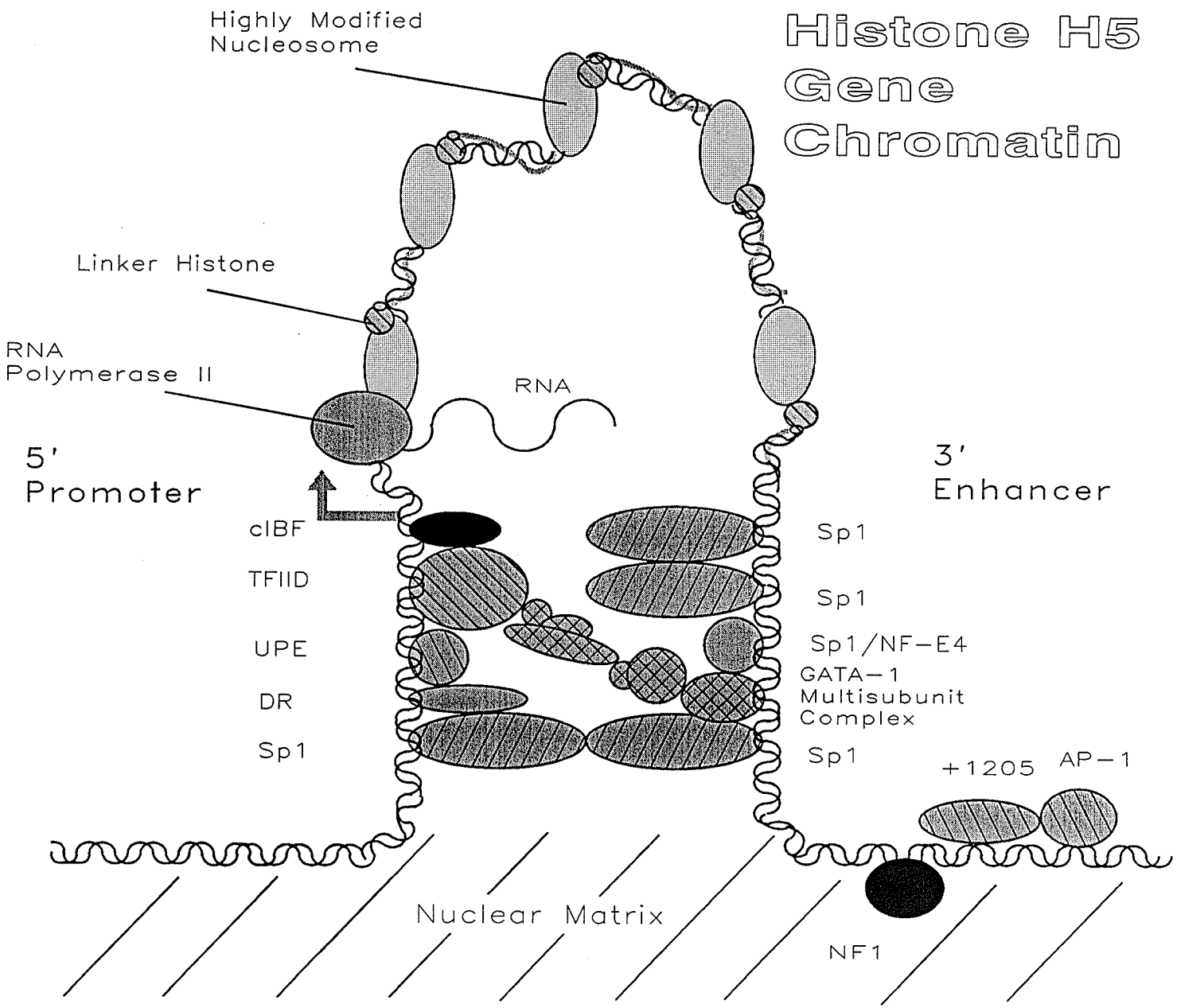
Histone H5 is expressed at the CFU-E stage of erythroid development, before the globin genes are expressed. Histone H5 expression continues through the terminal stages of erythroid differentiation, although its expression is decreased in mature erythrocytes. The mature erythrocyte contains factors which specifically repress histone H5 expression (Gomez-Cuadrado *et al.*, 1992; Bungert *et al.*, 1992b). Expression of histone H5 also occurs in the chicken embryo in primitive chicken erythrocytes (Pikaart *et al.*, 1991). Therefore histone H5 expression occurs prior to globin gene expression in a lineage dependent manner and in both the embryonic and the definitive erythrocytes. These latter cells have different ß-globin genes expressed. The transcription elements and the presence of three enhancer elements in histone H5 should reflect the need for expression in this wide range of development and lineage stages. The 3′ enhancer of the histone H5 gene has a strong binding site for GATA-1, which might be important for allowing expression of the histone H5 gene at the CFU-E stage when GATA-1 levels are still low. The

continued expression of histone H5 in definitive erythroid cells, may be also depend on the transcription factor NF-E4, which is only expressed in definitive cells. Other factors which might influence the ability of H5 to be expressed in the range of cell stages is that the promoter has a strong binding site for Sp1.

Looping Model of Histone H5 Gene Chromatin

The histone H5 enhancer activity has been localized in the region +851 to +1185 in undifferentiated HD3 cells by Rousseau *et al.* (1989) and +851 to +1064 in differentiated HD3 and in HD6 cells, a cell line with higher H5 mRNA levels than several other erythroid precursor cell lines by Trainor *et al.* (1987). The region +1065 to +1363 did not have enhancer activity (Trainor *et al.*, 1987). Our results show that the region in common, +851 to +1064, has two sites with moderate and weak binding affinity (TGGGC^{+939}GGGT and TAGGA^{+966}GGGC, respectively) for an Sp1-like protein and a binding site for GATA-1. In transient transfection studies using histone H5 expressing cells, Rousseau *et al.* (1989) demonstrated that to activate transcription the histone H5 3′ enhancer required the presence of the histone H5 promoter element (-38 to -90) which contains the GC box, a strong Sp1 binding site, and the UPE. Sp1 multimers bound at DNA sites separated by 1.8 kb will interact, resulting in a looping out of the intervening sequences (Pascal and Tjian, 1991; Mastrangelo *et al.*, 1991; Su *et al.*, 1991). We propose that Sp1 multimers bound at the promoter (GGGGC^{-79}GGGG) and enhancer (+939 and/or +966) interact to mediate the juxtapositioning of the enhancer and promoter elements (Figure 38) and that GATA-1 interacts with factors bound at the promoter. A model has also been proposed for the ß-globin gene, whereby the 3′

Figure 38. Model for positioning the 3′ enhancer next to the 5′ promoter through protein-protein interactions. The DNase I protected region at +1205 is indicated by the site of the DNase I protected region. This model suggests that the association of the Sp1 multimers mediates the juxtapositioning of the enhancer and promoter elements, bringing the GATA-1 multisubunit complex (and other factors) next to the initiation site. Adapted from Dr. Michael Hendzel.


Highly Modified Nucleosome
Histone H5 Gene Chromatin
Linker Histone
RNA Polymerase II
RNA
5' Promoter
3' Enhancer
clBF
TFIID
UPE
DR
Sp1
Sp1
Sp1
Sp1/NF-E4
GATA-1 Multisubunit Complex
Sp1
+1205
AP-1
Nuclear Matrix
NF1

enhancer is juxtapositioned next to the promoter elements through protein-protein interactions (Choi and Engel, 1988).

The key elements of the histone H5 gene looping model are the interaction of Sp1 bound in the promoter with Sp1 bound in the enhancer and interactions of GATA-1 with factors bound at the promoter. However, the data show that there were binding sites for other transcription factors in the region +1065 to +1365 and this region in histone H5 gene chromatin of chicken erythrocytes, is DNase I sensitive and has DNase I hypersensitive sites, suggesting that the region is accessible to transcription factors and has bound transcription factors respectively. It is probable that these binding sites are important to histone H5 gene expression *in vivo*. In the transient transfections the lack of a requirement for these sequences may have been because the DNA sequences were not integrated into the genome. Therefore the assays would not show the effects of chromatin and of interaction of the histone H5 chromatin with nuclear matrix. If on the other hand, the *in vitro* model which we suggest is placed into a physiological context, the other factors interacting with the promoter and 3′ enhancer are expected to perform a function.

The effect of the juxtapositioning is to bring the bound GATA-1 multisubunit complex and the other histone H5 gene chromatin bound factors next to the initiation site. The histone H5 3′ enhancer bound factors would therefore be able to affect transcription through direct physical contact. It has been suggested that the function of looping is to funnel the effect of transcription factors, bound at remote enhancers, through the promoter (Ptashne, 1988).

Renaud and Ruiz-Carrillo (1986) observed DNase I sensitivity within the histone H5 gene chromatin. The looping model incorporates the features of altered

chromatin structure observed on transcribed, DNase I sensitive chromatin. These features include core histone hyperacetylation and linker histone phosphorylation, both of which result in histones with less affinity for DNA and which is therefore reflected by an altered more open chromatin structure than bulk chromatin. Amounts of the general transcription repressor linker histone are reduced. Kamakaka and Thomas (1990) suggest that loss of even a few linker histones results in loss of cooperative chromatin condensation. The diagram includes DH sites where nucleosomes are absent and are bound by transcription factors. In the model Sp1 plays a major role in causing an interaction between the enhancer and promoter regions. Factors bound to GATA-1 may also act to bridge the two regions (Fong and Emerson, 1992). The gene chromatin is also diagrammed as interacting with the nuclear matrix which acts as a source of the transcription machinery, since RNA pol II is associated with it.

Within the promoter (Figure 38), the transcription factor binding sites (Sp1, DR, UPE, TFIID and cIBF) are occupied by their respective factors in the DNA looping model. The enhancers located at histone H5 DNase I hypersensitive sites I and III are also not diagrammed in the looping model. The factors binding these sites were not well characterized by Rousseau *et al.* (1993). It has been suggested that erythroid promoters require binding by at least three transcription factors in order to function (Walters and Martin, 1992). GATA, and one of CACCC or AP-1/NF-E2 elements were required for promoter activity of a minimal promoter in erythroid cells (Walters and Martin, 1992). A synergy between GATA-1 and CACCC factor in the human porphobilinogen gene promoter was demonstrated (Frampton et al., 1990). Although the histone H5 gene promoter does not have

binding sites for these elements, loss of the Sp1 binding site at -79 and the UPE element reduced activity of the promoters.

The looping model explains activation over large distances, the orientation independence of enhancers, and provides a mechanism whereby the effect of upstream promoter and enhancer elements can be funnelled through the basic transcription machinery. In this model, the transcription factors bound within the complex also affect transcription. For example, Sp1 will continue to stabilize the formation of productive initiation complex (Kollman et al., 1992). The model does not take into account the function of the UNE, an element which is bound by a ubiquitous factor. The diagram includes the factors identified within this research and additionally the factors bound at sites which were not included in this study but have been demonstrated by Rousseau *et al.* (1993). Some or all of the factors are expected to activate transcription in a synergistic manner as a result of multiple cooperative contacts between the proteins and the general transcription apparatus (Berk and Schmidt, 1990). The multiplicity of Sp1-bound sites would affect transcription synergistically as has been demonstrated at a modified thymidine kinase promoter with a downstream Sp1 binding site 1.8 kb distant (Su et al., 1991).

The role of the nuclear matrix associated factor NF1 would be to assist in the specific association of the nuclear matrix associated RNA pol II and RNA processing functions with the histone H5 gene. Renaud and Ruiz-Carrillo (1986) found that the region downstream of DH 7 (3′ of +1145) was resistant to DNase I digestion and that in run on experiments most of the polymerases engaged on the H5 gene terminate transcription after traversing the DH 7 region. These data may

reflect the proposed role for MAR factors in defining boundaries of transcriptionally active chromatin. Renaud and Ruiz-Carrillo (1986) speculate that DHS 7 together with the more compact chromatin structure downstream from it are involved in the process of transcription termination. When transcription is terminated the newly synthesized RNA would remain attached to the nuclear matrix for final processing and then transport to the cytoplasm (Warren and Cook, 1978). A function which some of the enhancer-bound transcription factors may then be involved in, is to immediately assist in recruiting the newly released RNA pol II and associated factors. In this way the factors could be transferred rapidly back to the initiation site.

A potential function of the factors with which GATA-1 interacts is to provide a bridge between the nuclear matrix and the enhancer. F6 protein is a nuclear matrix protein which is apparently a member of the GATA family (Vassetzky *et al.*, 1993). It's relationship if any to GATA-1 or its associated factors is unknown.

Role of GATA-1 in Transcription

The factors identified in this project which interact with GATA-1 to form multisubunit complexes are presumed to have a function in assisting or modulating GATA-1 action within the nucleus. Factors which may affect DNA-binding transcription factors have been grouped into three categories, adaptors, co-activators and accessory proteins (Martin, 1991). Adaptors form bridges between the transcription factors and the basal transcription machinery to affect transcription. Co-activators act with the transcription factors to affect transcription. Both factors make contact with the basal transcription. Accessory

proteins do not contact the basal transcription machinery, but instead they modulate the activity of transcription factors. Little is known of the mechanism of action of GATA-1. Therefore the roles in transcription played by GATA-1 and the factors of the GATA-1 multi-subunit complexes are not known.

The role of transcription factors may be to recruit or stabilize factors of the RNA pol II initiation complex. Therefore GATA-1 and/or GATA-1 multisubunit complexes may interact with the general transcription apparatus in this manner. The two general transcription factors which are thought to be affected by specific transcription factors are TFIID and TFIIB. GATA-1 complexes may interact with either of the factors. The present data do not rule out the possibility that the proteins associated with GATA-1 may be part of the TFIID complex. There are prior examples of transcription factors such as Sp1 and CTF-1 which make direct contact with factors of TFIID (Dussere and Mermod, 1992). Sp1 acts through the *Drosophila* 110 kDa TAF (TFIID associated factor) in order to affect transcription. It should be noted that the GATA-1 multisubunit complex C8 is greater that 450 kDa and that holo TFIID isolated from a number of species is 500-600 kDa (Dynlacht et al., 1991). In this regard it is also noteworthy that the histone H5 5′ promoter fragment (-350 to +212) but not Sp1 oligonucleotide were effective competitors for GATA-related complex C8 (data not shown). This suggests that either the UPE protein, DR protein, TFIID or clBF are part of the GATA-1 multisubunit complex C8. It has been suggested that a promoter requires at least one factor binding site other than a TATA box in order to respond to enhancers. In transient transfection studies using histone H5 expressing cells, Rousseau *et al.* (1989) demonstrated that to activate transcription the histone H5 3′ enhancer required the presence of the

histone H5 promoter element (-38 to -90) which contains a strong Sp1 binding site, the DR binding site and the UPE but not the TATA box and clBF binding site. An explanation for this observation has been proposed in that Sp1 factors bound at the promoter and the 3' enhancer could interact. This does not preclude an interaction between the UPE or DR bound factor, TFIID or clBF and a 3' enhancer-bound factor. The GATA-1 multisubunit complexes may interact with the UPE or DR bound factor. Competition with the histone H5- promoter, but not Sp1 was efficient at preventing formation of the GATA-1 multisubunit complex C8 in the EMSA. However, there is also data suggesting a transient interaction between GATA-1 and TFIID via an adaptor protein (Fong and Emerson, 1992).

Three activation domains were mapped in chicken GATA-1 (Yang and Evans, 1992). The third activation domain is located at amino acids 227 to 280 (Figure 3). Out of 53 amino acids, 12 are proline. This includes seven consecutive proline residues. It has been demonstrated that the proline-rich activation domain of the transcription factor CTF interacts with TAFs (TATA binding protein associated factors) of TFIID and thereby with the TATA binding protein (Tanese *et al.*, 1991). Similarly, GATA-1 might interact with TAFs. GATA-1 interaction with TFIID via the third activation domain does not preclude GATA-1 interaction with other proteins bound to the histone H5 promoter via either of the other activation domains mapped to GATA-1.

Data from this project which would suggest a role for GATA-1 multisubunit complexes in activation of transcription come from the preliminary study done with the uninduced and induced mouse erythroleukemia (MEL) cells. The uninduced cells did not express mature erythroid proteins and nuclear extracts from these cells did

not form GATA-1 multisubunit complexes in the EMSA. When the MEL cells were induced to differentiate, mature erythroid genes began to be expressed. EMSAs done with nuclear extracts prepared from these cells demonstrated the presence of GATA-1 containing protein complexes. This would suggest that expression of mature erythroid proteins requires GATA-1 multisubunit complexes. A functional assay to test activity of the GATA-1 containing protein complexes needs to be developed in order to prove or disprove this hypothesis. The possible functions of the GATA-1 multisubunit complexes would occur within the framework of the looping model.

GATA-1 Interaction with Proteins of the Multisubunit Complex

Both zinc fingers of GATA-1 proteins are highly conserved. The second finger plays a crucial role in binding DNA. In order to determine the activation domains of chicken GATA-1, mutant proteins were expressed in QT6 cells (Yang and Evans, 1992), which may have provided GATA-1 associated factors. Although the first finger also plays a role in binding to DNA, it has been identified as an activation domain of chicken GATA-1 and murine GATA-1 (Yang and Evans, 1992; Martin and Orkin, 1990). The conserved finger I may be responsible for GATA-1 interactions to form multisubunit protein complexes. The zinc fingers of GATA-1 are conserved in GATA-2 and GATA-3 and therefore may also form multi-subunit complexes. Alternatively, GATA-1 interactions with other proteins may be dependent on activation domains I, III and/or as yet unmapped regions of GATA-1.

GATA-1 Interactions with DNA Binding Sequences

The affinity of GATA-1 and GATA-1 containing protein complexes for the histone H5 3′ GATA binding sequence was compared to the affinity of a number of other GATA binding sequences (Figure 14, Table 1). The strength of binding was alphaG2 > rho, H5-GATA > ß-globin enhancer, ß-globin TATA/GATA. Plumb *et al.* (1989) also noted that the sequence alphaG2 bound mouse GATA-1 with higher affinity than the chicken ß-globin enhancer region, which has two GATA-1 binding sites including the one used here. Ko and Engel (1993) and Engel *et al.* (1992) purified and PCR-amplified sequences which bound GATA-1 in EMSAs, to gain an understanding of the sequences for which GATA-1 has affinity. The PCR-amplified sequence was $^{-2}A/C_{0.81}$ $A^{-1}{}_{0.63}$ GAT $^{+1}AA_{0.86}$ $C_{0.59}$ $A^{+4}{}_{0.59}$ (Engel *et al.*, 1992), where the superscript represents the position of the nucleotide relative to the central GAT as defined by Engel *et al.* (1992) and the subscript indicates the relative frequency of a particular nucleotide occurring at a given position. There are similarities and differences between expected affinities for GATA-1 of the sequences used in this study when compared to the PCR-amplified sequences.

The data of Omichinski *et al.* (1993b), which showed that the DNA-specific carboxy-terminal finger of GATA-1 did not interact with the DNA before -1 or beyond the +2 position, suggests that outside of -1 and +2 changes in DNA sequence should not have altered the binding of any of the oligonucleotides in the competition assays. However, the data of Ko and Engel (1993) suggest that sequence context is important. Yang and Evans (1993) demonstrated that methylation of nucleotides -4, -2, -1, G and A of GAT, +1, +2 and +4 interfered with GATA-1 binding to its cognate site. Thus finger I of GATA-1, which has been

shown to affect stability of DNA binding (Omichinski *et al.*, 1993a) may not only contact the nucleotides between positions -1 to +2, but additionally finger I may contact nucleotides outside of -1 to +2.

At positions +1 and +2 the sequences of all of the oligonucleotides conformed to the PCR-amplified sequences. At the -1 position the ß-globin TATA/GATA promoter sequence and the αG2 (mouse) sequences differed. The data of Ko and Engel (1993) showed that a T at the -1 position resulted in higher affinity binding, whereas the C at this position was allowed for binding by GATA-1 but did not result in greater binding. The data presented here taken with the data of Ko and Engel (1993) and Omichinski *et al.* (1993b) suggest that the C at position -1 of the ß-globin promoter TATA/GATA sequence may have affected the affinity of the sequence for GATA-1. It should also be noted that in the binding assay, TFIID/adaptor was present which also binds the oligonucleotide. This would effectively remove an amount of the oligonucleotide from solution and therefore affect the amount of the DNA available for competition with the H5-GATA for the GATA-1 and GATA-1 containing protein complexes.

The ß-globin enhancer (IV) sequence did not compete as effectively for binding of the GATA-1 and GATA-1 complexes as H5-GATA, alphaG2 and the rho sequence. The ß-globin enhancer (IV) sequence did not have the A, which was preferred at position +4, or the C preferred at position +3. Two of the oligonucleotides, the rho-globin promoter sequence and the alphaG2 sequence, had a G at the +4 position and also had relatively strong binding to GATA-1 complexes when compared to the other GATA-binding oligonucleotides. Both of these oligonucleotides also have a G which was not preferred at the +3 position. This is

contrary to the data of Ko and Engel (1993) which showed that GATA-1 preferred an A at the +4 position and a C at +3. These results may reflect the finding of Ko and Engel (1993) that sequence context is important for GATA-1 binding.

Methylation at -4 affects GATA-1 binding, however, Ko and Engel (1993) did not determine sequence preference at -4. The rho globin promoter had a G at this position, which may have assisted in a slightly higher affinity for GATA-1. There was not a clear effect for other nucleotides at this position, since H5-GATA and the ß-globin enhancer sequence each had a T at this position. The TATA/GATA and the alphaG2 sequences each had an A at this position. For each of these pairs of examples, one of the oligonucleotides had significantly more affinity for GATA-1 than the other.

These data overall demonstrate that progress has been made with respect to determining sequences with which GATA-1 is able to interact. However, the rules have not been clearly delineated to determine with which sequences GATA-1 will interact with high affinity. The nucleotides which create a strong GATA-1 binding site will do so by creating a better fit between GATA-1 and the nucleotides present in the DNA. The better fit may be created through specific DNA-GATA-1 interactions or by allowing the conformation of the DNA to be altered so that GATA-1 is better able to interact with it. For example, certain nucleotide sequences allow DNA to bend readily. Nucleotide sequence may play a role in the bending which occurs at the +1 to +2 position. It should be noted that in each of the competitions in Figure 14, a given competitor DNA affected GATA-1 containing protein complex binding to the same degree as it affected monomeric GATA-1 binding to the sequence.

GATA-1 is the only Protein of the GATA-1 Multisubunit Complexes that Interacts with DNA

These data provide evidence that GATA-1 is the only protein of the complexes that interacts with DNA and that the heterologous proteins do not affect GATA-1 binding to DNA. 1) The proteins which were cross-linked did not differ between the C2 (GATA-1 alone) and C5 or C8 enriched complexes. It would be expected that if proteins other than GATA-1 in the larger complexes made contact with the DNA, that they should become cross-linked by the ultraviolet light and would be seen on the SDS-PAGE gels. This was not seen when either the template or non-template strand was labelled. 2) The preparative DNase I protection assays, did not show a difference in DNase I protected patterns for each of the complexes. If the non-GATA-1 proteins of the GATA-1 protein complexes interact with DNA or affect GATA-1 binding to DNA, the DNase I protected region would have been expected to be different for the multisubunit complexes when compared to the DNase I protected region of GATA-1 alone. 3) Binding of the higher complexes was not affected by changes in the DNA sequence outside of the GATA binding site. This was illustrated by two experiments. The first, an experiment which demonstrated that the pattern of complexes which formed on H5-GATA was the same as the pattern formed on a number of oligonucleotides containing GATA binding sites (Figure 9, 18B). These oligonucleotides each have a GATA binding site, but have very different sequences outside of this core (Table 1). The second experiment which demonstrates that only GATA-1 interacts with DNA is that the aforementioned GATA binding oligonucleotides, when used in competition assays did not preferentially prevent binding of any of the GATA-1 mutlisubunit complexes

over GATA-1 (Figure 18A). If any of the non-GATA-1 proteins of the multisubunit complexes interacted with DNA, they would have been expected to show a DNA sequence preference. Both the EMSAs and the competition assays do not show this. Therefore our hypothesis is that GATA-1 alone interacts with DNA and the proteins which interact with GATA-1 to generate the multisubunit complexes do not.

Role of Histone H5 Binding Factors in DNase I Hypersensitive Site Formation

In chicken immature erythrocyte chromatin, DNase I hypersensitive (DH) site 5 maps at -117 ± 6 and DH site 5_L maps at -25 ± 6 in the promoter region (Renaud and Ruiz-Carrillo, 1986). Thus, the DH site 5 is located immediately 5′ to the GC box, while DH site 5_L co-maps with the protected regions II (the UPE) and III (TATA box) (Figure 37). It has been suggested that factors bound to enhancers physically interact with promoters to assist in clearing promoters of bound nucleosomes and thereby creating DH sites (Reitman et al., 1993). In the looping model we propose, Sp1 bound at the strong Sp1 binding site in the histone H5 promoter interacts with the enhancer which has a medium and several weak Sp1 binding sites.

One of the functions of upstream transcription factors may be to remove nucleosomes from the TATA box region in order to allow general transcription factors to interact with the promoter (Berk and Schmidt, 1990). Reitman *et al.* (1993) have demonstrated that in the ß-globin gene, both the promoter and the 3′ enhancer/LCR are required for preventing chromatin condensation. They propose a mutual physical interaction between the two regions similar to this model for histone H5. In the histone H5 gene chromatin looping model, Sp1 may play a direct

role in keeping the chromatin open. Sp1 not only activates transcription, but in cooperation with cofactors it also acts as an anti-repressor and displaces the general repressor histone H1. Given the similarities between the linker histones H1 and H5, Sp1 may also displace histone H5. In the linker histone rich environment of the maturing erythrocyte this may be an important role of Sp1, and which results in preventing repression of the gene chromatin. Other factors may prevent deposition of nucleosomes at the promoter and enhancer sites.

In the histone H5 enhancer, DHS 7_L ($+1003 \pm 12$) and 7_U($+1095 \pm 12$) co-map with DNase I protected regions III and V, respectively (Figure 37). Lowrey *et al.* (1992) demonstrated that interactions between ubiquitous factors (e.g. Sp1) and erythroid-specific nuclear proteins (e.g., GATA-1) were required to form a DNase I hypersensitive site within the human globin locus control region. The binding sites for GATA-1 and Sp1 (e.g., at CACCC; Yu *et al.*, 1991) are commonly found in the promoter and enhancers of globin and other erythroid-specific genes (e.g. porphobilinogen deaminase gene) (Yu *et al.* 1991, Frampton *et al.*, 1990; Evans *et al.*, 1990). Thus, Sp1 and GATA-1 may be involved in the formation of hypersensitive sites in chromatin within the histone H5 enhancer as well as the globin enhancer and promoter regions. It should be noted the CACCC sequence found at DNase I protected region IV is a weak Sp1 binding site, and this sequence may bind to other proteins such as the CACCC factor.

Developmental and Stage Specific Control of Histone H5 Gene Expression

It had been previously assumed that the levels of Sp1 are at similar levels throughout all stages of erythroid development (Jackson *et al.*, 1989). Our data

showed this was not the case, mature cells had lower levels of transcription factors Sp1 and GATA-1 and multi-subunit GATA-1 complexes than immature erythrocytes. NF1 was at similar amounts in mature and immature erythrocytes. The DR and UPE-binding protein were also at similar amounts in mature and immature erythrocytes, although Rousseau *et al.* (1993) determined that deletion of the UPE-binding sequence inactivated the promoter in non-differentiated HD3 cells, but had no effect in differentiated HD3 cells. This suggests that the UPE-binding protein affects transcription early in the differentiation of erythrocytes and may still be present, but is not functional as the cell matures.

The physiological implication of the reduction of GATA-1 multisubunit complexes and of Sp1 at regions I and III in the enhancer in mature erythrocytes is that in the looping model the positioning and function of the histone H5 enhancer could be seriously compromised. The low abundance of these proteins would also hinder the expression of the ß-globin gene (Evans *et al.*, 1990; Minie *et al.*, 1992a; 1992b). The histone H5 gene has a UNE which represses transcription of the gene in both non-differentiated and differentiated HD3 cells (Rousseau et al., 1993). Chicken mature erythrocytes contain the factor cIBR which binds to the transcription start site of the histone H5 gene and represses its expression (Gomez-Cuadrado *et al.*, 1992). cIBR is not present in early erythroid cells that express the histone H5 gene (Gomez-Cuadrado *et al.*, 1992). Further, the inactivation of the general transcription factor TFIIA may also contribute to the inactivation of the histone H5 in mature erythrocytes (Bungert *et al.*, 1992b). Together, these observations suggest that several mechanisms are involved in the specific repression of the histone H5 gene (as well as other erythroid-specific genes;

Jackson *et al.*, 1989) in mature erythrocytes.

Minie *et al.* (1992) demonstrated that GATA-1 and Sp1 levels decrease in chicken embryos after day 5, during which time the rho-globin gene ceases to be transcribed. Our data demonstrated that the GATA-1 multisubunit complexes were present at this early stage of development. This suggests that the ability of GATA-1 to form multisubunit complexes does not affect rho-globin expression throughout erythropoiesis.

Partial Purification of Protein of GATA-1 Containing Protein Complexes

Evans and Felsenfeld (1989) purified GATA-1. They did so using a number of chromatographic steps followed by electrophoresis on a denaturing polyacrylamide gel to separate the proteins of the final chromatographic step. Their first step was to concentrate proteins of crude nuclear extract derived from adult erythrocytes by ammonium sulphate precipitation. The protein was then chromatographed on double-stranded calf-thymus DNA-cellulose. A second chromatographic separation using a single-stranded DNA-cellulose column was performed. This step removed a major contaminant, a heteronuclear RNA protein with a molecular mass of 39 kDa which is the size of GATA-1. The last chromatographic step was achieved by passage over an affinity column consisting of concatamers of an oligonucleotide duplex containing GATA-1 binding site from the chicken ß-globin enhancer. Active fraction from this step was separated on a denaturing polyacrylamide gel. Individual protein bands were excised, eluted and renatured. GATA-1 activity was shown to be present as a 39 kDa protein.

We wished to purify GATA-1 containing protein complexes instead of GATA-

1. We were interested in the nature of the proteins which interacted with GATA-1. Specifically the complex we sought was complex C5. This complex was abundant and was the least complicated of the complexes other than the less abundant complex C4. Similar to Evans and Felsenfeld, extracts of mature chicken erythrocyte nuclei were used, however only C2 (monomeric GATA-1) activity was present after partial purification under conditions which preserved the integrity of complex C5 prepared from adult chicken immature erythrocytes. We do not know why this occurred and the experiment was not repeated. Therefore, the chromatographic separations were done using adult chicken immature erythrocytes. A number of different columns were used for the partial purification of complex C5. Heparin-agarose was used as the first column with activity eluting in the 0.4 M KCl fraction. This was followed by an S-sepharose column, which was eluted at 0.3 and 0.5 M NaCl after a 0.2 M NaCl wash. These columns both select for positively charged proteins. The 0.5 M fraction was used for subsequent fractionations, however, because the 0.3 M fraction had less total protein, it may be possible to purify complex C5 from this fraction. The 0.3 M NaCl fraction of S-sepharose was chromatographed over a phenyl-agarose column once, without successful recovery of activity. The fractionation might be repeated with altered conditions. Most of the GATA-1 containing protein complexes had little affinity for Q-sepharose, and were eluted in the flowthrough or displaced by 0.1 M NaCl, although a small portion of the C5 complex was displaced at 0.5 M NaCl. This suggests that there was little negative charge in the complexes and in GATA-1 which is known to be basic (Evans and Felsenfeld, 1989). Phenyl-agarose was frequently used as the second or third column for purification of GATA-1 containing protein complexes.

This column separated proteins on the basis of hydrophobicity. Complex C5 was separated from complexes C2 (GATA-1 monomer) and C8 in about 5% of the starting material. When the proteins from this fraction were analyzed on denaturing polyacrylamide gels, there were fewer proteins present than in the fractions from previous columns. A 54 kDa protein was enriched in the fraction with C5 activity.

Other columns used were double-stranded and single-stranded DNA-cellulose. The active protein fraction from these columns had less protein complexity than the starting material. The single-stranded DNA cellulose column was better for this than the double-stranded DNA-cellulose column. However, when the column steps were heparin-agarose to phenyl-agarose to double-stranded DNA-cellulose to single-stranded DNA cellulose, pure C5 was not obtained, since there was a fair number of proteins present, when the active fractions were analyzed by denaturing polyacrylamide gel electrophoresis (not shown).

Affinity column chromatography using concatamers of oligonucleotide bound to sepharose beads did not prove useful for decreasing the complexity of the proteins. A reason for this might be that the 0.3 M KCl step used by Evans and Felsenfeld (1989) was omitted, since a lot of GATA-1 activity eluted at this salt concentration. In retrospect, the 0.3 M KCl elution might have proved useful to rid the column of nonspecific proteins, since GATA-1 containing protein complex activity was also seen at 0.4 M KCl. It should however be noted that Evans and Felsenfeld (1989) used extract of adult chicken mature erythrocytes. In these cells transcriptional activity is generally lower than in adult chicken immature erythrocytes. Therefore, the starting protein complexity is different for mature erythrocytes than it is for immature erythrocytes. It should however be possible to

purify complex C5 in the future, using the fractionations described here as a building block for further purification.

Future Experiments

The results described in this research are from *in vitro* experiments. These data were useful for showing interactions which are possible on the histone H5 gene. However, it does not demonstrate the effect of the erythroid environment and histone H5 gene chromatin structure on accessibility of the transcription factors to the gene. To gain an understanding of the physiological interactions on the gene, *in vivo* DNase I protection assays of the H5 gene chromatin are required. *In vivo* DNase I protection assays will demonstrate whether the potential binding sites identified in this study are bound by factors within the chromatin structure of the histone H5 gene.

A number of potential factor binding sties which we identified were able to bind a number of related or unrelated transcription factors and/or transcription factors which were poorly characterized. Many of these factors have not been characterized well enough to be able to distinguish the particular factor which will interact with a specific DNA sequence. This fact is even more important when the binding sites are considered within the context of other transcription factors and chromatin present under physiological conditions in the cell. In order to fully understand the mechanics which occur in order to activate a gene, the particular transcription factors binding to the gene and the effects of the transcription factor *in vivo* need to be understood. The looping model which we propose needs to be tested. Electron micrograph techniques exist to demonstrate physical interactions

between Sp1 factors bound at distant sites (Su *et al.*, 1991; Mastrangelo *et al.*, 1991).

Little is known of the mechanism of action of chicken GATA-1 on transcription and on chromatin structure. We determined that the GATA-1 multisubunit complexes were able to form on a number of GATA-1 binding sequences. Therefore, the effect of GATA-1 multisubunit complexes on the histone H5 gene may also be found on other genes. The major GATA-1 containing protein complexes (C5 and C8) need to be purified and tested in defined transcription assays against monomeric GATA-1, to see if the complexes affect the amount of final RNA product. These assays would then allow the mechanism of action of GATA-1 and GATA-1 multisubunit complexes to be tested. It could be determined whether the factors increase the rate of formation or stability of initiation complex. Concurrently, it should be possible to determine the target of GATA-1 action. The two general transcription factors which are known to interact with certain transcription factors are TFIID and TFIIB. These are the factors which would be tested initially.

There have been suggestions that GATA-1 affects chromatin structure. For example GATA-1 in conjunction with CACCC factors or NF-1 are thought to form DH sites (Lowrey et al., 1992). This implies that the factors either displace nucleosomes actively (dynamic competition) or prevent formation of nucleosomes at the site during DNA replication (pre-emptive). *In vitro* assays could be employed to determine if GATA-1 or GATA-1 multisubunit complexes are 1) able to displace nucleosomes from DNA either alone or in concert with other factors when binding sites for the factors are present on the DNA or 2) require the factors bound to the

DNA followed by nucleosome binding for DH site formation.

Determining functions for the factors of the GATA-1 multi-subunit complexes should help to answer the question whether these complexes exist *in vivo*. It can be ruled out that the complexes are an artifact of the 1-acetyl-2-phenylhydrazine injections since comparable complexes were present in the nuclei of chicken embryo erythrocytes which did not receive any treatment prior to collecting erythrocytes from the embryos.

Factors that could affect binding of GATA-1 to accessory proteins could include post-translational modifications to the accessory proteins. The evidence (Evans and Felsenfeld, 1989) is that GATA-1 is not post-synthetically modified. However, it is possible that there are post-synthetic modifications to the GATA-1 accessory proteins and that these modifications affect multisubunit complex formation. Evidence that modifications occur comes from the column fractionation studies, in which GATA-1 containing protein complexes eluted over a number of salt concentrations from the Q-sepharose column, from the double-stranded and single-stranded DNA cellulose columns and from the S-sepharose column.

Cloning of the proteins of the GATA-1 multisubunit complexes will allow further characterization of the nature of the proteins, their interactions and effects on GATA-1 function. Deletion mutants of GATA-1 and of these proteins could be synthesized to determine the site and nature of interaction of these proteins. These mutants could also be used to show the effect of the proteins on cell function. The tissues and species specificity of the factors could also be determined to show whether the factors are ubiquitous.

Evans and Felsenfeld (1991) found that the chicken alpha-globin promoter

can be activated in chicken fibroblast cells by expression of a transfected GATA-1 coding plasmid. This would suggest that either the accessory proteins that interact with GATA-1 are found in more than one cell type (fibroblast cells), or that the accessory factors were not needed for activation in fibroblast cells. Another potential explanation is that the accessory proteins modulate GATA-1 activity. If they increase activity of GATA-1, then Evans and Felsenfeld (1991) may not have seen the true activation which GATA-1 is capable of. If on the other hand the accessory proteins decrease the activity of GATA-1, this may explain why Evans and Felsenfeld (1991) and other workers (Martin and Orkin, 1990; Frampton *et al.*, 1990; Walters and Martin, 1992; deBoer *et al.*, 1988) have found that GATA-1 binding sites alone are not capable of activating transcription in erythroid cells.

In summary this research has demonstrated potential binding sites in the chicken histone H5 promoter and 3′ enhancer for proteins of adult chicken immature erythrocytes. These data show that Sp1 and GATA-1, both of which form multisubunit proteins complexes as well as a number of other transcription factors were able to interact with the DNA. We proposed a model whereby Sp1 is responsible for juxtapositioning the enhancer and promoter. The proteins bound at these sites are then able to interact with the final result that the enhancer bound transcription factors are able to affect transcription from the promoter. The model was placed in the context of chromatin structure and in the chicken erythrocyte environment. These data suggest that this looping model would be compromised in nuclei of mature erythrocytes due to a decrease in Sp1, GATA-1 and GATA-1 multisubunit complexes.

References

Whole cell extracts which were prepared using buffers containing HEPES
! Nuclear extracts which were prepared using buffers which contained HEPES
* Nuclear extracts which were prepared using buffer D, a HEPES containing buffer
^ Nuclear extracts which were prepared using the Dignam method, in which there a dialysis against buffer D

Abmayr, S.M., Workman, J.L. and Roeder, R.G. (1988) The pseudorabies immediate early protein stimulates *in vitro* transcription by facilitating TFIID: promoter interactions. Genes Dev. **2**:542-553

Adamson, J.W. (1968) Erythropoietin / hematocrit relationship in normal and polycythemic man-implications of marrow regulation. Blood **32**:597-609

Affolter, M., Cote, J., Renaud, J. and Ruiz-Carrillo, A. (1987) Regulation of histone and $ß^A$-globin expression during differentiation of chicken erythroid cells. Mol. Cell. Biol. **7**:3663-3672

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D. (1989) Molecular Biology of the Cell. Garland Publishing Inc., New York

Alemany, J., Klement, J.F., Borras, T. and DePablo, F. (1992) DNA-binding factors which interact with the SP1 site of the chicken-delta-1-crystallin promoter are developmentally regulated. Biochem. Biophys. Res. Commun. **183**:659-665

Allan, J., Harborne, N., Rau, D.C. and Gould, H. (1982) Participation of core histone tails in the stabilization of the chromatin solenoid. J. Cell Biol. **93**:285-297

Allan, J.D., Cowling, G.J., Harborne, N., Cattini, P., Craigie, R. and Gould, H. (1981) Regulation of the higher-order structure of chromatin by histones H1 and H5. J. Cell Biol. **90**:279-288

Allan, J.D., Hartman, P.G., Crane-Robinson, C. and Avilés, F.X. (1980) The structure of histone H1 and its location in chromatin. Nature **288**:675-679

Allegra, P., Sterner, R., Clayton, D.F. and Allfrey, V.G. (1987) Affinity chromatographic purification of nucleosomes containing transcriptionally active DNA sequences. J. Mol. Biol. **196**:379-388

Allfrey, V.G. (1977) in Chromatin and Chromosome Structure (Li, H.J. and Eckhardt, R.A., eds.), Academic Press, New York. pp. 167-191

Allfrey, V.G. (1971) in Histones and Nucleohistones (Phillips, D.M.P., ed.), Plenum Publishing Co. London. pp.241-294

Allfrey, V.G., Faulkner, R. and Mirsky, A.E. (1964) Acetylation and methylation of histones and their possible role in regulation of RNA synthesis. Proc. Natl. Acad. Sci. USA. **51**:786-794

Allison, L.A., Wong, K.-C., Fitzpatrick, V.D., Moyle, M. and Ingles, C.J. (1988) The C-terminal domain of the largest subunit of RNA polymerase II of *Saccharomyces cerevisiae, Drosophila melanogaster* and mammals: a conserved structure with an essential function. Mol. Cell. Biol. **8**:321-329

Almer, A., Rudolph, H., Hinnen, A. and Horz, W. (1986) Removal of positioned nucleosomes from the yeast PHO5 promoter upon PHO5 induction releases additional upstream activating DNA elements. EMBO J. **5**:2689-2696

Anderson, G.M. and Freytag, S.O. (1991) Synergistic activation of a human promoter *in vivo* by transcription factor Sp1. Mol. Cell. Biol. **11**:1935-1943

Andrews, N.C., Erdjument-Bromage, H., Davidson, M.B., Tempst, P. and Orkin, S.H. (1993) Erythroid transcription factors NF-E2 is a haematopoietic-specific basic-leucine zipper protein. Nature **362**:722-728

Annunziato, A.T., Frado, L.-L.Y., Seale, R.L. and Woodcock, C.L.F. (1988) Treatment with sodium butyrate inhibits the complete condensation of interphase chromatin. Chromosoma **96**:132-138

! Aplan, P.D., Nakahara, K., Orkin, S.H. and Kirsch, I.R. (1992) The SCL gene product: a positive regulator of erythroid differentiation. EMBO J. **11**:4073-4081

Appels, R. and Wells, J.R.E. (1972) Synthesis and turnover of DNA-bound histones during maturation of avian red blood cell. J. Mol. Biol. **70**:425-434

Appels, R., Bolund, L. and Ringertz, R.N. (1974) Biochemical analysis of reactivated chick erythrocyte nuclei isolated from chick HeLa heterokaryons. J. Mol. Biol. **85**:182-190

Archer, T.K., Cordingley, M.G. Wolford, R.G. and Hager, G.L. (1991) Transcription factor access is mediated by accurately positioned nucleosomes on the mouse mammary tumor virus promoter. Mol. Cell. Biol. **11**:688-698

Aumont, F.L., Trudel, P. and Wall, L. (1993) Murine erythroleukemia cells contain two distinct GATA-binding proteins that have different patterns of expression during cellular differentiation. Differentiation **52**:169-176

Ausio, J., Dong, F. and van Holde, K.E. (1989) Use of selectively trypsinized nucleosome core particles to analyze the role of the histone "tails" in the stabilization of the nucleosome. J. Mol. Biol. **206**:451-464

Ausio, J. and van Holde, K.E. (1986) Histone hyperacetylation: its effects on nucleosome conformation and stability. Biochemistry **25**:1421-1427

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (1987) Current Protocols in Molecular Biology. Greene/Wiley, New York.

Aviles, F.J., Chapman, G.E., Kneale, G.G., Crane-Robinson, C. and Bradbury, E.M. (1979) The conformation of histone H5. Eur. J. Biochem. **88**:363-371

Ayer, D.E. and Dynan, W.S. (1990) A downstream-element-binding factor facilitates assembly of a functional preinitiation complex at the simian virus 40 major late promoter. Mol. Cell. Biol. **10**:3635-3645

Bading, H. (1988) Determination of the molecular weight of DNA-bound protein(s) responsible for gel electrophoretic mobility shift of linear DNA fragments exemplified with purified viral myb protein. Nucl. Acids Res. **16**:5241-5248

Ball, D.J., Gross, D.S. and Garrard, W.T. (1983) 5-methylcytosine is localized in nucleosomes that contain histone H1. Proc. Natl. Acad. Sci. USA **80**:5490-5494

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**:347-59

Barnhart, K.M., Kim, C.G. and Sheffery, M. (1989) Purification and characterization of an erythroid cell-specific factor that binds the murine α- and ß-globin genes. Mol. Cell. Biol. **9**:2602-2614

Barnhart. K.M., Kim, C.G., Banerji, S.S. and Sheffery, M. (1988) Identification and characterization of multiple erythroid cell proteins that interact with the promoter of the murine alpha-globin gene. Mol. Cell. Biol. **8**:3215-3226

Baron, M.H. and Maniatis, T. (1986) Rapid reprogramming of globin gene expression in transient heterokaryons. Cell **46**:591-602

Bartholomew, B., Dahmus, M.E. and Meares, C.F. (1986) RNA contacts subunits IIo and IIc in HeLa RNA polymerase II transcription complexes. J. Biol. Chem. **261**:14226-14231

Bartolomei, M.S., Halden, N.F., Cullen, C.R. and Corden, J.L. (1988) Genetic analysis of the repetitive carboxyl-terminal domain of the largest subunit of mouse RNA polymerase II. Mol. Cell. Biol. **8**:330-339

Bates, D.L. and Thomas, J.O. (1981) Histones H1 and H5: one or two molecules per nucleosomes? Nucl. Acids Res. **9**:5883-5894

Bennet, J.P., Brotherton, T.W. and Ginder, G.D. (1989) Nuclear matrix attachment regions of the chicken beta-type globin gene cluster. In Hemoglobin Switching Part B: Cellular and Molecular Mechanisms. ed. G. Stamatoyannopoulos. A.W. Nienhuis. pp 47-55. New York: Liss

Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Ovalbumin gene-sequence of putative control regions. Nucl. Acids Res. **8**:127-142

Berger, S.L., Cress, W.D., Cress, A., Triezenberg, S.J. and Guarente, L. (1990) Selective inhibition of activated but not basal transcription by the acidic activation domain of VP16: Evidence for transcriptional adaptors. Cell **61**:1199-1208

Bergman, M. and Ringertz, N. (1990) Gene expression pattern of chicken erythrocyte nuclei in heterokaryons. J. Cell. Sci. **97**:167-175

Bergman, M.G., Wawra, E. and Winge, M. (1988) Chicken histone H5 inhibits transcription and replication when introduced into proliferating cells by microinjection. J. Cell. Sci. **91**:201-209

Berk, A.J. and Schmidt, C.M. (1990) How do transcription factors work? Genes Dev. **4**:151-155

Berrios, M., Osherhoff, N. and Fisher, P.A. (1985) In situ localization of DNA topoisomerase II, a major polypeptide component of the *Drosophila* nuclear matrix fraction. Proc. Natl. Acad. Sci. USA **82**:4142-4146

^Berry, M., Grosveld, F. and Dillon, N. (1992) A single point mutation is the cause of the Greek form of hereditary persistence of fetal haemoglobin. Nature **358**:499-502

Beug, H., von Kirchbach, A., Doederlein, G., Conscience, J. and Graf, T. (1979) Chicken hematopoetic cells transformed by seven strains of defective avian leukemia viruses display three distinct phenotypes of differentiation. Cell **18**:375-390

Beug, H., Palmieri, S., Freudenstein, C., Zentgraf,H. and Graf, T. (1982) Hormone-dependent terminal differentiation *in vitro* of chicken erythroleukemia cells transformed by ts mutants of avian erythroblastosis virus. Cell **28**:907-919

Blau, H.M. and Baltimore, D. (1991) Differentiation requires continuous regulation - editorial. J. Cell Biol. **112**:781-783

Blau, H.M., Chiu, C.-P. and Webster, C. (1983) Cytoplasmic activation of human nuclear genes in stable heterocaryons. Cell **32**:1171-1180

Bode, J. and Maass, K. (1988) Chromatin domain surrounding the human interferon-beta gene as defined by scaffold-attached regions. Biochemistry **27**:4706-4711

Bode, J., Gomez-Lira, M.M. and Schroter, H. (1983) Nucleosomal particles open as the histone core becomes hyperacetylated. Eur. J. Biochem. **130**:437-445

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantitites of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**:248-254

Brady, H.J.M., Sowden, J.C., Edwards, M., Lowe, N. and Butterworth, P.H.W. (1989) Multiple GF-1 binding sites flank the erythroid specific transcription unit of the human carbonic anhydrase I gene. FEBS Letts. **257**:451-456

Brasch, K., Adams, G.H.M. and Neelin, J.M. (1974) Evidence for erythrocyte-specific histone modification and structural changes in chromatin during goose erythrocyte maturation. J. Cell Sci. **15**:659-677

Breant, B., Huet, J., Sentenac, A and Fromageot, P. (1983) Analysis of yeast RNA - Polymerases with subunit-specific antibodies. J. Biol. Chem. **258**:11968-11973

Breathnach, R. and Chambon, P. (1981) Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349-383

Briand, G., Kmiecik, D., Sautiere, P., Wouters, D., Borie-Loy, O., Biserte, G., Mazen, A. and Champagne, M. (1980) Chicken erythrocyte histone H5. FEBS Lett. **112**:147-153

Briegel, K., Lim, K.-C., Plank, C., Beug, H., Engel, J.D. and Zenke, M. (1993) Ectopic expression of a conditional GATA-2/estrogen receptor chimera arrests erythroid differentiation in a hormone-dependent manner. Genes Dev. **7**:1097-1109

Brown, J.L. and Ingram, V.M. (1974) Structural studies on chicken embryonic hemoglobins. J. Biol. Chem. **249**:3960-3972·

Bruns, G.A. and Ingram, V.M. (1973) The erythroid cells and hemoglobins of the chick embryo. Phil. Trans. R. Soc. London B **266**:225-305

Bungert, J., Kober, I., During, F. and Seifart, K.H. (1992a) Transcription factor eUSF is an essential component of isolated transcription complexes on the duck histone H5 gene and it mediates the interaction of TFIID with a TATA-deficient promoter. J. Mol. Biol. **223**:885-898

Bungert, J., Waldschmidt, R., Kober, I. and Seifart, K.H. (1992b) Transcription factor IIA is inactivated during terminal differentiation of avian erythroid cells. Proc. Natl. Acad. Sci. USA **89**:11678-11682

Bunick, D., Zandomeni, R., Ackerman, S. and Weinmann, R. (1982) Mechanism of RNA polymerase II-specific initiation of transcription *in vitro*: ATP required and uncapped runoff transcripts. Cell **29**:877-886

Buratowski, S., Hahn, S., Guarent, L. and Sharp, P.A. (1989) Five intermediate complexes in transcription initiation by RNA polymerase II. Cell **56**:549-561

Buratowski, S., Hahn, S., Sharp, P.A. and Guarent, L. (1988) Function of a yeast TATA element-binding protein in a mammalian transcription system. Nature **334**:37-42

Burgess, A. and Nicola, N. (1983) Growth factor and stem cells. Academic Press, Sydney.

Burton, Z.F., Ortolan, L.G. and Greenblatt, J. (1986) Proteins that bind to RNA polymerase II are required for accurate initiation of transcription at the adenovirus 2 major late promoter. EMBO J. **5**:2923-2930

Butler, P.J.G. (1988) in Chromosomes and Chromatin, Vol I, (Adolph, K. W., ed.), pp. 7-82 CRC Press, Pica Raton, FL

Cadena, D. and Dahmus, M.E. (1987) Messenger RNA synthesis in mammalian cells is catalyzed by the phosphorylated form of RNA polymerase II. J. Biol. Chem. **262**:12468-12474

Cai, H. and Luse, D.S. (1987) Transcription initiation by RNA polymerase II *in vitro*. Properties of preinitiation, initiation, and elongation complexes. J. Biol. Chem. **262**:298-304

Carcamo, J., Buckbinder, L. and Reinberg, D. (1991) The initiator directs the assembly of a transcription factor IID-dependent transcription complex. Proc. Natl. Acad. Sci. USA **88**:8052-8056

Carcamo, J., Lobos, S., Merino, A., Buckbinder, L., Weinmann, R., Nataragan, V. and Reinberg, D. (1989) Factors involved in specific transcription by mammalian RNA polymerase II. Role of factors IID and MLTF in transcription for the Adenovirus major late and IVa2 promoters. J. Biol. Biol. Chem. **264**:7704-7714

Carey, M., Leatherwood, J. and Ptashne, M. (1990a) A potent GAL4 derivative activates transcription at a distance *in vitro*. Science **247**:710-712

Carey, M., Lin, Y.-S., Green, M.R. and Ptashne, M. (1990b) A mechanism for synergistic activation of a mammalian gene by GAL4 derivatives. Nature **345**:361-364

Carrol, S.B. and Stollar, B.D. (1982) Inhibitory monoclonal-antibody to calf thymus RNA polymerase II blocks formation of enzyme-DNA complexes. P. Nas. Biol. **79**:7233-7237

Carthew, R.W., Samuels, M. and Sharp, P.A. (1988) Formation of transcription preinitiation complexes with an amanitin-resistant RNA polymerase II. J. Biol. Chem. **263**:17128-17135

^Catala, F., deBoer, E., Habets, G. and Grosveld, F. (1989) Nuclear protein factors and erythroid transcription of the human A gamma-globin gene. Nucl. acids Res. **17**:3811-3827

Catarina, J.J., Ryan, T.M., Pawlik, K.M., Palmiter, R.D., Brinster, R.L., Behringer, R.R. and Townes, T.M. (1991) Human ß-globin locus control region: analysis of the 5' DNase I hypersensitive site HS2 in transgenic mice. Proc. Natl. Acad. Sci. USA **88**:1626-1630

Caubet, J.-F., Mitjavila, M.-T., Dubart, A., Roten, D., Weil, S.C. and Vainchenker, W. (1989) Expression of the c-*fos* protooncogene by human and murine erythroblasts. Blood **74**:947-951

Cavallini, B., Faus, I., Matthes, H., Chipoulet, J.M., Winsor, B., Egly, J.-M. and Chambon,

P. (1989) Cloning of the gene encoding the yeast protein BTF1Y which can substitute for the human TATA box-binding factor. Proc. Natl. Acad. Sci. USA **86**:9803-9807

Cavallini, B., Huet, J., Plassat, J., Sentenac, A., Egly, J. and Chambon, P. (1988) A yeast activity can substitute for the HeLa cell TATA box factor. Nature **334**:77-80

Chambon, P. (1975) Eukaryotic nuclear-RNA polymerase - review. Annu. Rev. Biochem. **44**:613-638

Chapman, B.S. and Tobin, A.J. (1979) Distribution of developmentally regulated hemoglobins in embryonic erythroid populations. Dev. Biol. **69**:375-387

Chen, T.A. and Allfrey, V.G. (1987) Rapid and reversible changes in nucleosome structure accompany the activation, repression and superinduction of murine fibroblast protooncogenes c-*fos* and c-*myc*. Proc. Natl. Acad. Sci. USA **84**:5252-5256

^Chiba, T., Ikawa, Y. and Todokoro, K., (1991) GATA-1 transactivates erythropoietin receptor-mediated receptor gene, and erythropoietin receptor-mediated signals enhance GATA-1 gene expression. Nucleic Acids Res. **19**:3843-3848

Chiu, C.-P. and Blau, H.M. (1984) Reprogramming cell differentiation in the absence of DNA synthesis. Cell **37**:879-887

Chodosh, L. A., Baldwin, A. S., Carthew, R. W. and Sharp, P. A. (1988) Human CCAAT-binding proteins have heterologous sub-units. Cell **55**:11-24

Choi, O.-R. and Engel, J.D. (1988) Developmental regulation of ß-globin gene switching. Cell **55**:17-26

Chou, R.H., Churchill, J.R., Flubache, M.M., Mapstone, D.E., Jones, J. (1990) Identification of a nuclear matrix-associated region of the c-*myc* protooncogene and its recognition by a nuclear protein in the humna leukemia HL-60 cell-line. Cancer Res. **50**:3199-3206

Chuang, R.Y. and Chuang, L.F. (1987) The 180 KDa polypeptide contains the DNA-binding domain of RNA polymerase II. Biochem. Biophys. Res. Commun. **145**:73-80

Cockerill, P.N. and Garrard, W.T. (1986) Chromosomal loop anchorage of the Kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. Cell **44**:273-282

Collis, P., Antoniou, M. and Grosveld, F. (1990) Definition of the minimal requirements within the human beta-globin gene and the dominant control region for high level expression. EMBO J. **9**:233-240

Comai, L., Tanese, N. and Tjian, R. (1992) The TATA-binding protein and associated

factors are integral components of the RNA polymerase I transcription factor, SL1. Cell **68**:965-976

Conaway, J.W., Hanley, J.P., Garrett, K.P. and Conaway, R.C. (1991)

Transcription initiated by RNA polymerase II and transcription factors from liver. Structure and action of transcription factors epsilon and tau. J. Biol. Chem. **266**:7804-7811

Conaway, J.W. and Conaway, R.C. (1991) Initiation of eukaryotic messenger RNA synthesis. J. Biol. Chem. **266**:17721-17724

Conaway, J.W. and Conaway, R.C. (1989) A multisubunit transcription factor essential for accurate initiation by RNA polymerase II.

J. Biol. Chem. **264**:2357-2362

Conaway, J.W., Reines, D. and Conaway, R.C. (1990) Transcription initiated by RNA polymerase II and purified transcription factors from liver. Cooperative action of transcription factors tau and epsilon in initial complex formation. J. Biol. Chem. **265**:7552-7558

Conaway, J.W., Bond, M.W. and Conaway, R.C. (1987) An RNA polymerase II transcription system from rat liver: purification of an essential component. J. Biol. Chem. **265**:8293-8297

Corden, J.L. (1990) Tails of RNA polymerase II. Trends Biochem. Sci. **15**:383-387

Corden, J., Waslyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) Promoter sequences of eukaryotic protein-coding genes. Science **209**:1406-1414

Cordingley, M.G., Riegel, A.T. and Hager, G.L. (1987) Steroid-dependent interaction of transcription factors with the inducible promoter of mouse mammary tumor virus *in vivo*. Cell **48**:261-270

Cortes, P., Flores, O. and Reinberg, D. (1992) Factors involved in specific transcription by mammalian RNA polymerase II: purification and analysis of transcription factor IIA and identification of transcription factor IIJ. Mol. Cell. Biol. **12**:413-421

Croston, G.E., Laybourn, P.J., Paranjape, S.M. and Kadonaga, J.T. (1992) Mechanism of transcriptional antirepression by GAL4-VP16. Genes and Dev. **6**:2270-2281

Croston, G.E., Kerrigan, L.A., Lira, L.M., Marshak, D.R. and Kadonaga, J.T. (1991) Sequence-specific antirepression of histone H1-mediated inhibition of basal RNA polymerase II transcription. Science **251**:643-649

Dardick, I. and Setterfield, G. (1976) Volume of condensed chromatin in developing

primitive-line erythrocytes of chick. Exp. Cell Res. **100**:159-171

Davidson, I., Xiao, J.H., Rosales, R., Staub, A. and Chambon, P. (1988) The HeLa cell protein TEF-1 binds specifically and cooperatively to two SV40 enhancer motifs of unrelated sequence. Cell **54**:931-942

Davie, J.R. and Nickel, B.E. (1987) The ubiquitinated histone species are enriched in histone H1-depleted chromatin regions. Biochim. Biophys. Acta **909**:183-189

Davie, J.R., Saunders, C.A., Walsh, J.M. and Weber, S.C. (1981) Histone modifications in the yeast S-cerevisiae. Nucleic Acids Res. **9**:3205-3216

Davie, J.R. and Saunders, C.A. (1981) Chemical composition of nucleosomes among domains of calf thymus chromatin differing in micrococcal nuclease accessibility and solubility properties. J. Biol. Chem. **256**:2574-2580

Davison, B.L., Egly, J., Mulvihill, E.R. and Chambon, P. (1983) Formation of stable preinitiation complexes between eukaryotic class B transcription factors and promoter sequences. Nature **301**:680-686

ˆdeBoer, E., Antoniou, M., Mignotte, V., Wall, L. and Grosveld, F. (1988) The human ß-globin promoter: nuclear protein factors and erythroid specific induction of transcription. EMBO J. **7**:4203-4212

Delcuve, G.P. and Davie, J.R. (1989) Chromatin structure of erythroid-specific genes of immature and mature chicken erythrocytes. Biochem. J. **263**:179-186

Demczuk, S., Donovan, M., Franklin, G. and Ohlsson, R. (1990) Order of probe and nuclear protein extract addition can determine specificity of protein-DNA complexes in tested mobility shift assays. Nucl. Acids Res. **19**:677-678

Dexter, T. and Spooncer, E. (1987) Growth and differentiation in the hemopoietic system A. Rev. Cell Biol. **3**:423-441

Dierks, P., van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J. and Weissmann, C. (1983) Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit beta-globin gene in mouse 3T6 cells. Cell **32**:695-706

Dietrich, J., Teissere, M., Job, C. and Job, D. (1986) Effect of salts on abortive and productive elongation catalysed by wheat germ RNA polymerase II. Nucleic Acids Res. **14**:1583-1597

Dignam, J.D., Martin, P.L., Shastry, B.S. and Roeder, R.G. (1983a) Eukaryotic gene transcription with purified components. Methods Enzymol **101**:582-598

Dignam, J.D., Lebovitz, R.M. and Roeder, R. (1983b) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucl. Acids

Res. **11**:1475-1489

Dingwall, C. and Laskey, R.A. (1991) Nuclear targeting sequences - a consensus? Trends Biochem. Sci **16**:478-481

Disela, C., Glineur, C., Bugge, T., Sap, J., Stengl, G., Dodgson, j., Stunnenberg, H., Beug, H. and Zenke, M. (1991) v-*erb*A overexpression is required to extinguish c-erbA function in erythroid cell differentiation and regulation of the *erb*A target gene CAII. Genes Dev. **5**:2033-2047

Doenecke, D. and Gallwitz, D. (1982) Acetylation of histones in nucleosomes. Mol. Cell Biochem. **44**:113-128

Dorn, A., Bollekens, J., Staub, A., Benoist, C. and Mathis, D. (1987a) A multiplicity of CCAAT box-binding proteins. Cell **50**:863-872

Dorn, A., Durand, B., Marfing, C., Le Meur, M., Benoist, C. and Mathis, D. (1987b) The conserved MHC class II boxes - x and y - are transcriptional control elements and specifically bind nuclear proteins. Proc. Natl. Acad. Sci. USA **84**:6249-6253

Drew, H.R. and McCall, M.J. (1987) Structural analysis of a reconstituted DNA containing three histone octamers and histone H5. J. Mol. Biol. **197**:485-511

Dusserre, Y. and Mermod, N. (1992) Purified cofactors and histone H1 mediate transcriptional regulation by CTF/NF1. Mol. Cell. Biol. **12**:5228-5237

Dworetzky, S.I., Wright, K.L., Fey, E.G., Penman, S., Lian, J.B., Stein, J.L. and Stein, G.S. (1992) Sequence-specific DNA-binding proteins are components of a nuclear matrix-attachment site. Proc. Natl. Acad. Sci. USA **89**:4178-4182

Dynan, W.S. (1989) Modularity in promoters and enhancers. Cell **58**:1-4

Dynan, W.S. and Tjian, R. (1983) Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. Cell **32**:669-680

Dynlacht, B.D., Hoey, T. and Tjian, R. (1991) Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. Cell **66**:563-576

Earnshaw, W.C., Halligan, B., Cooke, C.A., Heck, M.M. and Liu, L.F. (1985) Topoisomerase II is a structural component of mitotic chromosome scaffolds. J. Cell. Biol. **100**:1706-1715

Eaves, A.C. and Eaves, C.J. (1984) Erythropoiesis in culture. Clin. Haematol. **13**:371-391

Egly, J.M., Miyamoto, N.G., Moncolin, V. and Chambon, P. (1984)

Is actin a transcription initiation factor for RNA polymerase B?

EMBO J. **3**:2363-2371

Eisenmann, D.M., Dollard, C. and Winston, F. (1989) SPT15, the gene encoding the yeast TATA binding factor TFIID, is required for normal transcription *in vivo*. Cell **58**:1183-1191

Eisenman, R.N., Tachiban, C.Y., Abrams, H.D. and Hann, S.R. (1985) V-*myc*-encoded and c-*myc* encoded proteins are associated with the nuclear matrix. Mol. Cell. Biol. **5**:114-126

Elgin, S.C.R. (1988) The formation and function of DNase I hypersensitive sites in the process of gene activation. J. Biol. Chem. **263**:19259-19262

Emerson, B.M., Nickol, J.M. and Fong, T. (1989) Erythroid-specific activation and derepression of the chick beta-globin promoter *in vitro*. Cell **57**:1189-1200

Emerson, B.M., Nickol, J.M., Jackson, P.D. and Felsenfeld, G. (1987) Analysis of the tissue-specific enhancer at the 3' end of the chicken adult ß-globin gene. Proc. Natl. Acad. Sci. USA **84**:4786-4790

Emerson, B.M., Lewis, C.D. and Felsenfeld, G. (1985) Interaction of specific nuclear factors with the nuclease-hypersensitive region of the chicken adult beta-globin gene: Nature of the binding domain. Cell **41**:21-30

Emerson, B.M. and Felsenfeld, G. (1984) Specific factor conferring nuclease hypersensitivity at the 5' end of the chicken adult beta-globin gene. Proc. Nat. Acad. Sci. USA **81**:95-99

Enver, T., Raich, N., Ebens, A.J. Papayannopoulou, T., Constantini, F. and Stamatoyannopoulos, G. (1990) Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice.

Nature **344**:309-313

^Evans, T. and Felsenfeld, G. (1991) *Trans*-activation of a globin promoter in nonerythroid cell. Mol. Cell. Biol. **11**:843-853

Evans, T., Felsenfeld, G. and Reitman, M. (1990) Control of globin gene transcription. Annu. Rev. Cell. Biol. **6**:95-124

^Evans, T. and Felsenfeld, G. (1989) The Erythroid-Specific Transcription Factor Eryf1: A New Finger Protein. Cell **58**:877-885

^Evans, T., Reitman, M. and Felsenfeld, G. (1988) An erythrocyte-specific DNA binding factor recognizes a regulatory sequence common to all chicken globin genes. Proc. Natl. Acad. Sci. USA **85**:5976-5980

Faisst, S. and Meyer, S. (1992) Compilation of vertebrate-encoded transcription factors.

Nucl. Acids Res. **20**:3-26

Fascher, K.D., Schmitz, J. and Horz, W. (1990) Role of *trans*-activating proteins in the generation of active chromatin at the PHO5 promoter in S. cerevisiae. EMBO J. **9**:2523-2528

Felsenfeld, G. (1992) Chromatin as an essential part of the transcriptional mechanism. Nature **355**:219-224

Felsenfeld, G. and McGhee, J.D. (1986) Structure of the 30 nm chromatin fiber. Cell **44**:375-377

Finch, J.T. and Klug, A (1976) Solenoidal model for superstructure in chromatin. Proc. Natl. Acad. Sci. USA **73**:1897-1901

Fire, A., Samuels, M. and Sharp, P.A. (1984) Interactions between RNA polymerase II, factors, and template leading to accurate transcription. J.Biol. Chem. **259**:2509-2516

^Fischer, D-K. and Nowock, J. (1990) The T -> C substitution at - 198 ^Agamma-globin gene associated with the British form of HPFH generates overlapping recognition sites for two DNA-binding proteins. Nucl. Acids Res. **18**:5685-5693

Fisher. E.A. and Felsenfeld. G. (1986) Comparison of the folding of beta-globin and ovalbumin gene containing chromatin isolated from chicken oviduct and erythrocytes. Biochemistry **25**:8010-8016

Flanagan, P.M., Kelleher, R.J., Sayre, M.H., Tschochner, H. and Kornberg, R.D. (1991) A mediator required for activation of RNA polymerase II transcription *in vitro*. Nature **30**:436-438

Flores, O., Lu, H., Killeen, M., Greenblatt, J., Burton, Z.F. and Reinberg, D. (1991) The small subunit of transcription factor TFIIF recruits RNA polymerase II into the preinitiation complex. Proc. Natl. Acad. Sci. **88**:9999-10003

Flores, O., Maldonado, E., Burton, Z., Greenblatt, J. and Reinberg, D. (1988) Factors involved in specific transcription by mammalian RNA polymerase II. RNA polymerase II-associating protein 30 is an essential component of transcription factor IIF. J. Biol. Chem. **263**:10812-10816

* Fong, T.C. and Emerson, B.M. (1992) The erythroid-specific protein cGATA-1 mediates distal enhancer activity through a specialized ß-globin TATA box. Genes Dev. **6**:521-532

Forrester, W.C., Takegawa, T., 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. Nucl. Acids Res. **15**:10159-10177

Frampton, J., Walker, J., Plumb, M. and Harrison, P.R. (1990) Synergy between the NF-E1 erythroid-specific transcription factor and the CACCC factor in the erythroid-specific promoter of the human porphobilinogen deaminase gene. Mol. Cell. Biol. **10**:3838-3842

Franza, B.R., Rauscher, F.J., Josephs, S.F. and Curran, T. (1988) The *Fos* complex and *Fos*-related antigens recognize sequence elements that contain AP-1 binding sites. Science **239**:1150

Fraser, P., Hurst, J., Collis, P. and Grosveld, F. (1990) DNase I hypersensitive sites 1,2 and 3 of the human ß-globin dominant control region direct position-independent expression. Nucl. Acids Res. **18**:3503-3508

Fukamachi, H., Saito, T., Tojo, A., Kitamura, T., Urabe, A. and Takaku, F. (1987) Binding of erythropoietin to CFU-E derived from fetal mouse liver cells. Exp. Hematol. **15**:833-837

Gallarda, J.L., Foley, K.P., Yang, Z. and Engel, J.D. (1989) The beta-globin stage selector element factor is erythroid-specific promoter/enhancer binding protein NF-E4. Genes Dev. **3**:1845-1859

Garcia, J., Wu, F. and Gaynor, R. (1987) Upstream regulatory regions required to stabilize binding to the TATA sequence in an adenovirus early promoter. Nucleic Acids Res. **15**:8367-8385

Gaubatz, J.W. and Chalkley, R. (1977) Distribution of H-1 histone in chromatin digested by micrococcal nuclease. Nucl. Acids Res. **4**:3281-3301

Gasaryan, K. (1982) Genome activity and gene expression in avian erythroid cells. Int. Rev. Cytology **74**:95-126

Gasser, S.M., Amati, B.B. Cardenas, M.E. and Hofmann, J.F.-X. (1989) Studies on scaffold attachment sites and their relation to genome function. Int. Rev. Cytol **119**:57-96

Gasser, S.M. and Laemmli, U.S. (1987) A glimpse at chromosomal order - review. Trends Genetics **3**:16-22

Getzenberg, R.H., Pienta, K.J., Ward, W.S. and Coffey, D.S. (1991) Nuclear Structure and the three-dimensional organization of DNA. J. Cell. Biochemistry **47**:289-299

Giglioni, B., Comi, P., Ronchi, A., Mantovani, R. and Ottolenghi, S. (1989) The same nuclear proteins bind the proximal CACCC box of the human beta-globin promoter and a similar sequence in the enhancer. Biochem. Biophys. Res. Comm. **164**:149-55

Gomez-Cuadrado, A., Rousseau, S. Renaud, J. and Ruiz-Carrillo, A. (1992) Repression of the H5 histone gene by a factor from erythrocytes that binds to the region of

transcription initiation. EMBO J. **11**:1857-1866

^Gong, Q.-H and Dean, A. (1993) Enhancer-dependent transcription of the epsilon-globin promoter requires promoter-bound GATA-1 and enhancer-bound AP-1/NF-E2. Mol. and Cell Biol. **13**:911-917

^Gong, Q.-H., Stern, J. and Dean, A. (1991) Transcriptional role of a conserved GATA-1 site in the human ϵ-globin gene promoter. Mol. Cell. Biol. **11**:2558-2566

Goyal, N., Knox, J. and Gronostajski, R.M. (1990) Analysis of multiple forms of nuclear factor I in human and murine cell lines. Mol. Cell Biol. **10**:1041-1048

Grachev, M.A., Hatrmann, G.R., Maximova, T.G., Mustaev, A.A., Schaffner, A.R., Sieber, H. and Zaychikov, E.F. (1986) Highly selective affinity labelling of RNA polymerase B (II) from wheat germ.

FEBS Lett. **200**:287-290

Gralla, J.D. (1991) Transcriptional control - Lessons from an *E. coli* promoter data base. Cell **66**:415-418

Gralla, J.D. (1990) Promoter recognition and mRNA initiation by *Escherichia coli* σ 70 Methods Enzymol. **185**:37-53

Grasso, J.A. and Woodward, J.W. (1966) Relationship between RNA synthesis and hemoglobin synthesis in amphibian erythropoiesis - cytochemical evidence. J. Cell Biol. **31**:279-294

Greenberg, M.E. and Ziff, E.B. (1984) Stimulation of 3T3 cells induces transcription of the c-*fos* proto oncogene. Nature **311**:433-438

Greenblatt, J. (1991a) RNA polymerase-associated transcription factors. Trends Biochem. Sci. **16**:408-411

Greenblatt, J. (1991b) Roles of TFIID in transcriptional initiation by RNA polymerase II. Cell **66**:1067-1070

Greenleaf, A.L., Haars, R. and Bautz, E.K.F. (1976) *In vitro* proteolysis of a large subunit of drosophila-melanogaster RNA polymerase-B. FEBS Lett. **71**:205-208

Gross, D.S. and Garrard, W.T. (1988) Nuclease hypersensitive sites in chromatin. Ann. Rev. Biochem. **57**:159-197

Gross, D.S. and Garrard, W.T. (1987) Poising chromatin for transcription. Trends Biochem. Sci. **12**:293-297

Grosschedl, R. and Birnstiel, M.L. (1980a) Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion

mutants *in vivo*. P. Nas. Biol. **77**:1432-1436

Grosschedl, R. and Birnstiel, M.L. (1980b) Spacer DNA sequences upstream of the T-A-T-A-A-A-T-A sequence are essential for promotion of H2A histone gene-transcription *in vivo*. P. Nas. Biol. **77**:7102-7106

Grosveld, F., van Assendelft, G.B., Greaves, D.R. and Kollias, B. (1987) Position-independent, high-level expression of the human beta-globin gene in transgenic mice. Cell **51**:975-985

Groudine, M. and Weintraub, H. (1981) Activation of globin genes during chicken development. Cell **24**:393-401

Gumucio, D.L., Rood, K.L., Blanchard-McQuate, K. L., Gray, T.A., Saulino, A. and Collins, F.S. (1991) Interaction of Sp1 with the human gamma globin promoter: binding and transactivation of normal and mutant promoters. Blood **78**:1853-1863

^Gumucio, D.L., Lockwood, W.K., Weber, J.L., Saulino, A.M., Delgrosso, K., Surrey, S., Schwartz, E., Goodman, M. and Collins, F.S. (1990) The - 175 -> C mutation increases promoter strength in erythroid cells: Correlation with evolutionary conservation of binding sites for two *trans*-acting factors. Blood **75**:756-761

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

Gundelfinger, E.D. (1983) Interaction of nucleic acids with the DNA-dependent RNA polymerases of *Drosophila*. FEBS Lett. **157**:133-138

Ha, I., Lane, W. and Reinberg, D. (1991) Cloning of a human gene encoding the general transcription initiation factor TFIIB. Nature **352**:689-695

Hahn, S., Buratowski, S., Sharp, P.A. and Guarente, L. (1989) Cell Isolation of the gene encoding the yeast TATA binding protein TFIID: a gene identical to the SPT15 suppressor of Ty element insertions. Cell. **58**:1173-1181

Hai, T., Horikoshi, M., Roeder, R.G. and Green, M.R. (1988) Analysis of the role of the transcription factor ATF in the assembly of a functional preinitiation complex. Cell **54**:1043-1051

Ham, J., Steger, G. and Yaniv, M. (1992) How do eukaryotic activator proteins stimulate the rate of transcription by RNA polymerase II?

FEBS Lett. **307**:81-86

Hamburger, V. and Hamilton, H.L. (1951) A series of normal stages in the development of the chick embryo. J. Morph. **88**:49-92

Hannon, R., Evans, T., Felsenfeld, G. and Gould, H. (1991) Structure and promoter activity of the gene for the erythroid transcription factor GATA-1. Proc. Natl. Acad. Sci. USA **88**:3004-3008

Hanson, J.C. and Ausio, J. (1992) Chromatin dynamics and the modulation of genetic activity. TIBS **17**:187-191

Harrison, P.R., Plumb, M., Frampton, J., Chambers, I., Llewellyn, D., Chester, J., Macleod, K., Fleming, J., O'prey, J., Walker, M., and Wainwright, H. (1988) *Cis* and *trans* control of erythroid cell-specific gene expression during erythropoiesis. J. Cell Sci. Suppl. **10**:145-155

Harrison, P.R. (1982a) Stem cell regulation in erythropoiesis. Nature **295**:454-455

Harrison, P.R. (1982b) Regulation of differentiation in retrovirus-induced erythroleukaemias. Cancer Surveys **1**:232-277

Hartmann, P.G., Chapman, G.E., Moss, T. and Bradbury, E.M. (1977) Studies on role and mode of operation of very lysine-rich histone H-1 in eukaryote chromatin: 3 structural regions of histone H-2 molecule. Eur. J. Biochem. **77**:45-51

Hawley, D.K. and Roeder, R.G. (1985) Separation and partial characterization of three functional steps in transcription initiation by human RNA polymerase II. J.Biol. Chem. **260**:8163-8172

Hayes, J.J. and Wolffe, A.P. (1992) Histones H2A/H2B inhibit the interaction of transcription factor IIIA with the Xenopus borealis somatic 5S RNA gene in a nucleosome. Proc. Natl. Sci. USA **89**:1229-1233

Hebbes, T.R., Thorne, A.W., Clayton, A. L. and Crane-Robinson, C. (1992) Histone acetylation and globin gene switching. Nucleic Acids Res. **20**:1017-1022

Heberlein, C., Fischer, K.-D., Stoffel, M., Nowock, J., Ford, A., Tessmer, U. and Stocking, C. (1992) The gene for erythropoietin receptor is expressed in multipotential hematopoietic and embryonal stem cells: evidence for differentiation stage-specific regulation. Mol. Cell. Biol. **12**:1815-1826

Hendzel, M.J. and Davie, J.R. (1992) Nuclear distribution of histone deacetylase: a marker enzyme for the internal nuclear matrix. Biochim. Biophys. Acta **1130**:307-311

Hendzel, M.J., Delcuve, G.P. and Davie, J.R. (1991) Histone deacetylase is a component of the internal nuclear matrix. J. Biol. Chem. **266**:21936-21942

Hill, C.S. and Thomas, J.O. (1990) Core histone-DNA interactions in sea urchin sperm interacts with linker DNA. Eur. J. Biochem. **187**:145-153

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. and Pease, L.R. (1989) Site-directed

mutagenesis by overlap extension using the polymerase chain reaction. Gene **77**:51-59

Hoey, T., Weinzierl, R.O., Gill, G., Chen, J.L., Dynlacht, B.D. and Tjian, R. (1993) Molecular cloning and functional analysis of *Drosophila* TAF110 reveal properties expected of coactivators. Cell **72**:247-260

Horikoshi, M., Yamamoto, T., Ohkuma, T., Weil, P.A. and Roeder, R.G. (1990) Analysis of structure-function relationships of yeast TATA box binding factor TFIID. Cell **61**:1171-1178

Horikoshi, M., Wang, C.K., Fujii, H., Cromlish, J.A., Weil, P.A. and Roeder, R.G. (1989a) Cloning and structure of a yeast gene encoding a general transcription initiation factor TFIID that binds to the TATA box. Nature **341**:299-303

Horikoshi, M., Wang, C.K., Fujii, H., Cromlish, J.A., Weil, P.A. and Roeder, R.G. (1989b) Purification of a yeast TATA box-binding protein that exhibits human transcription factor IID activity. Proc. Natl. Acad. Sci. USA **86**:4843-4847

Horikoshi, M., Carey, M.F., Kakidani, H. and Roeder, R.G. (1988a) Mechanism of action of a yeast activator: direct effect of GAL4 derivatives on mammalian TFIID-promoter interactions. Cell **54**:665-669

Horikoshi, M., Hai, T., Lin, Y.-S., Green, M.R. and Roeder, R.G. (1988b) Transcription factor ATF interacts with the TATA factor to facilitate establishment of a preinitiation complex. Cell **54**:1033-1042

Horikoshi, M., Sekimizu, K. and Natori, S. (1985) Conformational change of DNA binding subunit of RNA polymerase II on binding to DNA.

Biochem. Biophys. Res. Commun. **129**:141-147

Horikoshi, M., Tamura, H., Sekimizu, K., Obinata, M. and Natori, S. (1983) Identification of the DNA binding subunit of RNA polymerase II from Ehrlich ascites tumor cells. J. Biochem. Tokyo. **94**:1761-1767

Imai, B.S., Yau, P., Baldwin, J.P., Ibel, K., May, R.P. and Bradbury, E.M. (1986) Hyperacetylation of core histones does not cause unfolding of nucleosomes. Neutron scatter data accords with disc shape of the nucleosome. J. Biol. Chem. **261**:8784-8792

Imataka, H., Sogawa, K., Yasumoto, K.-I., Kikuchi, Y., Sasano, K., Kobayashi, A., Hayami, M. and Fujii-Kuriyama, Y. (1992) Two regulatory proteins that bind to the basic transcription element (BTE), a GC box sequence in the promoter region of the rat P-4501A1 gene. EMBO J **11**:3663-3671

Ingles, C.J., Shales, M., Cress, W.D., Triezenberg, S. and Greenblatt, J. (1991) Reduced binding of TFIID to transcriptionally compromised mutants of VP16. Nature

351:588-590

Ingram, V.M. (1981) Hemoglobin switching in amphibians and birds. In Hemoglobins in development and differentiation. (ed. G. Stamatoyannopoulos and A.W. Nienhuis), pp. 147-160. New York: Alan R. Liss, Inc.

Ingram, V.M. (1972) Embryonic red blood cell formation. Nature **234**:338-339

Inostroza, J.A., Mermelstein, F.H., Ha, I., Lane, W.S. and Reinberg, D. (1992) Drl, a TATA-binding protein-associated phosphoprotein and inhibitor of class II gene transcription. Cell **70**:477-489

Jackson, D.A., Dickinson, P. and Cook, P.R. (1990) The size of chromatin loops in HeLa cells. EMBO J. **9**:567-571

Jackson, D.A, and Cook, P.R. (1985) Transcription occurs at a nucleoskeleton. EMBO J. **4**:919-925

^Jackson, P.D., Evans, T., Nickol, J.M. and Felsenfeld, G. (1989) Developmental modulation of protein binding to beta-globin gene regulatory sites within chicken erythrocyte nuclei. Genes Dev. **3**:1860-1873

Jackson, S.P. and Tjian, R. (1988) O-glycosylation of eukaryotic transcription factors: Implications for mechanisms of transcriptional regulation. Cell **55**:125-133

Jane, S.M., Ney, P.A., Vanin, E.F., Gumucio, D.L. and Neinhuis, 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 ß-promoter. EMBO J. **11**:2961-2969

^Jarman, A.P., Wood, W.G., Sharpe, J.A., Gourdon, G., Ayyub, H. and Higgs, D.R. (1991) Characterization of the major regulatory element upstream of the human α-globin cluster. Mol. Cell Biol. **11**:4679-4689

Jarman, A.P., Higgs, D.R. (1988) Nuclear scaffold attachment sites in the human globin gene complexes. EMBO J. **7**:3337-3344

Johnson, F.B. and Krasnow, M.A. (1992) Differential regulation of transcription preinitiation complex assembly by activator and repressor homeodomain proteins. Genes and Dev. **6**:2177-2189

Johnson, L.M., Kayne, P.S., Kahn, E.S. and Grunstein, M. (1990)

Genetic evidence for an interaction between SIR3 and histone H4 in the repression of the silent mating loci in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **87**:6286-6290

Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelley, T.J. and Tjian, R. (1987) A cellular

DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell **48**:79-89

Jones, K.A., Yamamoto, K.R. and Tjian, R. (1985) Two distinct transcription factors bind to the HSV thymidine kinase promoter *in vitro*. Cell **42**:559-572

Jones, K.A. and Tjian, R. (1985) Sp1 binds to promoter sequences and activates herpes simplex virus 'immediate-early' gene transcription *in vitro*. Nature **317**:179-182

Kadonaga, J.T. (1990) Assembly and disassembly of the *Drosophila* RNA polymerase II complex during transcription. J. Biol. Chem. **265**:2624-2631

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

Kadonaga, J.T., Jones, K.A. and Tjian, R. (1986) Promoter-specific activation of RNA polymerase II transcription by Sp1. Trends Biochem. **11**:20-23

Kalandadze, A.G., Bushara, S.A., Vassetzky, T.S. and Razin, S.V. (1990) Characterization of DNA pattern in the site of permanent attachment to the nuclear matrix located in the vicinity of a replication origin. Biochem. Biophys. Res. Commun. **168**:9-15

Kamakaka, R.T. and Thomas, J.O. (1990) Chromatin structure of transcriptionally competent and repressed genes. EMBO J. **9**:3997-4006

Kassavetis, G.A., Joazeiro, C.A., Pisano, M., Geiduschek, E.P., Colbert, T., Hahn, S. and Blanco, J.A. (1992) The role of the TATA-binding protein in the assembly and function of the multisubunit yeast RNA polymerase III transcription factor, TFIIIB. Cell **71**:1055-1064

Katagiri, F., Yamazaki, K., Horikoshi, M., Roeder, R.G. and Chua, N.-H. (1990) A plant DNA-binding protein increases the number of active preinitiation complexes in a human *in vitro* transcription system. Genes and Dev. **4**:1899-1909

Kawamura, H., Nagata, K., Masamune, Y. and Nakanishi, Y. (1993) Phosphorylation of NF-1 *in vitro* by cdc2 kinase. Biochem. Biophys. Res. Comm. **192**;1424-1431

Kelleher, R.J., Flanagan, P.M. and Kornberg, R.D. (1990) A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. Cell **61**:1209-1215

Kemper, B., Jackson, P.D. and Felsenfeld, G. (1987) Protein-binding sites within the 5' DNase I-hypersensitive region of the chicken alpha^{D}-globin gene. Mol. Cell. Biol. **7**:2059-2069

Kim, C.G., Swendeman, S.L., Barnhart, K.M. and Sheffery, M. (1990) Promoter elements and erythroid cell nuclear factors that regulate α-globin gene transcription

in vitro. Mol. Cell. Biol. **10**:5958-5966

Kim, W.-Y. and Dahmus, M.E. (1986) Immunochemical analysis of mammalian RNA polymerase II subspecies. Stability and relative *in vivo* concentration. J. Biol. Chem. **261**:14219-14225

Kimura, T., Mills, F.C., Allen, J. and Gould, H. (1983) Selective unfolding of erythroid chromatin in the region of the active beta-globin gene. Nature **306**:709-712

Klehr, D., Maass, K., Bode, J. (1991) Scaffold-attached regions from the human interferon beta domain can be used to enhance the stable expression of genes under the control of various promoters. Biochemistry **30**:1264-1270

Knezetic, J.A., and Felsenfeld, G. (1989) Identification and characterization of a chicken α-globin enhancer. Mol. Cell. Biol. **9**:893-901

Knezetic, J.A., Jacob, G.A. and Luse, D.S. (1988) Assembly of RNA polymerase II preinitiation complexes before assembly of nucleosomes allows efficient initiation of transcription on nucleosomal templates. Mol. Cell Biol. **8**:3114-3121

Ko, L.J. and Engel, J.D. (1993) DNA-binding specificities of the GATA transcription factor family. Moll. Cell. Biol. **13**:4011-4022

Kollmar, R., Lindstrom, M.J. and Farnham, P.J. (1992) Heat sensitivity and Sp1 activation of complexes formation at the Syrian hamster carbamoyl-phosphate synthase (Glutamine-hydrolysing)/aspartate carbamoyltransferase/dihydroorotase promoter *in vitro*. J. Biol. Chem. **267**:385-391

Kornberg, R.D. (1977) Structure of chromatin. Ann. Rev. Biochem. **46**:931-954

Kozmik, Z., Urbanek, P. and Paces, V. (1990) Albumin improves formation and detection of some specific protein-DNA complexes in the mobility shift assay. Nucl. Acids Res. **18**:2198

Krieg, P.A., Robins, A.J., D'Andrea,R. and Wells, J.R.E. (1983) The chicken H5 gene is unlinked to core and H1 histone genes. Nucl. Acids Res. **11**:619-627

Kruse, U. Qian, F. and Sippel, A.E. (1991) Identification of a fourth nuclear factor I gene in chicken by cDNA cloning: NFI-x. Nucl. Acids Res. **19**:6641

Kuhl, D., de la Fuente, J., Chaturvedi, M., Parimoo, S., Ryals, J., Meyer, F. and Weissmann, C. (1987) Reversible silencing of enhancers by sequences derived from the human IFN-alpha promoter. Cell **50**:1057-1069

Kuramochi, S., Ikawa, Y. and Todokoro, K. (1990) Characterization of murine erythropoietin receptor genes. J. Mol. Biol. **216**:567-575

^ Lacronique, V., Boquet, D., Lopez, S., Kahn, A. and Raymondjean, M. (1992) *In vitro*

and *in vivo* protein-DNA interactions on the rat erythroid-specific L' pyruvate kinase gene promoter. Nucl. Acids Res. **20**:5669-5676

Landes, G.M., Villeponteau, B., Pribyl, T.M. and Martinson, H.G. (1982) Haemoglobin switching in chickens: Is the switch initiated post-transcriptionally. J. Biol. Chem. **257**:11008-11014

Laybourn, P.J. and Kadonaga, J.T. (1992) Threshold phenomena and long-distance activation of transcription by RNA polymerase II. Science **257**:1682-1684

Laybourn, P.J. and Kadonaga, J.T. (1991) Role of nucleosomal cores and histone H1 in regulation of transcription by RNA polymerase II. Science **254**:238-245

Laybourn, P.J. and Dahmus, M.E. (1990) Phosphorylation of RNA polymerase IIA occurs subsequent to interaction with the promoter and before the initiation of transcription. J. Biol. Chem. **265**:13165-13173

Laybourn, P.J. and Dahmus, M.E. (1989) Transcription-dependent structural changes in the C-terminal domain of mammalian RNA polymerase subunit IIa/o. J. Biol. Chem. **264**:6693-6698

Lazarides, E. (1987) From genes to structural morphogenesis: The genesis and epigenesis of a red blood cell. Cell **51**:345-356

Lee, D.Y., Hayes, J.J., Pruss, D. and Wolffe, A.P. (1993) A positive role for histone acetylation in transcription factor access to nucleosomal DNA. Cell **72**:73-84

Lee, D.K., Dejong, J., Hashimoto, S., Horikoshi, M. and Roeder, R.G. (1992) TFIIA induces conformational changes in TFIID via interactions with the basic repeat. Mol. Cell. Biol. **12**:5189-5196

^ Lemarchandel, V., Ghysdael, J., Mignotte, V., Rahuel, C. and Romeo, P.-H. (1993) GATA and Ets *cis*-acting sequences mediate megakaryocyte-specific expression. Mol. Cell. Biol. **13**:668-676

Lennard, A.C. and Thomas, J.O. (1985) The arrangement of H5 molecules in extended and condensed chicken erythrocyte chromatin. EMBO J. **4**:3455-3462

Letovsky, J. and Dynan, W.S. (1989) Measurement of the binding of transcription factor Sp1 to a single GC box recognition sequence. Nucl. Acids Res. **17**:2639-2653

Levywilson, B. and Fortier, C. (1989) The limits of the DNase I-sensitive domain of the human apolipoprotein B gene coincide with the locations of chromosomal anchorage loops and define the 5' and 3' boundaries of the gene. J. Biol. Chem. **264**:21196-21204

Lewis, C.D., Clark, S.P., Felsenfeld, G. and Gould, H. (1988) An erythrocyte-specific proteins that binds to the poly (dG) region of the chicken beta-globin gene

promoter. Genes Dev. **2**:863-873

Lewis, M.K. and Burgess, R.R. (1982) Eukaryotic RNA polymerases in The Enzymes, 3rd ed., vol. 15B, P.D. Boyer, Ed. (New York: Academic Press), pp 109-153

Li, X.-Y., Mantovani, R., van Juijsduijnen, R.H., Andre, I., Benoist, C. and Mathis, D. (1992) Evolutionary variation of the CCAAT-binding transcription factor NF-Y Nucleic Acids Res. **20**:1087-1091

Li, R., Knight, J.D., Jackson, S.P., Tjian, R. and Botchan, M.R. (1991) Direct interaction between Sp1 and the BPV enhancer E2 protein mediates synergistic activation of transcription. Cell **65**:493-505

Libertini, L.J., Ausio, J., van Holde, K.E. and Small, E.W. (1988)

Histone hyperacetylation: its effects on nucleosome core particle transitions. Biophys. J. **53**:477-487

Lin, H.J., Anagnou, N.P., Rutherford, T.R., Shimada, T. and Nienhuis, A.W. (1987) Activation of the human beta-globin promoter in K562 cells by DNA sequences 5' to the fetal gamma- or embryonic zeta-globin genes. J. Clin. Invest. **80**:374-380

Lin. Y.-S. and Green, M.R. (1991) Mechanism of action of an acidic transcriptional activator *in vitro*. Cell **64**:971-981

Lin, Y.-S., Ha, I., Maldonado, E., Reinberg, D. and Green, M.R. (1991) Binding of general transcription factor TFIIB to an acidic activating region. Nature **353**:569-571

Lin, Y.-S., Carey, M.F., Ptashne, M. and Green, M.R. (1988) GAL4 derivatives function alone and synergistically with mammalian activators *in vitro*. Cell **54**:659-664

Lobanenkov, V.V., Nicolas, R.H., Adler, V.V., Paterson, H., Klenova, E.M., Polotskaja, A.V. and Goodwin, G.H. (1990) A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-*myc* gene. Oncogene **5**:1743-1753

Lobo, S.M., Tanaka, M., Sullivan, M.L. and Hernandez, N. (1992) A TBP complex essential for transcription from TATA-less but not TATA-containing RNA polymerase III promoters is part of the TFIIIB fraction. Cell **71**:1029-1040

Lohr, D. and Hereford, L. (1979) L-yeast chromatin is uniformly digested by DNAse-1. Proc. Natl. Acad. Sci. USA **76**:4285-4288

Loidl, P. (1988) Towards an understanding of the biological function of histone acetylation. FEBS Lett. **227**:91-95

Loidl, P. and Grobner, P. (1987) Postsynthetic acetylation of histones during the cell cycle: a general function for the displacement of histones during chromatin

rearrangements. Nucleic Acids Res. **15**:8351-8366

Lopez-Rodas, G., Brosch, G., Georgieva, E.I., Sendra, R., Franco, L. and Loidl, P. (1993) Histone deacetylase: A key enzyme for the binding of regulatory proteins to chromatin. FEBS **317**:175-180

Lorch, Y., LaPointe, J.W. and Kornberg, R.D. (1992) Initiation on chromatin templates in a yeast RNA polymerase II transcription system. Genes Dev. **6**:2282-2287

^Lowrey, C.H., Bodine, D.M. and Nienhuis, A.W. (1992) Mechanism of DNase I hypersensitive site formation within the human globin locus control region. Proc. Natl. Acad. Sci. USA **89**:1143-1147

Lu, H., Flores, O., Weinmann, R. and Reinberg, D. (1991) The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. Proc. Natl. Acad. Sci. USA **88**:10004-10008

Ma, L., Weeda, G., Jochemsen, A.G., Bootsma, D., Hoeijmakers, J.H.J. and van der Eb, A.J. (1991) Molecular and functional analysis of the XPBC/ERCC-3 promoter: transcription activity is dependent on the integrity of an SP1-binding site. Nucleic Acids Res. **20**:217-224.

Maldonado, E., Ha, I., Cortes, P., Weis, L. and Reinberg, D. (1990) Factors involved in specific transcription by mammalian RNA polymerase II: Role of transcription factors IIA, IID, and IIB during the formation of a transcription-competent complex. Mol. Cell. Biol. **10**:6335-6347

Margottin, F., Dujardin, G., Gerard, M., Egly, J.-M., Uuet, J. and Sentenac, A. (1991) Participation of the TATA factor in transcription of yeast U6 gene by RNA polymerase C. Science **251**:424-426

Martin, D.I.K. and Orkin, S.H. (1990) Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-1/Eryf1. Genes Dev. **4**:1886-1898

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

Martin, K.J. (1991) The interactions of transcription factors and their adaptors, coactivators and accessory proteins. Bio Essays **13**:499-503

Mastrangelo, I.A., Courey, A.J., Wall, J.S., Jackson, S.P. and Hough, P.V.C. (1991) DNA looping and Sp1 multimer links: a mechanism for transcriptional synergism and enhancement. Proc. Natl. Acad. Sci. USA **88**:5670-5674

Matsui, T., Segall, J., Weil, P.A. and Roeder, R.G. (1980) Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J. Biol. Chem. **255**:11992-11996

Matthews, H.R. (1988) in Chromosomes and Chromatin, Vo: I (Adolph, K.W.,ed.)pp 3-32, CRC Press, Boca Raton, FL

Matthews, H.R. and Waterborg, J.H. (1985) in The Enzymology of Post-translational Modification of Proteins (Freedman, R.B. and Hawkins, H.C., eds.)pp 125-185 Academic Press, London.

^ Max-Audit, I., Eleouet, J.-F. and Romeo, P.-H (1993) Transcriptional regulation of the pyruvate kinase erythroid-specific promoter. J. Biol. Chem. **268**:5431-5437

Maxam, A. and Gilbert, W. (1980) Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499-560

McDonagh, K.T., Lin, H.J., Lowrey, C.H., Bodine, D.M. and Nienhuis, A.W. (1991) The upstream region of the human gamma-globin gene promoter. Identification and functional analysis of nuclear protein binding sites. J. Biol. Chem. **266**:11965-11974

McGhee, J.D., Nickol, J.M, Felsenfeld, G. and Rau, D.C. (1983)

Higher order structure of chromatin: orientation of nucleosomes within the 30 nm chromatin solenoid is independent of species and spacer length. Cell **33**:831-841

McGhee, J.D. and Felsenfeld, G. (1980) Nucleosome structure. Annu. Rev. Biochem. **49**:1115-1156

McKnight, S. and Tjian, R. (1986) Transcriptional selectivity of viral genes in mammalian cells. Cell **46**:795-805

Meiklejohn, A.L., and Gralla, J.D. (1989) Activation of the lac promoter and its variants. J. Mol. Biol. **207**:661-673

Meisterernst, M., Roy, A.L., Lieu, H.M. and Roeder, R.G. (1991) Activation of class II gene transcription by regulatory factors is potentiated by a novel activity. Cell **66**:981-993

Meisterernst, M. and Roeder, R.G. (1991) Family of proteins that interact with TFIID and regulate promoter activity. Cell **67**:557-567

Meisterernst, M., Horikoshi, M. and Roeder, R.G. (1990) Recombinant yeast TFIID, a general transcription factor, mediates activation by the gene-specific factor USF in a chromatin assembly assay. Proc. Natl. Acad. Sci. USA **87**:9153-9157

Mermod, N., O'Neill, E.D., Kelly, T.J. and Tjian, R. (1989) The proline-rich transcriptional activator of CTF/NF-1 is distinct from the replication and DNA binding domain. Cell **58**:741-753

Migliaccio, A.R., Migliaccio, G., D'Andrea, A., Baiocchi, M., Crotta, S., Nicolis, S.,

Ottolenghi, S. and Adamson, J.W. (1991) Response to erythropoietin in erythroid subclones of the factor-dependent cell line 32D is determined by translocation of the erythropoietin receptor to the cell surface. Proc. Natl. Acad. Sci. USA **88**:11086-11090

^Mignotte,V., Wall, L., deBoer,E., Grosveld,F. and Romeo, P.-H. (1989a) Two tissue-specific factors bind the erythroid promoter of the human porphobilinogen deaminase gene. Nucl. Acids Res **17**:37-54

^Mignotte, V., Elequet, J.F., Raich, N. and Romeo, P.-H. (1989b) *cis*- and *trans*-acting elements involved in the regulation of the erythroid promoter of the human porphobilinogen deaminase gene. Proc. Natl. Acad. Sci. USA **86**:6548-6552

Miller, A.M. and Nasmyth, K.A. (1984) Role of DNA replication in the repression of silent mating type loci in yeast. Nature **312**:247-251

^Minie, M.E., Kimura, T. and Felsenfeld, G. (1992a) The developmental switch in embryonic rho-globin expression is correlated with erythroid-lineage differences in transcription factor levels. Development **115**:1149-1164

Minie, M.E., Clark, D., Trainor, C., Evans, T., Reitman, M., Hannon, R., Gould, H. and Felsenfeld, G. (1992b) Developmental regulation of globin gene expression. J. Cell Sci. **16**:15-20

Mirkovitch, J., Mirault, M.E. and Laemmli, U.K. (1984) Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. Cell **39**:223-232

Mitchell, P.M. and Tjian, R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins.

Science **245**:371-378

Moss, B.A., Joyce, W.G. and Ingram, V.M. (1973) Histones in chick embryonic erythropoiesis. J. Biol. Chem. **248**:1025-1031

Mura, C.V. and Stollar, B.D. (1984) Interaction of H1 and H5 histones with polynucleotides of ß- and Z- conformation. Biochemistry **23**:6141-6152

Mura, C.V., Setterfield, G. and Neelin, J.M. (1982) Immunofluorescent localization of lysine-rich histones in isolated nuclei from adult and embryonic chicken erythrocytes. Can. J. Biochem. **60**:215-223

Myers, R., Tilly, K. and Maniatis, T. (1986) Fine structure genetic analysis of a beta-globin promoter. Science **232**:613-618

Nakajima, N., Horikoshi, M. and Roeder, R.G. (1988) Factors involved in specific transcription by mammalian RNA polymerase II: Purification, genetic specificity, and

TATA box-promoter interactions of TFIID. Mol. Cell. Biol. **8**:4028-4040

Neelin, J. M. (1968) Common histone of avian erythropoiesis. Can. J. Biochem. **46**:241-247

Neelin, J.M., Callahan, P.X., Lamb, D.C. and Murray, K. (1964) The histones of chicken erythrocyte nuclei. Can. J. Biochem. **42**:1743-1752

Nelson, P.P., Albright, S.C., Wiseman, J.M. and Garrard, W.T. (1979) Reassociation of histone-H1 with nucleosomes. J. Biol. Chem. **254**: 1751-1760

^Ney, P., Sorrentino B., McDonagh, K. and Nienhuis, A. (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

Nickol, J.M. and Felsenfeld, G. (1988) Bidirectional control of the chicken beta- and epsilon-globin genes by a shared enhancer. Proc. Natl. Acad. Sci. USA **85**:2548-2552

^Nicolis, S., Bertini, C., Ronchi, A., Crotta, S., Lanfranco, L., Moroni, E., Giglioni, B. and Ottolenghi, S. (1991) An erythroid specific enhancer upstream to the gene encoding the cell-type specific transcription factor GATA-1. Nucl. Acids Res. **19**:5285-5291

Noguchi, T., Fukumoto, H., Mishina, T. and Obinata, M. (1987) Factors controlling induction of commitment of murine erythroleukemia (TSA8) cells to CFU-E (colony forming unit-erythroid). Development **101**:169-174

Noll, M. and Kornberg, R.D. (1977) Action of micrococcal nuclease on chromatin and location of histone H1. J. Mol. Biol. **109**:393-404

Nonet, M., Scafe, C., Sexton, J. and Young, R.A. (1987) Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. Mol. Cell. Biol. **7**:1602-1611

Ohkuma, Y., Sumimoto, H., Hoffmann, A., Shimasaki, S., Horikoshi, M. and Roeder, R.G. (1991) Structural motifs and potential T homologies in the large subunit of human general transcription factor TFIIE. Nature **354**:398-401

Olins, A.L., Carlson, R.D., Wright, E.B. and Olins, D.E. (1976) Chromatin *v* bodies: isolation, subfractionation and physical characterization. Nucl. Acids Res. **3**:3271-3291

Omichinski, J.G., Trainor, C., Evans, T., Gronenborn, A.M., Clore, G.M. and Felsenfeld, G. (1993a) 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

Omichinski, J.G., Clore, G.M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S.J. and Gronenborn, A.M. (1993b) NMR structure of a Specific DNA complex of Zn-containing DNA binding domain of GATA-1. Science **261**:438-448

Orkin, S.H. (1992) GATA-binding transcription factors in hematopoietic cells. Blood **80**:575-581

Orkin, S.H. (1990) Globin gene regulation and switching: Circa 1990. Cell **63**:665-672

Papayannopoulou, T., Brice, M. and Stamatoyannopoulos, G. (1986) Analysis of human hemoglobin switching in MEL X human fetal erythroid cell hybrids. Cell **46**:469-476

Pardoll, D.M., Vogelstein, B. and Coffey, D.S. (1980) A fixed site of DNA replication in eukaryotic cells. Cell **19**:527-536

Pascal, E. and Tjian, R. (1991) Different activation domains of Sp1 govern formation of multimers and mediate transcriptional synergism. Genes Dev. **5**:1646-1656

Paulson, J.R. and Laemmli, U.K. (1977) The structure of histone-depleted metaphase chromosomes. Cell **12**:817-828

! Perkins, N.D., Nicolas, P.H., Plumb, M.A. and Goodwin, G.H. (1989) The purification of an erythroid protein which binds to enhancer and promoter elements of haemoglobin genes. Nucl. Acids Res. **17**:1299-1314

Perlmann, T. and Wrange, O. (1988) Specific glucocorticoid receptor binding to DNA reconstituted in a nucleosome. EMBO J. **7**:3073-3079

Peters, L.L., Andrews, N.C., Eicher, E.M., Davidson, M.B., Orkin, S.H. and Lux, S.E. (1993) Mouse microcytic anaemia caused by a defect in the gene encoding the globin enhancer-binding protein NF-E2 Nature **362**:768-770

Peterson, M.G. and Tjian, R. (1992) Transcription. The tell-tail trigger [news comment] Nature **358**:620-621

Peterson, M.G., Inostroza, J., Maxon, M.E., Flores, O., Admon, A., Reinberg, D. and Tjian, R. (1991) Structure and functional properties of human general transcriptional factor TFIIE. Nature **354**:369-373

Peterson, M.G., Tanese, N., Pugh, B.F. and Tjian, R. (1990) Functional domains and upstream activation properties of cloned human TATA binding protein. Science **248**:1625-1630

Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.-F, D'Agati, V., Orkin, S.H. and Constantini, F. (1991) Erythroid differentiation in chimeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature **349**:257-260

Pham, T.A., Hwung, Y.-P., McDonnell, D.P. and O'Malley, B.W. (1991)

Transactivation functions facilitate the disruption of chromatin structure by estrogen receptor derivatives *in vivo*. J. Bio. Chem. **266**:18179-18187

Phi-Van, L., Von Kries, J.P., Ostertag, W. and Stratling, W.H. (1990) The chicken lysozyme 5' matrix attachment region increases transcription from a heterologous promoter in heterologous cells and dampens position effects on the expression of transfected genes. Mol. Cell Biol. **10**:2302-2307

Phi-Van, L. and Stratling, W.H. (1988) The matrix attachment regions of the chicken lysozyme gene co-map with the boundaries of the chromatin domain. EMBO J. **7**:655-664

^Philipsen, S., Talbot, D., Fraser, P. and Grosveld, F. (1990) The ß-globin dominant control region: Hypersensitive site 2. EMBO J. **9**:2159-2167

Picard, D. and Schaffner, W. (1985) Cell type preference of immunoglobulin κ and λ gene promoters. EMBO J. **4**:2831-2838

Pikaart, M., Irving, J. and Villeponteau, B. (1991) Decline in histone H5 phosphorylation during erythroid senescence in chick embryos Mech. of Ageing and Dev. **59**:189-195

Pina, B., Bruggemeier, U. and Beato, M. (1990) Nucleosome positioning modulates accessibility of regulatory proteins to the mouse mammary tumor virus promoter. Cell **60**:719-731

! Plumb, M., Frampton, J., Wainwright, H., Walker, M., Macleod, K., Goodwin, G. and Harrison, P. (1989) GATAAG; a *cis*-control region binding an erythroid-specific nuclear factor with a role in globin and non-globin gene expression. Nucl. Acids Res. **17**:73-91

Plumb, M.A., Lobanenkov, V.V., Nicolas, R.H., Wright, C.A., Zavou, S. and Goodwin, G.H. (1986) Characterizations of chicken erythroid nuclear proteins which bind to the nuclease hypersensitive regions upstream of the beta^{A}- and beta^{H}-globin genes. Nucl. Acids Res. **14**:7675-7693

Plumb, M.A., Nicolas, R.H., Wright, C.A. and Goodwin, G.H. (1985) Multiple sequence-specific DNA binding activities are eluted from chicken nuclei at low ionic strengths. Nucl. Acids Res. **13**:4047-4065

Postnikov. Y.V., Shick, V.V., Belyavsky, A.V., Khrapko, K.R., Brodolin, K.L., Nilolskaya, T.A. and Mirzabekov, A.D. (1991) Distribution of high mobility group proteins 1/2, E and 14/17 and linker histones H1 and H5 on transcribed and non-transcribed regions of chicken erythrocyte chromatin. Nucleic Acids Res. **19**:717-725

Price, D.H., Sluder, A.E. and Greenleaf, A.L. (1987) Fractionation of transcription factors

for RNA polymerase II from *Drosophila* Kc cell nuclear extracts. J. Biol. Chem. **262**:3244-3255

Privalsky, M.L, Sharif, M. and Yamamoto, K.R. (1990) The viral erbA oncogene protein, a constitutive repressor in animal cells, is a hormone-regulated activator in yeast. Cell **63**:1277-1286

Ptashne, M. and Gann, A.F. (1990) Activators and targets. Nature **346**:329-331

Ptashne, M. (1988) How eukaryotic transcriptional activators work.

Nature **335**:683-689

Ptashne, M. (1986) Gene regulation by proteins acting nearby and at a distance. Nature **322**:697-701

Pugh, B.F. and Tjian, R. (1992) Diverse transcriptional functions of the multisubunit eukaryotic TFIID complex. J. Biol. Chem. **267**:679-682

Pugh, B.F. and Tjian, R. (1991) Transcription from a·TATA-less promoter requires a multisubunit TFIID complex. Genes Dev. **5**:1935-1945

Pugh, B.F. and Tjian, R. (1990) Mechanism of transcriptional activation by Sp1: evidence for coactivators. Cell **61**:1187-1197

Puigdomènech, P., Paulau, J. and Crane-Robinson, C. (1980) Structure of sea-urchin-sperm histone-PHI-1 (H-1) in chromatin and in free solution. Trypsin digestion and spectroscopic studies. Eur. J. Biochem. **104**:263-270

^Purucker, M,. Bodine, D., Lin, H., McDonagh, K. and Nienhuis, A.W. (1990) Structure and function of the enhancer 3' to the human ^gamma globin gene. Nucl. Acids Res. **18**:7407-7415

^Rahuel, C., Vinit, M-A., Lemarchandel, V., Cartron, P-J and Romeo, P-H. (1992) Erythroid-specific activity of the glycophorin B promoter requires GATA-1 mediated displacement of a repressor. EMBO J. **11**:4095-4102

Ramanathan, B. and Smerdon, M.J. (1989) Enhanced DNA repair synthesis in hyperacetylated nucleosomes. J. Biol. Chem. **264**:110226-110234

Ranish, J.A., Lane, W.S. and Hahn, S. (1992) Isolation of two genes that encode subunits of the yeast transcription factor IIA. Science **255**:1127-1129

Rappaport, J. and Weinmann, R. (1987) Purine triphosphate beta-gamma bond hydrolysis requirements for RNA polymerase II transcription initiation and elongation. J. Biol. Chem. **262**:17510-17515

Rauscher, F.J., Sambucetti, L.C., Curran, T., Distel, R.J. and Spiegelman, B.M. (1988a)

Common DNA binding site for *Fos* protein complexes and transcription factor AP-1. Cell **52**:471-480

Rauscher, F.J., Cohen, D.R., Curran, T., Bos, T.J., Vogt, P.K., Bohmann, D., Tjian, R. and Franza, B.R. (1988b) *Fos*-associated protein p39 is the product of the *jun* proto-oncogene. Science **240**:1010-1016

Reeves, R. (1984) Transcriptionally active chromatin. Biochem. Biophys. Acta **782**:343-393

Reinberg, D., Horikoshi, M. and Roeder, R.G. (1987) Factors involved in specific transcription in mammalian RNA polymerase II: functional analysis of initiation factors IIA and IID and identification of a new factor operating at sequences downstream of the initiation site. J. Biol. Chem. **262**:3322-3330

Reinberg, D., and Roeder, R.G. (1987) Factors involved in specific transcription by mammalian RNA polymerase II: Purification and functional analysis of initiation factors IIB and IIE. J. Biol. Chem. **262**:3310-3321

Reitman, M., Lee, E., Westphal, H. and Felsenfeld, G: (1993) An enhancer/locus control region is not sufficient to open chromatin. Mol. Cell. Biol. **13**:3990-3998

Reitman, M., Lee, E., Westphal, H. and Felsenfeld, G. (1990) Site-independent expression of the chicken β^A-globin gene in transgenic mice. Nature **348**:749-752

Reitman, M. and Felsenfeld, G. (1988) Mutational analysis of the chicken beta-globin enhancer reveals two positive-acting domains. Proc. Natl. Acad. Sci. USA **85**:6267-6271

Renaud, J. and Ruiz-Carrillo, A. (1986) Fine analysis of the active H5 gene chromatin chicken erythrocytes at different stages of differentiation. J. Mol. Biol. **189**:217-226

Renz, M., Nehls, P. and Hozier, J. (1977) Involvement of histone H-1 in organization of chromosome fiber. Proc. Natl. Acad. Sci. USA **74**:1879-1883

Ridsdale, J.A., Hendzel, M.J., Delcuve, G.P. and Davie, J.R. (1990)

Histone acetylation alters the capacity of the H1 histones to condense transcriptionally active/competent chromatin. J. Biol. Chem. **265**:5150-5156

Rigaud, G., Roux, J., Pictet, R. and Grange, T. (1991) *In vivo* footprinting of rat TAT gene: dynamic interplay between the glucocorticoid receptor and a liver-specific factor. Cell **67**:977-986

Ringertz, N.R., Nyman, U. and Bergman, M. (1985) DNA replication and H5 histone exchange during reactivation of chick erythrocyte nuclei in heterokaryons. Chromosoma **91**:391-396

Ringertz, N.R. and Bolund, L. (1974) The nucleus during avian erythroid differentiation. In The Cell Nucleus, vol. 3 (ed. H. Busch), pp 417-446. San Francisco: Academic Press.

Riva, M., Schaffner, A.R.,Sentenac, A., Hartmann, G.R., Mustaev, A.A., Zaychikov, E.F. and Grachev, M.A. (1987) Active site labelling of the RNA polymerases A, B, and C from yeast. J. Biol. Chem. **262**:14377-14380

Roberts, S.B., Emmons, S. and Childs, G. (1989) Nucleotide sequences of *Caenorhabditis elegans* core histone genes. Genes for different histone classes share common flanking sequence elements. J. Mol. Biol. **206**:567-577

Rocha, E., Davie, J.R., vanHolde, K.E. and Weintraub, H. (1984) Differential salt fractionation of active and inactive genomic domains in chicken erythrocyte. J. Biol. Chem. **259** :8558-8563

Roeder, R.G. (1991) The complexes of eukaryotic transcription initiation: Regulation of preinitiation complex assembly. Trends Biochem. Sci. **16**:402-408

Roeder, R.G. (1976) Eukaryotic nuclear RNA polymerase. In RNA Polymerase, R. Losick and M. Chamberlin, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 285-329

Romanoff, A.L. (1960) The Avian Embryo: Structural and functional development (New York: Macmillan Company).

Roninson, I.B. and Ingram, V.M. (1983) Expression and partial DNA sequence of the chicken beta H-globin gene. J. Biol. Chem. **258**:802-809

Rosenfeld, P.J. and Kelley, T.J. (1986) Purification of NF-1 by DNA recognition site affinity chromatography. J. Biol. Chem. **261**:1398-1408

Roth, S.Y. and Allis, C.D. (1992) Chromatin condensation: does histone H1 dephosphorylation play a role? TIBS **17**:98-98

! Rousseau, S., Asselin, M., Renaud, J. and Ruiz-Carrillo, A. (1993) Transcription of the H5 gene is regulated by three differentiation-specific enhancers. Mol. Cell. Biol. **13**:4904-4917

! Rousseau, S., Renaud, J. and Ruiz-Carrillo, A. (1989) Basal expression of the histone H5 gene is controlled by positive and negative *cis*-acting sequences. Nucl. Acids Res. **17**:7495-7511

Roy, A.L., Meisterernst, M., Prognonec, P. and Roeder, R.G. (1991) Cooperative interaction of an initiator-binding transcription factor and the helix-loop-helix activator USF. Nature **354**:245-248

Ruiz-Carrillo, A., Affolter, M. and Renaud,J. (1983) Genomic organization of the genes

coding for the six main histones of the chicken: complete sequence of the H5 gene. J. Mol. Biol. **170**:843-859

Ruiz-Carrillo, A., Waugh, L.J., Littau, V.C. and Allfrey, V.G. (1974) Changes in histone acetyl content and in nuclear non-histone protein composition of avian erythroid cells at different stages of maturation. J. Biol. Chem. **249**:7358-7368

Rupp, R.A.W., Kruse, U., Multhaup, G., Gobel, U., Beyreuther, K. and Sippel, A.E. (1990) Chicken NF1/TGGCA proteins are encoded by at least three independent genes. Nucl. Acids Res. **18**:2607-2616

Rupp, R.A., Nicolas, R.H., Borgmeyer, U., Lobanenkov, V.V., Plumb, M.A., Sippel, A.E. and Goodwin, G.H. (1988) TGGCA protein is present in erythroid nuclei and binds within the nuclease-hypersensitive sites 5′ of the chicken beta H- and beta A-globin genes. Eur. J. Biochem. **177**:505-511

Ruppert, S., Wang, E.H. and Tjian, R. (1993) Cloning and expression of human TAFII250 : A TBP-associated factor implicated in cell-cycle regulation. Nature **362**:175-179

Sambucetti, L.C. and Curran, T. (1986) The *Fos* protein complex is associated with DNA in isolated nuclei and binds to DNA cellulose. Science **234**:1417-1419

Samuels, M. and Sharp, P.A. (1986) Purification and characterization of a specific RNA polymerase II transcription factor. J. Biol. Chem. **261**:2003-2013

Samuels, M., Fire, A. and Sharp, P.A. (1982) Purification and characterization of factors mediating accurate transcription by RNA polymerase II. J. Biol. Chem. **257**:14419-14427

Santoro, C., Mermod, N., Andrews, P.C. and Tjian, R. (1988) A family of human CCAAT-box-binding proteins active in transcription and DNA replication: cloning and expression of multiple cDNAs. Nature **334**:218-224

Sarnow, P., Hearing, P., Anderson, C.W., Reich, H. and Levine, A.J. (1982) Identification and characterization of an immunologically conserved adenovirus early region 11,000 MR protein and its association with the nuclear matrix. J. Mol. Biol. **162**:565-583

Sawadogo, M. and Sentenac, A. (1990) RNA polymerase B (II) and general transcription factors. Annu. Rev. Biochem. **9**:711-754

Sawadogo, M. (1988) Multiple forms of the human gene-specific transcription factor USF.II. DNA binding properties and transcriptional activity of the purified HeLa USF. J. Biol. Chem. **263**:11994-12001

Sawadogo, M. and Roeder, R.G. (1985) Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative *in vitro* assay. Proc. Natl. Acad. Sci. USA **82**:4394-4398

Sawadago, M., and Roeder, R.G. (1984) Energy requirements for specific transcription initiation by the human RNA polymerase II system. J. Biol. Chem. **259**:5321-5326

Schaufele, F., Cassil, J.A., West, B.L. and Reudelhuber, T. (1990) Resolution by diagonal gel mobility shift assays of multisubunit complexes binding to a functionally important element of the rat growth hormone gene promoter. J.Biol. Chem. **265**:14592-14598

Scherrer, K., Marcaud, L., Zajdela, F., London, J.M. and Gross, F. (1966) Patterns of RNA metabolism in a differentiated cell - a rapidly labelled unstable 60S RNA with messenger properties in duck erythroblasts. Proc. Natl. Acad. Sci. USA **56**:1571-1578

Schmid, A., Fascher, K.-D. and Horz, W. (1992) Nucleosome disruption at the yeast PHO5 promoter upon PHO5 induction occurs in the absence of DNA replication. Cell **71**:853-864

Schmidt, M.C., Zhou, Q. and Berk, A.J. (1989a) Sp1 activates transcription without enhancing DNA-binding activity of the TATA box factor. Mol. Cell. Biol. **9**:3299-3307

Schmidt, M.C., Kao, C.C., Pei, R. and Berk, A.J. (1989b) Yeast TATA-box transcription factor gene. Proc. Natl Acad. Sci. USA **86**:7785-7789

Schule, R., Umesono, K., Mangelsdorf, D.J., Bolado, J., Pike, J.W. and Evans, R.M. (1990) *Jun-Fos* and receptors for vitamins A and D recognize a common response element in the human osteocalcin gene.

Cell **61**:497-504

Schultz, M.C., Reeder, R.H. and Hahn, S. (1992) Variants of the TATA-binding protein can distinguish subsets of RNA polymerase I, II, and III promoters. Cell **69**:697-702

^Schwartzbauer, G., Schlesinger, K. and Evans, T. (1992) Interaction of the erythroid transcription factor cGATA-1 with a critical autoregulatory element. Nucl. Acids Res. **20**:4429-4436

Schweiger, H.G. (1962) Pathways of metabolism in nucleate and anucleate erythrocytes. Int. Rev. Cytol. **13**:135-201

Segers, A., Muyldermans, S. and Wyns, L. (1991) The interaction of histone H5 and its globular domain with core particles, depleted chromatosomes, polynucleosomes, and a DNA decamer. J. Biol. chem. **266**:1502-1508

Sentenac, A. (1985) Eukaryotic RNA polymerases. CRC Crit. Rev. Biochem. **18**:31-90

Serfling, E., Jasin, M. and Schaffner, W. (1985) Enhancers and eukaryotic gene-transcription - review. Trends Genet **1**:224-230

Shannon, M.F., Wigley, P.L. and Wells, J.R.E. (1985) Histone H5 and H1° cross-reacting material is restricted to erythroid cells in chicken. FEBS Lett. **186**:180-186

Simmen, K.A., Bernves, J., Parry, H.D., Stunnenberg, H.G., Berkenstam, A., Cavalini, B., Egly, J.-M. and Mattaj, I.W. (1991) TFIID is required for *in vitro* transcription of human U6 gene by RNA polymerase III. EMBO J. **10**:1853-1862

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

Simpson, R.T. (1978a) Structure of chromatin containing extensively acetylated H3 and H4. Cell **13**:691-699

Simpson, R.T. (1978b) Structure of chromatosome, a chromatic particle containing 160-base pairs of DNA and all histones. Biochemistry **17**:5524-5531

Sinclair, G.D. and Brasch, K. (1975) Nucleated erythrocyte - model of cell-differentiation. -> Review. Rev. Can. Biol. **34**:287-303

Smale, S.T., Schmidt, M.C., Berk, A.J. and Baltimore, D. (1990) Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirement for mammalian transcription factor IID. Proc. Natl. Acad. Sci. USA **87**:4509-4513

Small, J.V. and Davies, H.G. (1972) Erythropoiesis in yolk sac of early chick-embryo - electron-microscope and microspectrophotometric study. Tissue Cell **4**:341-352

Sopta, M., Burton, Z. and Greenblatt, J. (1989) Structure and associated DNA-helicase activity of a general transcription initiation factor that binds polymerase II. Nature **341**:410-414

Sotirov, N. and Johns, E.W. (1972) Quantitative difference in the content of the distal F2C between chicken erythrocytes and erythroblasts. Exp. Cell Res. **73**:13-16

Sperling, L. and Klug, A. (1977) X-ray studies on native chromatin. J. Mol. Biol. **112**:253-263

! Sposi, M.N., Zon, L.I., Care, A., Valtieri, M., Testa, U., Gabianelli, M., Mariani, G., Bottero, L., Mather, C., Orkin, S.H. and Peschle, C. (1992) Cell cycle-dependent initiation and lineage-dependent abrogation of GATA-1 expression in pure differentiating hematopoietic progenitors. Proc. Natl. Acad. Sci. USA **89**:6353-6357

Staufenbiel, M. and Deppert, W. (1983) Different structural systems of the nucleus are targets for SV40 large T-antigen. Cell **33**:173-181

Sterner, R., Boffa, L.C., Chen, T.A., and Allfrey, V.G. (1987) Cell cycle-dependent changes in conformation and composition of nucleosomes containing human

histone gene sequences. Nucleic Acids Res. **15**:4375-4391

Stief, A., Winter, D.M., Stratling, W.H. and Sippel, A.E. (1989) A nuclear DNA attachment element mediates elevated and position-independent gene activity. Nature **341**:343-345

Strauss, E. and Orkin, S. (1992) Human ß-globin locus control region hypersensitive site 3: *In vivo* protein-DNA interactions. Proc. Natl. Acad. Sci. USA **89**:5809-5813

Strauss, E., Andrews, N.C., Higgs, D.R. and Orkin, S. (1992) *in vivo* footprinting of the human *α*-globin locus upstream regulatory element by guanine/adenine ligation-mediated PCR. Mol. Cell. Biol. **12**:2135-2142

Strickland, W.N., Strickland, M., Brandt, W.F., Von Holt, C., Lehmann, A. and Wittmann-Liebold, B. (1980) The primary structure of histone-H1 from sperm of the sea-urchin *parechinus-angulosus*. 2. Sequence of the C-terminal CNBR peptide and the entire primary structure. Eur. J. Biochem. **104**:567-578

Stringer, K.F., Ingles, C.J. and Greenblatt, J. (1990) Direct and selective binding of an acidic transcriptional activation domain to the TATA-box factor TFIID. Nature **345**:783-786

Struhl, K. (1987) Promoters, activator proteins and the mechanism of transcriptional initiation in yeast. Cell **49**:295-297

Su, W., Jackson, S., Tjian, R., and Echols, H. (1991) DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1. Genes Dev. **5**:820-826

Sumimoto, H., Ohkuma, Y., Sinn, E., Kato, H., Shimasaki, S., Horikoshi, M. and Roeder, R.G. (1991) Conserved sequence motifs in the small subunit of human general transcription factor TFIIE. Nature **354**:401-404

Sun, J.-M., Hendzel, M.J. and Davie, J.R. (1992a) Nuclear matrix proteins bind very tightly to specific regions of the chicken histone H5 gene. Biochem. Cell Biol. **70**:822-829

Sun, J.-M. Wiaderkiewicz,R. and Ruiz-Carrillo, A. (1989) Histone H5 in the control of DNA synthesis and cell proliferation. Science **245**:68-71

Sun, J.M., Almi, Z., Lurz, R. and Ruiz-Carrillo, A. (1990) Replacement of histone H1 by H5 *in vivo* does not change the nucleosome repeat length of chromatin but increases its stability. EMBO J. **9**:1651-1658

Sung, M.T. and Freelender, E.F. (1978) Studies of *in vivo* phosphorylation of histone H5. Biochemistry **17**:1884-1890

Sung, M.T., Harford, J., Bundman, J. and Vidalakas, G. (1977) Metabolism of histone in

avian erythroid cells. Biochemistry **16**:279-285

^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

Taggart, A.K., Fisher, T.S. and Pugh, B.F. (1992) The TATA-binding protein and associated factors are components of pol III transcription factor TFIIIB. Cell **71**:1015-1028

Takada, R., Nakatani, Y., Hoffmann, A., Kokubo, T., Hasegawa, S., Roeder, R. and Horikoshi, M. (1992) Identification of human TFIID components and direct interaction between a 250-kDa polypeptide and the TATA box-binding protein (TFIID tau). Proc. Natl. Acad. Sci. USA **89**:11809-11813

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

^Talbot, D., Philipsen, S., Fraser, P. and Grosveld, F. (1990) Detailed analysis of the site 3 reg ion of the human ß-globin dominant control region. EMBO J. **9**:2169-2178

Tanese, N., Pugh, B.F. and Tjian, R. (1991) Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. Genes and Dev. **5**:2212-2224

Taylor, I.C.A., Workman, J. L., Schuetz, T.J. and Kingston, R.E. (1991) Facilitated binding of GAL4 and heat shock factor to nucleosomal templates: Differential function of DNA-binding domains. Genes Dev. **5**:1285-1298

Theveny, B., Bailly, A., Rauch, C., Delain, E. and Milgrom, E. (1987) Association of DNA-bound progesterone receptors. Nature **329**:79-81

Thoma, F. and Kollar, Th. (1981) Unravelled nucleosomes, nucleosome beads and higher-order structures of chromatin: Influence of non-histone components on histone H1. J. Mol. Biol. **149**:709-733

Thoma, F., Kollar, Th. and Klug, A. (1979) Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. J. Cell Biol. **83**:403-427

Thomas, J.O. and Wilson, C.M. (1986) Selective radiolabelling and identification of a strong nucleosome binding site on the globular domain of histone H5. EMBO J. **5**:3531-3537

Thomas, J.O., Rees, C. and Pearson, E.C. (1985) Histone H5 promotes the association of condensed chromatin fragments to give pseudo-higher-order structures. Eur. J. Biochem. **147**:143-151

Till, J.E. and McCulloch, E.A. (1980) Hematopoeitic stem cell differentiation. Biochem. Biophys. Acta. **605**:431-459

Till, J.E., McCulloch, E.A. and Sininovitch, L. (1964) Stochastic model of stem cell proliferation based on growth of spleen colony-forming cells. Proc. natl. Acad. Sci. USA **51**:29-36

Timmers, H.T.M. and Sharp, P.A. (1991) The mammalian TFIID protein is present in two functionally distinct complexes. Genes and Dev. **5**:1946-1956

Tobin, A.J., Hansen, D.A., Chapman, B.S., McCabe, J.B. and Septor, E.A. (1981) Globin gene expression during chicken development. In Hemoglobins in Development and Differentiation (ed. G. Stamatoyannopoulos and A.W. Nienhuis), pp. 203-213. New York: Alan R. Liss, Inc.

Tooze, J. and Davies, H.G. (1967) Light- and electron-microscope studies on spleen of newt *Triturus cristatus* - fine structure of erythropoietic cells. J. Cell. Sci. **2**:617-638

Torres-Martinez, S. and Ruiz-Carrillo, A. (1982) Nucleosomes containing histones H-1 or H-5 are closely interspersed in chromatin. Nucl. Acids Res. **10**:2323-2335

Townes, T.M. and Behringer, R.R. (1990) Human globin locus activation region (LAR): role in temporal control. Trends Genet. **6**:219-223

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

Trainor, C.D. and Engel, J.D. (1989) Transcription of the chicken histone H5 gene is mediated by distinct tissue-specific elements within the promoter and the 3' enhancer. Mol. Cell. Biol. **9**:2228-2232

Trainor, C.D., Stamler, S.J. and Engel, J.D. (1987) Erythroid-specific transcription of the chicken histone H5 gene is directed by a 3' enhancer. Nature **328**:827-830

! Trentesaux, C., Nyoung, M.N., Aries, A., Morceau, F., Ronchi, A., Ottolenghi, S., Jardiller, J-C. and Jeannesson, P. (1993) Increased expression of GATA-1 and NFE-2 erythroid-specific transcription factors during aclacinomycin-mediated differentiation of human erythroleukemic cells. Leukemia **7**:452-457

Tsai, S., Strauss, E. and Orkin, S.H. (1991) Functional analysis and *in vivo* footprinting implicate the erythroid transcription factor GATA-1 as a positive regulator of its own promoter. Genes Dev. **5**:919-931

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

Tsai, S.Y., Tsai, M.J., Kops, L.E., Minghetti, P.P. and O'Malley, B.W. (1981) Transcription factors from oviduct and hela-cells are similar. J. Biol. Chem. **256**:13055-13059

Tuan, D., Solomon, W., London, I.M. and Lee, D.P. (1989) An erythroid-specific, developmental-stage-independent enhancer far upstream of the human "beta-like-globin" genes. Proc. Natl. Acad. Sci. USA **86**:2554-2558

Urban, M.K., Neelin, J.M. and Betz, T.W. (1980) Correlation of chromatin composition with metabolic changes in nuclei of primitive erythroid cells from chicken embryos. Can. J. Biochem. **58**:726-731

Usheva, A., Maldonado, E., Goldring, A., Lu, H., Houbavi, C., Reinverg, D. and Aloni, Y. (1992) Specific interaction between the nonphosphorylated form of RNA polymerase II and the TATA-binding protein. Cell **69**:871-881

Usuda, Y., Kubota, A., Berk, A.J. and Handa, H. (1991) Affinity purification of transcription factor IIA from HeLa cell nuclear extracts. EMBO J. **10**:2305-2310

van Driel, R., Humbel, B. and de Jong, L. (1991) The Nucleus: A black box being opened. J. Cell. Biochemistry **47**:311-316

Van Dyke, M.W., Sawadogo, M. and Roeder, R.G. (1989) Stability of transcription complexes on class II genes. Mol. Cell. Biol. **9**:342-344

Van Dyke, M.W., Roeder, R.G. and Sawadogo, M. (1988) Physical analysis of transcription preinitiation complex assembly on a class II gene promoter. Science **241**:1335-1338

van Holde, K.E. (1988) in Chromatin (Rich, A., Ed.), Springer-Verlag, New York

van Wijnen, A.J., Bidwell, J., Lian, J.B., Stein, J.L., and Stein, G.S. (1992) Stairway assays: rapid localization of multiple protein/DNA interaction sites in gene-regulatory 5' regions. Biotechniques **12**:400-407

* Vassetzky, Y.S., de Moura Gallo, C.V., Bogdanova, A.N., Razin, S.V. and Scherrer, K. (1993) The sequence-specific nuclear matrix binding factor F6 is a chicken GATA-like protein. Mol. Gen. Genetic. **238**:309-314

Vogelstein, B., Pardo., D.M. and Coffey, D.S. (1980) Supercoiled loops and eukaryotic DNA-replication. Cell **22**:79-85

Waitz, W. and Loidl, P. (1991) Cell cycle dependent association of c-*myc* protein with the nuclear matrix. Oncogene **6**:29-35

^Wall, L., deBoer, E. and Grosveld, F. (1988) The human ß-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. Genes Dev. **2**:1089-1100

Walters, M. and Martin, D.I.K. (1992) Functional erythroid promoters created by interaction of the transcription factor GATA-1 with CACCC and AP-1/NFE-2 elements. Proc. Natl. Acad. Sci. **89**:10444-10448

Wang, W., Carey, M. and Gralla, J.D. (1992a) Polymerase II promoter activation: closed complex formation and ATP-driven start site opening. Science **255**:450-453

Wang, W., Gralla, J.D. and Carey, M. (1992b) The acidic activator GAL4-AH can stimulate polymerase II transcription by promoting assembly of a closed complex requiring TFIID and TFIIA. Genes and Development **6**:1716-1727

Warren, A.C. and Cook, P.R. (1978) Supercoiling of DNA and nuclear conformation during cell-cycle. J. Cell Sci. **30**:211-226

Weatherall, D.J., Clegg, J.B. and Wood, W.G. (1976) A model for the persistence of reactivation of fetal hemoglobin production. Lancet **2**:660-663

Weintraub, H. (1985) Assembly and propagation of repressed and depressed chromosomal states. Cell **42**:705-711

Weintraub, H. (1978) The nucleosome repeat length increases during erythropoiesis in the chick. Nucl. Acids Res. **5**:1179-1188

Weinzierl, R.O.J., Dynlacht, B.D. and Tjian, R. (1993) Largest subunit of *Drosophila* transcription factor IID directs assembly of a complex containing TBP and a coactivator. Nature **362**:511-517

Weisbrod, S. (1982) Active Chromatin. Nature **297**:289-295

Westin, G., Gerster, T., Muller, M.M., Schaffner, G. and Schaffner, W. (1987) OVEC, a versatile system to study transcription in mammalian cells and cell-free extracts. Nucl. Acids Res. **15**:6787-6798

White, J., Brou, C., Wu, J., Lutz, Y., Moncollin, V. and Chambon, P. (1992) The acidic transcriptional activator GAL-VP16 acts on preformed template-committed complexes. EMBO J. **11**:2229-2240

White, R.J. and Jackson, S.P. (1992) Mechanism of TATA-binding protein recruitment to a TATA-less class III promoter. Cell **71**:1041-1053

#Whitelaw, E., Tsai, S.-F., Hogben, P. and Orkin, S.H. (1990) Regulated expression of globin chains and the erythroid transcription factor GATA-1 during erythropoiesis in the developing mouse. Mol. Cell. Biol. **10**:6596-6606

Whitelaw, E., Hogben, P., Hanscombe, O. and Proudfoot, N.J. (1989) Transcriptional promiscuity of the human alpha-globin gene. Mol. Cell. Biol. **9**:241-251

Whitlock, J.P. Jr. and Stein, A. (1978) Folding of DNA by histones which lack their NH_2-

terminal regions. J. Biol. Chem. **253**:3857-3861

Widom, J. and Klug, A. (1985) Structure of the 300A chromatin filament: X-ray diffraction from oriented samples. Cell **43**:207-213

Wigley, P.L. and Wells, J.R.E. (1987) H5 gene-specific *trans*-activation by nuclear extracts from avian erythroid cells. Mol. Cell. Biol. **7**:3853-3856

Williams, A.F. (1972) DNA synthesis in purified populations of avian erythroid cells. J. Cell Sci. **10**:27-46

Wolffe, A.P. (1990) Transcription complexes. Prog. Clin. Biol. Res. **322**:171-186

Wolffe, A.P. and Brown, D.D. (1988) Developmental regulation of two 5S ribosomal RNA genes. Science **241**:1626-1632

Workman, J.L. and Kingston, R.E. (1992) Nucleosome core displacement *in vitro* via a metastable transcription factor-nucleosome complex. Science **258**:1780-1784

Workman, J.L., Taylor, E.C.A. and Kingston, R.E. (1991) Activation domains of stably bound GAL4 derivatives alleviate repression of promoters by nucleosomes. Cell **64**:533-544

Workman, J.L., Abmayr, S.M., Cromlish, W.A. and Roeder, R.G. (1988) Transcription by the immediate early protein of pseudorabies virus during *in vitro* nucleosome assembly. Cell **55**:211-219

Workman, J.L. and Roeder, R.G. (1987) Binding of transcription factor TFIID to the major late promoter during *in vitro* nucleosome assembly potentiates subsequent initiation by RNA polymerase II. Cell **51**:613-622

Wu, R.S., Panusz, H.T., Hatch, C.L. and Bonner, W.M. (1986) Histones and their modification. CRC Crit. Rev. Biochem. **20**:201-263

Yaguchi, M., Roy, C., Dove, M. and Seligy, V. (1977) Amino-acid sequence homologies between H-1 and H-5 histones. Biochem. Biophys. Res. Commun. **76**:100-106

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 transcription factor NF-E1 multigene family. Genes Dev. **4**:1650-1662

Yang, H.-Y. and Evans, T. (1992) Distinct roles for the two cGATA-1 finger domains. Mol. Cell. Biol. **12**:4562-4570

Young, R.A. (1991) RNA polymerase II. Annu. Rev. Biochem. **60**:689-715

Youssoufian, H., Zon, L.I., Orkin, S.H., D'Andrea, A.D. and Lodish, H.F. (1990) Structure and transcription of the mouse erythropoietin receptor gene. Mol. Cell. Biol.

10:3675-3682

^Yu, C.-Y., Motamed, K., Chen, J., Bailey, A.D. and Shen, C.-K.,J. (1991) The CACC box upstream of human embryonic epsilon globin gene binds Sp1 and is a functional promoter element *in vitro* and *in vivo*. J. Biol Chem. **266**:8907-8915

^Yu, C.-Y., Chen, J., Lin, L.-I., Tam, M. and Shen, C.-K.,J. (1990) Cell type-specific protein-DNA interactions in the human zeta-globin upstream promoter region: Displacement of Sp1 by the erythroid cell-specific factor NF-E1. Mol. Cell. Biol. **10**:282-294

Zawel, L. and Reinberg, D. (1993) Initiation of transcription by RNA polymerase II: A multi-step process. Prog. Nucl. Acids Res. Mol. Biol. **44**:67-108

Zehring, W.A., Lee, J.M., Weeks, J.R., Jokerst, R.S. and Greenleaf, A.L. (1988) The C-terminal repeat domain of RNA polymerase II largest subunit is essential *in vivo* but is not required for accurate transcription initiation *in vitro*. Proc. Natl. Acad. Sci. USA **85**:3698-3702

Zenke, M., Munoz, A., Sap, J., Vennstrom, B. and Beug, H. (1990) v-erbA oncogene activation entails the loss of hormone-dependent regulator activity of c-erbA. Cell **61**:1035-1049

Zenke, M., Kahn, P., Disela, C., Vennstrom, B., Leutz, A., Keegen, K., Hayman, M., Choi, H.-R., Yew, N.S., Engel, J.D. and Beug, H. (1988) V-*erb*A specifically suppresses transcription of the avian erythrocyte anion transporter (band 3) gene. Cell **52**:107-119

Zentgraf, H., Scheer, U. and Franke, W.W. (1975) Characterization and localization of RNA synthesized in mature avian erythrocytes. Exp. Cell Res. **96**:81-95

Zhou, Q., Boyer, T.G. and Berk, A.J. (1993) Factors (TAFs) required for activated transcription interact with TATA box-binding protein conserved core domain. Genes and Dev. **7**:180-187

Zhou, Q., Lieberman, P.M., Boyer, T.G. and Berk, A.J. (1992) Holo-TFIID supports transcriptional stimulation by diverse activators and from a TATA-less promoter. Genes and Dev. **6**:1964-1974

Zhou, Q., Schmidt, M.C. and Berk, A.J. (1991) Requirement of acidic amino acid residues immediately amino-terminal to the conserved domain of *Saccharomyces cerevisiae* TFIID. EMBO J. **10**:1843-1852

Zon, L.I., Mather, C., Burgess, S., Bolce, M.E. Harland, R.M. and Orkin, S.H. (1991a) Expression of GATA-binding proteins during embryonic development in Xenopus laevis. Proc. Natl Acad. Sci. USA **88**:10642-10646

Zon, L.I., Youssoufian, H., Mather, C., Lodish, H.F. and Orkin, S.H. (1991b) Activation

of the erythropoietin receptor promoter by transcription factor GATA-1. Proc. Natl. Acad. Sci. USA **88**:10638-10641

Zon, L.I. Gurish, M.F., Stevens, R.L., Mather, C., Reynolds, D.S., Austen, K.F. and Orkin, S.H. (1991c) GATA-binding transcription factors in mast cells regulate the promoter of the mast cell carboxypeptidase A gene. J. Biol. Chem. **266**:22948-22953

Zon,L.I., Tsai, S.-F. Burgess, S., Matsudaira,P., Bruns, G.A.P. and Orkin, S.H. (1990) The major human erythroid DNA-binding protein (GF-1): primary sequence and localization of the gene to the X chromosome. Proc. Natl. Acad. Sci. USA **87**:668-672