UCSF
UC San Francisco Electronic Theses and Dissertations

Title
The rat alpha 1-acid glycoprotein (gene structure, sequence analysis, function prediction and identification of regulatory region for glucocorticoid induction)

Permalink
https://escholarship.org/uc/item/7pm5158k

Author
Liao, Yu-Cheng J

Publication Date
1986

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

The Rat Alpha 1-Acid Glycoprotein
(Gene structure, sequence analysis, function prediction and identification of regulatory region for glucocorticoid induction)

by

Yu-Cheng J. Liao

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

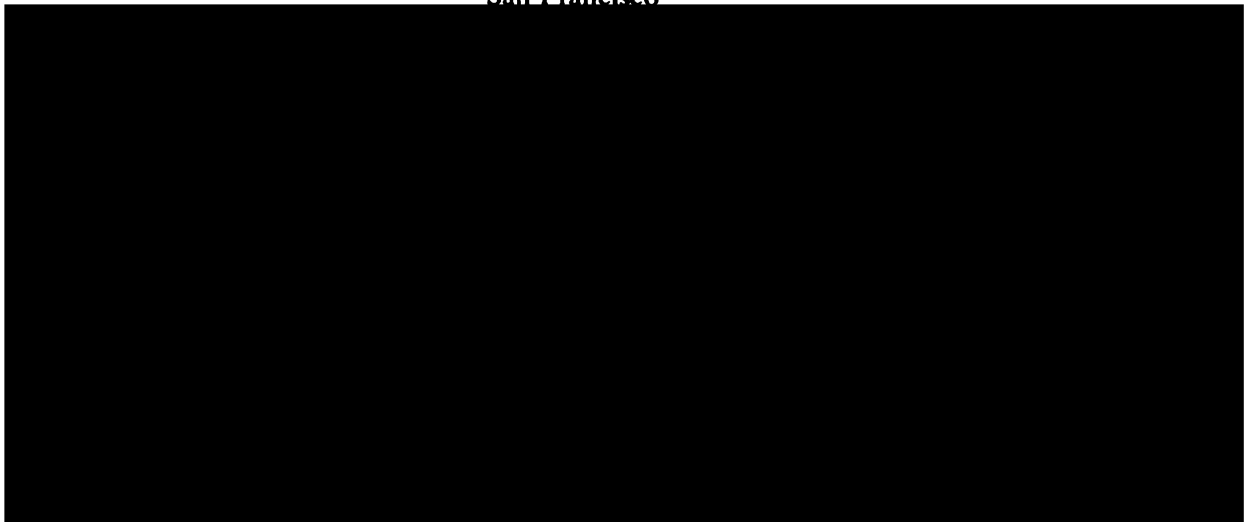
in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco



. . . .

Date University Librarian

Degree Conferred: MAR 30 1986 . . . .

©1986

YU-CHENG JORDAN LIAO

All Rights Reserved

ABBREVIATIONS

ACTH	Adrenocorticotropic hormone
α_1-AGP	Alpha 1-acid glycoprotein
bp	Base pair
CRP	C-reactive protein
CAT	Chloramphenicol acetyltransferase
cDNA	Complimentary DNA
Ci	Curie
ddNTP	dideoxy-nucleotide-5'-triphosphate
dNTP	2'-deoxy-nucleotide-5'triphosphate
DNAse	deoxyribonuclease
ddGTP	dideoxy guanisine triphosphate
ddATP	dideoxy adenosine triphosphate
ddTTP	dideoxy thymidine triphosphate
ddCTP	dideoxy cytidine triphosphate
ds	double strand
EGF	Epidermal growth factor
EDTA	ethylene diaminetetracetic acid, disodium salt
HTC	Hepatoma cell
IL-1	Interleukin-1
IPTG	isopropyl-1-β-D-thiogalactopyranoside
mRNA	Messenger RNA
MMTV	Mouse mammary tumor virus
nRNA	Nuclear RNA
PBS	phosphate-buffered saline

SAA	Serum amyloid A protein
SDS	Sodium dodecyl sulfate
ss	single strand
SSC	Standard saline citrate
Tris	Tris (Hydroxymethyl) aminomethane
UV	ultraviolet
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

ACKNOWLEDGMENTS

I would like to thank my supervisor, Dr. Edward A. Smuckler. In the beginning, he showed me the interest and significance of a single acute phase protein gene and from there he led me into an opening world of molecular biology and disease research. He deserves total credit for initiating my study and guiding it toward completion.

I shall extend my gratitude to Dr. John M. Taylor for his pioneering work on α_1-AGP and the gift of cDNA probes. I also would like to thank Dr. James H. McKerrow and Dr. Leon Levintow for their valuable advice and suggestions. Finally, I would like to thank Ms. Claudia Schumann for her assistance in the preparation of this thesis.

ABSTRACT

The Rat α_1-Acid Glycoprotein Gene

Yu-Cheng J. Liao

The work in this dissertation describes the isolation and characterization of the rat α_1-AGP gene and its relationship to the cDNA and gene of human α_1-AGP. α_1-AGP is one of the acute phase proteins. Its prominent elevation and convenient induction in whole animal by subcutaneous injection of turpentine or addition of glucocorticoid in tissue culture respectively make it a very useful model to study the various steps of gene regulation and protein biosynthesis.

The 4.8 kilobase (kb) nucleotide sequence of the rat genomic clone lambda rAGP-40 contains the entire rat α_1-AGP gene including "TATA" box, polyadenylylation signal "AATAAA", six exons and several potential glucocorticoid receptor binding sites. By linking the 5' region of the rat AGP gene with chloramphenicol acetyltransferase cDNA, this 5' end sequence was shown to be responsible for a 50-fold promotion of gene expression.

During this project, cloning human α_1-AGP cDNA, and sequencing of its gene was also being carried out by Luciana et al (1985) of European Molecular Biology Laboratory, Heidelberg, Germany. Based on the partially published data, a preliminary comparison were made between the human and rat α_1-AGP cDNA and gene for studying the possible mechanism of intron splicing and structural relationships.

Using both the nucleotide and amino acid consensus sequences, an extensive comparison between the rat and human α_1-AGP was carried out to identify sequence and structural conservation which might give clues of potential functions of α_1-AGP. Finally, the available gene sequence was reconstructed to allow a fusion of its 5'end region with an assayable gene to identify its promotion effect by glucocorticoid induction.

TABLE OF CONTENTS

Thesis Title Page i

Abbreviations ii

Acknowledgments iv

Abstract v

Table of Contents vii

List of Figures x

CHAPTER 1: GENERAL INTRODUCTION

1.1 Introduction 1

1.2 Plasma proteins 4

1.3 Acute phase reaction 6

1.4 The α_1-acid glycoprotein 10

1.5 Working hypothesis 14

CHAPTER 2: ISOLATION AND CHARACTERIZATION OF THE RAT α_1-AGP GENE

2.1 Introduction 18

2.2 Methods and materials 19

2.2.1 Preparation of genomic DNA 19

2.2.2 Genomic library construction 19

2.2.3 Selection and isolation of the genomic clone 20

2.2.4 Preparation of bacteriophage lambda DNA 23

2.2.5 Restriction mapping and Southern blot analysis 24

2.3 Discussion 33

CHAPTER 3: NUCLEOTIDE SEQUENCING OF THE RAT α_1-AGP GENE

3.1 Introduction 38

3.2 Methods and materials 39

3.2.1 Materials and sequencing techniques 39

3.2.2 Sequencing strategies 52

3.3 Results 53

3.4 Discussion 54

CHAPTER 4: SEQUENCE ANALYSIS

4.1 Introduction 62

4.2 Methods and materials 65

4.3 Results 65

4.4 Discussion 121

CHAPTER 5: REGULATION OF THE RAT α_1-AGP GENE

5.1 Introduction 136

5.2 Methods and materials 137

5.3 Results 139

5.4 Discussion 146

APPENDIX A 150

APPENDIX B 157

APPENDIX C 165

REFERENCES 177

LIST OF FIGURES

Fig.No.	Description	Page No.
Fig. 2.1.	An agarose gel electrophoresis of restriction endonuclease -digested DNAs of both rat AGP lambda 40 and rat liver.	25
Fig. 2.2.	An agarose gel replica of Fig. 2.1.	27
Fig. 2.3.	Southern analysis of recombinant lambda (AGP lambda 40) and rat cellular DNA (rat hepatoma cell).	31
Fig. 2.4.	The mapping of the rat α_1-AGP gene.	36
Fig. 3.1.	A schematic illustration of M-13 sequencing method.	41
Fig. 3.2.	Representation of the life cycle of E. coli bacteriophage M-13.	44
Fig. 3.3.	Diagrammatic representation of the M-13 bacteriophage vector.	46
Fig. 3.4.	A schematic representation of cloning a DNA fragment into M-13 bacteriophage.	49
Fig. 3.5.	The sequencing strategy - 3 steps.	55
Fig. 4.1.	A homology plot of the rat α_1-AGP vs. itself.	67
Fig. 4.2.	A homology plot of the rat α_1-AGP vs. itself (with leader signal.	69
Fig. 4.3.	A. A self-homology plot of the human pre-α_1-AGP amino acid sequences. B. A self-homology plot of the mature rat α_1-AGP amino acid sequences.	71
Fig. 4.4.	cDNA and amino acid sequence of the rat α_1-AGP.	73
Fig. 4.5.	A. A summarizing figure representing the counting of translational triplet codes for the rat pre-α_1-AGP. B. The rat pre-α_1-AGP represented by a single letter amino acid abbreviation.	75
Fig. 4.6.	A. Circulating form of rat α_1-AGP. B. An entire amino acid sequence of the rat α_1-AGP in abbreviations.	77
Fig. 4.7.	Nucleotide and amino acid sequences of the human pre-α_1-AGP cDNA and protein.	79

Fig. 4.8.	A. Translation count of human pre-α_1-AGP. B. Abbreviated amino acid sequences of the human pre-α_1-AGP.	81
Fig. 4.9.	A. Translation count of mature human α_1-AGP. B. Abbreviated amino acid sequences of the mature human AGP.	83
Fig. 4.10.	A comparison of amino acid residues of human and rat pre-α_1-AGP.	85
Fig. 4.11.	The amino acid comparison between human and rat pre-α_1-AGP classified into exonic segments.	87
Fig. 4.12.	A. A homology plot of the amino acid residues of the rat and human pre-α_1-AGP. B. Same as above, except plotting stringency is reduced slightly from 7:7 to 6:7.	89
Fig. 4.13.	C. Same as Fig. 4.8, except that the stringency is reduced to 5:7. D. Same as above, except that the plotting stringency is reduced to 4:7.	91
Fig. 4.14.	E. The plotting stringency is 3:7. F. The plotting stringency is 2:7.	93
Fig. 4.15.	Comparisons of human and rat α_1-AGP glycosylation sites.	95
Fig. 4.16.	Nucleotide sequence homology between the rat and human α_1-AGP cDNA.	97
Fig. 4.17.	A homology plot of nucleotide sequences of human and rat α_1-AGP cDNA.	99
Fig. 4.18.	Homology plot of the amino acid residues of the rat and human pre-α_1-AGP with matching stringency of 5:7.	101
Fig. 4.19.	Same as above, plotting stringency increased to 6:7.	103
Fig. 4.20.	Same as above, plotting stringency increased to highest level.	105
Fig. 4.21.	A comparison of human and rat α_1-AGP gene based on sizes of exons and introns.	107
Fig. 4.22.	A comparison of splicing junctions of human and rat α_1-AGP gene.	109
Fig. 4.23.	A hydropathy representation of the mature human α_1-AGP.	111
Fig. 4.24.	A hydropathy figure of the mature rat α_1-AGP.	113

Fig. 4.25.	Comparison of rat's and human α_1-AGP hydropathy.	115
Fig. 4.26.	A hydropathy plot of human pre-α_1-AGP.	117
Fig. 4.27.	A hydropathy plot of rat pre-α_1-AGP.	119
Fig. 5.1.	Construction of pAGP-CAT 1.	142
Fig. 5.2.	Assay of CAT activity in mammalian cells.	144
Appendix CI.	Map of the rat α_1-AGP gene	165
Appendix CII.	DNA sequence of the rat α_1-AGP gene.	167
Appendix CIII.	The complete nucleotide sequences of the rat α_1-AGP gene with comprehensive restriction sites.	169

CHAPTER 1: GENERAL INTRODUCTION

1.1 Introduction

The biosynthesis of eukaryotic proteins is an extremely sophisticated process involving multiple steps. Each of these steps can be seen as a regulatory point at a particular phenotype for expression (Brown, 1981). Some of the regulatory mechanisms which operate at the transcriptional, post-transcriptional and translational stages of protein biosynthesis have been identified and reviewed (Darnell, 1982; Brown, 1981). In general, the synthesis of mRNA is mediated in part by promotor-specific transcription-activating factors, which are DNA-binding proteins capable of discriminating between distinctive DNA sequence elements found in the promotor regions of different genes. The binding of these factors to DNA activate other components of the transcriptional machinery, including the RNA polymerase, to initiate transcription with accurate selectivity (Dynan et al, 1985). Most promotors have a concensus sequence 'TATAAA' called Goldberg-Hogness or TATA box, located 25-30 base pairs (bp) upstream from the transcriptional start (Breathnach et al, 1981). Further upstream from the transcriptional start sites may contain one or more additional promotor elements such as 'CCAAT' or 'GGGCGG'. In short, upstream promotor elements are the major region for the activation of promotor-specific transcription factors. Besides promotors, there are other regulatory elements termed enhancers which modulate the activity of many promotors. The enhancer is a specific DNA sequence located either upstream or downstream from the transcription start site at variable distances of up to 1000 bp from the promotor. It is becoming clear that promotor-specific transcription factors interact with promotor sequences and enable eukaryotic RNA polymerase II (which lacks any inherent ability of recognizing the promotor) to start mRNA synthesis.

Phenotypic expression in eukaryotes is regulated not only at the level of selective transcription, but also during the processing, splicing, and specific release of nuclear RNA (nRNA) species into cytoplasm for translation (Lewin, 1981; Brown, 1981). Post-trancriptional events, including 5'-end capping, intron splicing, 3'-end polyadenylation, and the transportation of nRNA cross the nuclear envelope are particularly difficult to study because of technical obstacles and lack of information about nuclear organization.

The nucleus is the largest intracellular organelle, housing the genomic program and controls for the issuing messages for cellular activity. The boundary between the nuclear envelope and the cytoplasm is one of the most strategic sites for intercepting and interpreting vital messages controlling eukaryotic phenotypes. Beyond the periphery of the nucleus, most mRNA's will not return and must enter the phase of kinetic competition for translation (Walden et al, 1981).

Investigators have speculated that the nuclear envelope plays an important role in regulating the release of nRNA's. This function can be altered by biological toxins, hormones, and chemical agents. In some cases, disturbances may lead to neoplastic transformation.

The microscopic detection of aberrant changes around the nucleus of tumor cells is a major task of the pathologist. The basic criteria of neoplastic cytology are "pleomorphism" (variation in the size and shape of cells and nuclei), indented and folded nuclear borders, nuclear hyperchromasia (dense; basophilic; coarse chromatin pattern), variation in the thickness of the nuclear membrane (irregular chromatin clumping around nuclear margins), nucleolar changes (multiplicity, enlargement, irregularity), abnormal chromosome figures, and an increased nuclear/cytoplasmic ratio (Smuckler et al., 1982).

Changes in nuclear transport functions have been induced and observed in experiments on chemical carcinogenesis. In primary hepatoma of the rat induced by Azo dye, nRNA transportation is altered (Shearer, 1972). The releasing fidelity is decreased, ill-processed nRNA's are found in the cytoplasm, and some released RNA's are identical to those found in regenerating adult rat liver after 30% hepatectomy or in intact livers of newborn rats (Shearer, 1972). In rats treated with thioacetamide, a hepatocarcinogen, energy-dependent nRNA transport is reduced and energy-independent diffusing nRNA release is increased. Ill-processed nRNA is again found in the cytoplasm (Smuckler et al., 1974). Since these alterations exist before and persist after tumor formation, they might be related to the mechanism of carcinogenesis.

Current approaches to the study of nRNA processing and release include the intranuclear injection of RNA into living cells (Stacy, 1981; Wicken, 1980; DeRoberts, 1982), the infection of cells with viruses of known DNA sequences (Subraman, 1983), and the use of prelabeled nRNA to study release from isolated nuclei in vitro (Blobel, 1966). Each method has its strong point, but lacks certain advantages of the other methods.

The method employed in our laboratory and by many other investigators is the in vitro release of nRNA from isolated rat liver nuclei. This method is technically convenient; and by administering specific metabolic substrates, biochemical analogs, chemical carcinogens, industrial toxins, etc., we can study the complex interrelationships among designated agents and the resultant effects on nuclei. The main criticism of this approach is that there might be built-in artifacts caused by the mechanical and chemical denaturation of nuclei during the isolation procedure (Clawson, 1980; Agutter, 1983).

Clawson and Smuckler (1980) have reported that the fidelity of nRNA release might be altered during nuclear incubation in vitro. Agutter et al. (1983)

have questioned the validity of using an in vitro assay of release because the exact alterations have not been completely verified. The in vitro assay of release is a unique and irreplaceable system for studying chemical carcinogensis, it is essential to determine its specificity and fidelity in comparison to events in vivo.

In order to determine the specificity and fidelity of nRNA release both in vitro and in vivo, a model system utilizing a highly inducible, measurable and well-characterized gene with determined DNA sequence must be utilized. So far, such a system is lacking. During the search among suitable gene systems, particular attention was paid to certain serum proteins due to their convenient assay and potential inducibility.

1.2 Plasma Proteins

The plasma proteins are the major extracellular components of the circulatory system (reviewed by Putham, 1975). Their functions are diverse including maintenance of colloid osmotic pressure, pH and electrolyte balance; transport of metal ions, fatty acids, steroid, hormones, drugs, etc.; availability as a nutritional source for tissues; hemostasis and the prevention of thrombosis; and defense against microbial invasion. During a disease state, the plasma proteins fluctuate in respect to their major components, such as albumin and the immunoglobins, and minor components, such as many of the α-globulins as well as transitory metabolic enzymes. It is the interest in plasma protein changes during disease states which has led to extensive studies and innumerable discoveries of the properties and functions of plasma proteins.

Plasma proteins are usually classified in different ways according to the methods of detection. Conventionally, plasma proteins are classed into the five groups of albumin (59%), α_1- (5%), α_2- (8%), β- (12%) and lambda-globulin (16%). The recent techniques of starch gel electrophoresis, immunoelectrophoresis,

chromatography and gel filtration have resulted in better resolution and identification of the plasma proteins. This increased resolution revealed many subgroups which were previously classified inappropriately. For example, the α_1-glycoprotein, α_1-lipoprotein and the heptoglobins, which were grouped under the α-globulins, actually bear nor familiar relationship with one another. Similarly, immunoglobulins were resolved into three major classes of IgG, IgA, and IgM.

Historically, the plasma proteins were first named and defined by their solubility in water and salt solutions, then grouped into five (albumin, α_1-, α_2-, β-, lambda-globulin) or six major classes according to their electrophoretic mobility at a restricted pH. Subgroupings were also defined in terms of solubility in cold ethanol immunoelectrophoresis, starch gel electrophoresis, and other specific isolation procedures.

None of the described procedures have been useful in identifying functions and systematic attempts to classify all the plasma proteins have failed. The difficulties of systematic nomenclature are compounded if more than one functional name is given to a single protein. For example, there are 13 additional designations given for α_1-acid glycoprotein, of which orosomucoid is the most commonly used. The term, α_1-acid glycoprotein, will be used throughout this dissertation despite the possible confusion with other α-globulins.

α-Globulins are proteins with very diverse functions, physical properties and chemical structures. α-Globulins share many common properties which enhance the difficulty in their separation and result in problems in nomenclature. They are mostly glycoproteins, usually having high carbohydrate content—as high as 45% for α_1-AGP. However, very little is known about their function. Most α-globulins are generally of low molecular weight, ranging from 21,000 to 60,000 except α_2-macroglobulin (725,000) and haptoglobin (100,000).

In 1965, none of the interior amino acid sequences of plasma proteins of any species were available although some terminal sequences of amino and carboxyl ends of several proteins were determined. Then, within one decade, tremendous progress in the sequence analysis of human and mouse immunoglobins was achieved. Substantial progress has also extended to many other human plasma proteins including haptoglobin α chain, serum albumin, pre-albumin, apolipoproteins, and α_1-acid glycoproteins. Although functions of most plasma proteins are unknown, they are highly conserved among species. They must play crucial roles in animal survival. Comparative structural study of plasma proteins of various species has revealed evolutionary branching, for the human and mouse immunoglobulins, for the albumin of four species, and for the rat and human prothrombins. Unfortunately, the primary structure of most plasma proteins of man or animal remain to be elucidated.

One of the main problems of plasma protein research is the identification of function. α_1-AGP is a protein that has been intensively studies and the entire amino acid sequence of human and rat α_1-AGP have been determined both by method of protein chemistry (Schmid, 1973) and modern recombinant DNA technology (Ricca, 1981). Even its gene structure in man (Luciana, 1985, unpublished) and rat (Liao, 1985) has recently been determined, and it is one of the most thoroughly studied plasma protein. Yet its function is still unknown.

1.3 Acute phase reactions

Originally the term "acute phase response" was given to the pattern of alteration in concentration of a number of plasma proteins (Table I) which resulted from a variety of different types of infection, inflammation, or tissue damage (Pepys, 1983).

Table I

SERUM PROTEIN LEVELS DURING THE ACUTE PHASE RESPONSE

	Elevated	Decreased
Proteinase inhibitors	α_1-Antitrypsin	Inter α-antitrypsin
	α_1-antichymotrypsin	
Coagulation proteins	Fibrinogen	
	Prothrombin	
	Factor VIII	
	Plasminogen	
Complement proteins	C1s	Properdin (P)
	C2, B	
	C3, C4, C5	
	C9	
	C56	
	CIINH	
Transport proteins	Haptoglobin	
	Hemopexin	
	Ceruloplasmin	
	Ferritin	
Lipoproteins		High-density lipoprotein
		Low-density lipoprotein
Others	C-reactive protein	Prealbumin
	Serum amyloid A	
	α_1-AGP	
	Fibronectin	
	Gc globulin	

In 1941, Avery and his colleagues introduced the term "acute phase" to refer to serum containing C-reactive protein (CRP) collected from patients who were acutely ill with infectious disease (Abernathy and Avery, 1941; MacLeod and Avery, 1941). CRP was hence known as "acute phase protein". When more specific assays for individual plasma proteins were developed, it was realized that the concentration of a number of other plasma proteins were also raised in acute phase sera and they are now collectively designated as acute phase proteins.

The elevated acute phase proteins are an important part of a systemic response to injuries. This systemic response includes metabolic alterations such as fever; leukocytosis; increased protein degradation and gluconeogenesis; negative nitrogen balance despite increased total protein synthesis; increased biosynthesis of glucagon, insulin, ACTH, cortisol, catecholamines, growth hormones, TSH, T4, aldosterone and vasopressin; increased serum copper concentration, and decreased serum zinc and iron concentration. Many of these changes are caused by a variety of regulatory mediators which are transiently released after injury (Sipe and Rosetreich, 1981; Kushner, 1982; Bornstein, 1982; Kampschmidt et al., 1982). Historically, the term "acute phase response" was primarily applied to the acute phase plasma protein changes, however, the meaning has been extended to a general description of overall systemic reactions which cover cellular, neurological, biochemical, endocrine and other metabolic changes.

The diverse stimuli which initiate the acute phase response include chemical or physical trauma; chemical, toxic or allergic inflammation; bacteriol, viral, fungal or parasitic infection; ischemic necrosis; and malignant neoplasia. The

common denominator is cellular injury or death. Other potent agents such as abrin and Gram-negative bacterial endotoxin can also provoke major acute phase responses without causing tissue damage. This may be due to their direct activation or to stimulation to macrophages. The study of macrophage products as important signals to initiate acute phase response has led to the conclusion that any substance which is capable of triggering the release of stimulating factors from macrophage is able to cause acute phase response, even when no real tissue injury has materialized. Recently, it was demonstrated that a peptide product of mononuclear phagocytic cells designated as interleukin-1 (IL-1) is involved in triggering the synthesis of several acute phase proteins and fever.

The acute phase responses which are represented by altered plasma protein levels and various metabolic, endocrinological and physiological changes following tissue injury, infection, or inflammation are observed in all homiothermic animals. The elevated levels of coagulation proteins, complement proteins, transport proteins, and proteinase inhibitors are obviously promoting resolution and repair which are beneficial to the recovery of the animal. Nevertheless, the functions of many other proteins are still unclear.

In most cases, the increased circulating levels of acute phase proteins during the acute phase response are mainly due to increased _de novo_ biosynthesis. The majority of the acute phase proteins are produced by hepatocytes. Other cells such as fibroblast cells, lymphocytes, monocytes, and polymorphs also synthesize several acute phase proteins. In experimental studies, the increased hepatocyte synthesis of acute phase proteins is a result of increased number of hepatocytes participating in the active synthesis. The gradually expanding involvement of hepatocytes from periportal to central region of liver engaging in active production of acute phase proteins may be interpreted as an increased diffusing gradient of stimulating mediators during clinical course. The molecular

mechanism of increased hepatocyte synthesis of acute phase proteins seem to involve by the increased abundance of specific species of mRNA. In mouse serum, amyloid A protein (SAA) mRNA increases 500- to 1000-fold (Morrow et al., 1981). For rat, the α_1-acid glycoprotein mRNA increases 90-fold (Ricca et al., 1982) and for baboon α_1-antitrypsin mRNA, 80-fold (Chandra et al., 1981). The elevated mRNA accumulations of acute phase proteins may be due to increased transcription, increased mRNA stability, and increased nucleocytoplasmic transportation of specific mRNA species. At the peak of the acute phase response, the rat α_1-AGP mRNA accounts for 2.7% of the total liver mRNA (Ricca et al., 1982), while mouse SAA protein synthesis constitutes 2.5% of total hepatic protein biosynthesis (Morrow et al., 1981).

1.4 The α_1-acid glycoprotein

α_1-Acid glycoprotein (α_1-AGP) is a globulin in normal animal plasma. Its unusual properties such as very high carbohydrate content, large numbers of sialyl residues, peripheral macroheterogeneity, very acidic isoelectric point, high aqueous solubility, numerous amino acid substitutions, significant degree of homology with immunoglobulins, and drastic elevation during acute phase response distinguish it from other serum proteins. It is noted that the carbohydrate moiety of α_1-AGP constitutes approximately 10% of the protein-bound carbohydrate of normal plasma.

The very acidic isoionic point and unusual solubility of α_1-AGP, even in the presence of trichloroaceteic acid, make the preparation of α_1-AGP a relative easy task. Procedures for isolation and purification of α_1-AGP are mainly electrophoresis and ion-exchange chromatography. Solubility procedures are also commonly used. Based on the procedure developed by Winzler and colleagues (Weimer et al., 1950) which involved precipitation of most other protein at acidic

PH values and increasing ammonium sulfate concentration, α_1-AGP has been isolated in the homogeneous state from different animal species including chimpanzee, rabbit, rat, ox, sheep and human. At present, due to the large scale of preparation of the human immunoglobulins and albumin by low temperature-low salt-ethanol procedure of Cohn (1946), α_1-AGP became an important side product. The yield of this protein is about 450 mg per liter of plasma and commercial preparations are inexpensive.

α_1-AGP is a single polypeptide with a molecular weight of 23,000. It also has 5 carbohydrate side chains which account for about 45-55% of the total molecular mass of 45,000. This globulin is found to contain no cholesterol, phospholipids, fatty acids and hexuronic acid. The isoionic point of α_1-AGP is found to be at pH 3.53, due to its large number of sialyl residues, and is the lowest of the human plasma proteins (Schmid et al., 1962).

The amino acid sequence of human α_1-AGP (which consists of 181 residues) was completely elucidated by techniques of CNBr[3] and subsequent tryptic cleavage and chymotryptic digestion of desialyzed protein (Schmid et al., 1971, 1973a; Ikenaka et al., 1971, 1972; Kaufmann and Schmid, 1962). It appears that certain amino acid substitutions were present in the human population. In the human, the 5 carbohydrate units of α_1-AGP (Fig. 2) were found to be linked to the asparagine residues 15,38,54,75, and 85. In rat, the N-glycosidical linkages are at asparagine residues 7, 16, 58, 76, 86 and 116. In the human the two disulfide bonds found in α_1-AGP were determined to be at residue 5 to residue 147 and residue 72 to 164 respectively by conventional techniques (Schmid et al., 1974). Although the definite second structure of α_1-AGP has not yet been completely determined, preliminary assessment of the conformations from optical rotatory dispersion and circular dichroism measurements, has shown it has 70% β-structure and negligible α-helical structure.

On starch gel electrophoresis near the isoelectric point of pH 5, desialyzed human α_1-AGP shows two separated bands. Since all glucosamine residues are N-acetylated, the desialyzed carbohydrate moiety contains no electrostatic charges and any differences in the electrophoretic mobilities must be due to differences in the electrostatic net change of the polypeptide chain. Hence, it is suspected that a variant form of α_1-AGP is present in each individual.

There are three types of patterns observed in the human population: SS, FF and FS types. The SS type individual possesses a major component of α_1-AGP of slow electrostatic mobility and a minor component of fast electrostatic mobility. The FF type individual contains major α_1-AGP of fast electrostatic mobility and minor amount of the slow electrostatic mobility species. The FS type individual has an equal amount of α_1-AGP of both slow and fast electrostatic mobility (Schmid, 1965). The presence of these three phenotypes suggest genetic transmission of two different genotypes. Further studies using desialyzed α_1-AGP isolated from a series of twins and normal individual who were in acute phase reaction demonstrated that the patterns remained constant and established the theory of genetic transmission (Tokita et al., 1966; Yoshizaki et al., 1969). A pedigree study of α_1-AGP variant types demonstrated that they are autosomal traits with codominant expression. The gene frequencies of α_1-AGP variant types in different ethnic groups were also determined in the same study (Johnson et al., 1969). At least one experiment using proteolytic digestion under well-controlled conditions revealed that difference in at least one peptide pair. Sequence analysis of this pair demonstrated an amino acid substitution of arginine for glutamine at residue 20 (Nimberg et al, 1971). Sequence analysis of pooled human plasma indicates at least 22 amino acid substitutions at other undetermined residue locations. Silent variants, i.e., substitution of a neutral amino acid by another neutral amino asid, also takes place in α_1-AGP. Such amino acid

substitutions will not change the electrostatic net charge, therefore, resulting in no difference in electrostatic mobility.

Elevated blood level of α_1-AGP has been known for decades to be associated with various unrelated disease or physiological states such as cancer, pneumonia, and rheumatoid arthritis (Winzler, 1955), pregnancy (Adams and Wacher, 1968), and induced inflammation (Jamieson et al., 1972a,b). After a long period of observation, it was concluded that any acute response caused by physical injury (major operations); bacterial, viral, fungal, or protozoan infections; lupus erythematosus; rheumatoid arthritis; or neoplasia, etc., resulted in prompt increase of the serum levels of this protein. Experimental subcutaneous injection of turpentine to produce sterile abscesses in the rat results in a 90- to 100-fold increase of α_1-AGP mRNA in hepatic cytoplasm, and its translational product becomes a major serum glycoprotein (Ricca et al., 1981). Exposure of rat hepatoma cells in culture to glucocorticoid hormones results in a similar induction of α_1-AGP (Vannice et al., 1983; Baumann et al., 1983) and reaches a 500-fold mRNA in hepatoma cytoplasm (unpublished observation). Although its precise biological function is unknown, the protein is highly conserved and maintained, even during severe starvation. Synthesis mainly occurs in the liver. Additional observations pertaining to the possible biological function of α_1-AGP have been reported as influencing the blood clotting mechanism primarily at the prothrombin activation level (Yamashina, 1976; Das, 1962). _In vitro_ studies indicate that α_1-AGP inhibits platelet aggregation (Snyder et al., 1976), lymphocyte transformation (Chill et al., 1977), cell-collagen interaction (Franzblau et al., 1976), and the invasion of malaria protozoa in erythrocytes (Friedman, 1983). The resemblance of α_1-AGP to erythrocyte membrane sialoglycoproteins, the homology of its amino acid sequence in the carboxyl terminal region with that in IgG chain constant region (Schmid et al., 1973), and its induction by

glucocorticoid hormones suggest that is might be involved with the immune system and might act both as a protective decoy during parasitism and as an immunosuppressant during acute inflammation in response to tissue damage.

During the translational processing of α_1-AGP in the hapatocyte cytosol, the carbohydrate moiety is established by transferring an N-acetylglucosaminyl residue to an asparaginyl residue during elongation of polypeptide chain, even if this chain is still on the ribosome. Additional monosaccharide residues are further incorporated when the nascent glycoproteins are passing the rough and smooth endoplasmic reticulum. Finally, the heteropolysaccharide unit additions are completed in the Golgi complex by transferring of sialyl residue (Jamieson and Ashton, 1973). In the step of heteropolyssacharide addition, variable pathways or "errors" occur and lead to branched heteropolysaccharide chains which may also be responsible for observable microheterogeneity or polymorphism in pooled human plasma. The secreted α_1-AGP has a half-life of 5.5 days (Winzler, 1965; Simkin and Jamieson, 1968) and its integrity is constantly monitored by the parenchymal cells of the liver. The terminal galactose appears to play the major role in hepatic recognition. Once desialysed, α_1-AGP has a half-life of only 2 minutes (Morell et al., 1971), apparently is recognized and taken up by the hepatocyte through sialylprotein receptors and subsequently internalized for destruction.

1.5 Working hypothesis

As described previously, the study of nRNA splicing, maturation and nucleocytoplasmic transportation require a well-characterized model system of a particular animal gene with known DNA sequenses. This gene should be highly inducible and measurable within a short period of time. It also should be expressed in a manipulatable laboratory animal host (preferably highly inbred to

avoid polymorphism). The gene product must be easily accessible for continuous monitoring without sacrificing the animal (ideally a serum protein). This gene shall be representative and simple, that is containing multiple but not too many introns for interpreting the blotting results (albumin α_1-fetoprotein and α_1-macroglobulin are too large and complex). Detailed studies show none of the currently available genes meet these requirements. However, the rat α_1-AGP gene is an excellent candidate, but the gene had not been isolated and DNA sequence had not been determined. The rat α_1-AGP gene was found to be specifically induced by glucocorticoids and the result can be repeated in an isolated, well-controlled in vitro system—the rat hepatoma culture. The amino acid sequences of human α_1-AGP has been completely determined. The amino acid sequences of the rat α_1-AGP was deduced from its cDNA and antibody against the rat α_1-AGP was also available (Ricca et al., 1981). The total length of the rat α_1-AGP cDNA is less than 800 bp, a speculative estimate of its genomic fragment would be 4000 to 8000 bp—making cloning and sequencing a relatively difficult but potentially attainable task.

The experimental scheme consisted of isolating the rat α_1-AGP gene, establishing the genomic DNA sequence, and subcloning various intron and exon fragments as probes for northern hybridization analysis. In northern blotting analysis, I am attempting RNA electrophoresis with a 0.1 mm to 0.2 mm ultra thin acrylamide gel. Isolated cytoplasmic RNA's and RNA's released in vitro will be denatured and run with known ssDNA fragments of M-13 bacteriophage, generated by the chain termination method. Since this gel will offer 1-nucleotide-difference resolution, ill-processed and released nRNA's will be quickly determined, and unspliced introns might be closely estimated because of their sizes.

The next step will be the observation of post-transcriptional events including determination of whether intron splicing is a random event or an orderly

sequence of events. The sites of nRNA processing of α_1-AGP will be detected by the use of specific exon and intron probes, either by direct in situ hybridization, or indirectly by in vivo hybridization. Both of the proposed techniques should permit microscopic observation of the post-transcriptional events when combined with autoradiography, immunofluorescence, and heavy metal labeling. cDNA's of exons, introns, or fragments overlapping both regions will be applied directly to microsectioned tissue, or predelivered within liposomes to intact tissue in vivo.

The last stage will be to test whether cDNA and/or mRNA delivered by liposomes will cause designated biological activities in vivo. cDNA of α_1-AGP, when delivered by liposomes into hepatocyte cytoplasm, may hybridize with the mRNA counterpart to form a double-stranded structure and terminate protein synthesis. The results can be easily monitored by measuring the α_1-AGP level in serum. mRNA of liver oncogenes will be transcribed on oncogene cDNA with newly available riboprobe kits and delivered to hepatocytes. By using this method, we can bypass the long waiting period of cancer induction by carcinogens and observe tumor formation without altering chromosomes or turning on cellular proto-oncogenes.

We can ask further questions: e.g., Will tumorogenesis induced by this method be abolished when mRNA of certain oncogenes is withdrawn? Can cDNA of certain oncogenes hybridize with mRNA to stop the promotion and initiation of tumorogenesis? This approach might be useful both in research, prevention and treatment of diseases. The post-transcriptional manipulation, if successful, will enable us to pursue a wide-range of testing in both normal and abnormal biological systems and to establish the relation between one gene product and its specific effect(s) without altering the host genome.

Within the large scope of this experimental scheme my project involved isolating, characterizing and sequencing the rat α_1-AGP gene at this stage. The

effort of identifying its regulatory elements and the comparison of its sequences to two functionally unrelated genes with homologous sequences was also pursued.

The significant induction of α_1-AGP makes its promotor region a valuble tool of improving certain low-yielded biotechnology products. In order to assay the promotor's effect, the promotor will be reconstructed with suitable foreign genes to assay its effects.

The unusual conservation and homology of α_1-AGP with immunoglobin, EGF receptor has led to the pursuit of its evolutional origin. In this thesis, its amino acid and nucleotide sequences in relationship to the human α_1-AGP cDNA and gene are presented in order to make guesses as to its functions.

CHAPTER 2: ISOLATION AND CHARACTERIZATION OF THE RAT α_1-AGP GENE

2.1 Introduction

The usage of bacteriophage lambda gene libraries to isolate genomic clones harboring DNA fragments homologous to cDNA sequences has made it possible to study eukaryotic genes. Elucidation of the structure and organization of many eukaryotic genes reveals that a large number of protein-coding genes are interrupted by intervening sequences or introns. A close inspection of the structure of eukaryotic genes leads to the development of theories on their evolution, expression, mechanism of splicing and other post-transcriptional events.

In order to establish a model system for studying gene expression it is desirable to characterize and sequence a highly inducible gene. The rat α_1-AGP as described previously possesses advantages for such study.

In order to isolate and determine the organization of the rat α_1-AGP gene, its genomic clones were isolated and sequenced. The results provide us information on its molecular structure, the possible evolutional relationship with other genes and insights on diverging movement of this gene among different species. The isolated gene also gives us opportunities to study mechanisms of expression through gene reconstruction and transfection experiments.

The work mentioned in this chapter discusses the isolation and characterization of a genomic clone (lambda rAGP-40) which contains the entire rat α_1-AGP gene.

2.2 Methods and materials

2.2.1 Preparation of genomic DNA

High molecular weight liver DNA was extracted from an adult male Sprague-Dawley rat (Simonson Laboratory, Gilroy, CA) essentially as described by Blin and Stafford (1976). Five to ten grams of rat liver was cut into small pieces by scissors in NKM buffer (0.15 M NaCl; 5 mM KCl; 2 mM $MgCl_2$) and then homogenized by 10 to 12 strokes in a tissue homogenizer. The cells were washed in NKM buffer and then lysed by adding 2 mM $MgCl_2$ solution (Sargent et al., 1977).

2.2.2 Genomic library construction

A bacteriophage lambda genomic library of Sprague Dawley rat was constructed by ligating a partial Eco RI digest of high molecular weight DNA (15-20 kb) into Eco RI-cleaved Charon 4A DNA and then packaging (Seargent et al, 1979). The partial Eco RI digest was carried out by aliquoting the prepared rat liver DNA into different batch tubes and digesting with Eco RI restriction enzyme (Boehringer-Mannheim) to different degrees of cleavage. The degree of cleavage was evaluated by comparing one-third and one-fifth parallel digestions of Eco RI sites of equivalent amount of bacteriophage lambda DNA under identical conditions. The partially digested DNA fragments were centrifuged and sedimented through a 10-30% sucrose gradient and fractioned. The DNA with size between 10 and 20 kb was recovered by ethanol precipitation then resuspended in TE solution. The final concentration and total yield were measured by UV (260 nm) spectrometry. A fraction (2.5 μg) of this recovered rat liver DNA was ligated with 8.5 μg of lambda Charon 4A cloning vector (Sternberg et al., 1977; Blatter et al., 1977). The ligated DNA was packaged in vitro by using extract

from defective lambda lysogens supplied by N. Sternburg (Sargent et al., 1977) and prepared as described by Hohn and Murray (8 of PNAS 1979). As a result, a genomic library was constituted with a complexity of 2,000,000 independent clones.

2.2.3 Selection and isolation of genomic clones

Bacteriophages from this genomic library were titered and plated out at density of 20 thousand bacteriophage plaques per plate. 500,000 bacteriophage lambda clones (equivalent to 10 genomes) were screened by in situ hybridization using nitrocellulose paper (Schleicher and Schuell, Inc., Keene, NH) as transferring carrier as essentially modified from methods of Maniatis et al. (1982). In this simplified method 20 µl of bacteriophage lambda equivalent to 20,000 bacteriophage particles was mixed in a 2057 Falcon polystyrene tube (Becton Dickinson Labware, Oxnard, CA) with 0.3 ml saturated bacterial broth (E. coli strain DP50) and 0.3 ml of 100 mM $MgCl_2$, $CaCl_2$ then incubated room temperature for 10 minutes. Eleven ml of molten NZYDT agarose (Appendix A) (first molten in microwave oven then cooled in 55°C water bath for at least 30 minutes) was mixed with the bacteriophages and DP50 E. coli hosts before pouring into a 150-mm petri dish containing NZYDT top agar (Appendix A). It is essential to dry the plate in a 37°C incubator for at least 3 hours with open lid to ensure a dry and firmly attachable surface for molten agarose before pouring top agarose. Plates were incubated in 37°C for overnight or 10 to 11 hours and the bacteriophage plaques were allowed to reach the size of about 1 mm diameter. The plates were chilled at 4°C for 1 hour to harden the surface. Nitrocellulose filters were marked by water resistant ink and placed evenly on the surface of the top agarose for 1 minute. Four asymmetric holes forming a "L" figure were punctured through nitrocellulose filters to the bottom of agarose to mark

alignment points for precise matching. The nicrocellulose filters were then peeled off from agarose plate with a blunt end forceps and placed on the surface of 3M paper (Whatman) soaked with a denaturing solution (1.5 M NaCl, 0.5 m NaOH) for 1 minute. The filters with liberated and attached bacteriophage DNA were transferred to another 3M paper saturated with neutralizing solution (1.5 M NaCl, 0.5 M Tris, Cl [pH 8.0]) for at least 5 minutes then rinsed in 2 X SSPE (0.3 M NaCl, 20 nM Na H_2PO_4 [pH 7.4], 2 mM EDTA [pH 7.4]) in order to wash off excess salt. After being rinsed, filters were air-dried and wrapped between sheets of 3M paper of similar size and baked in a vacuum oven at 80°C for 2 hours to fix DNA permanently on the nitrocellulose. These filters were stored in room temperature after baking.

The 792 bp <u>Pst</u>-1 <u>Pst</u>-1 DNA fragment derived from rat α_1-AGP cDNA plasmid pAGP663 (Ricca et al., 1981) was eluted from agarose gel electrophoresis and was used as a hybridizing probe after labeling with [α-^{32}P] by the nick-translation or random priming method. The nick translation kits were purchased from Amersham Corp., Arlington Heights, IL, and used according to manufacturer's specifications (Appendix B). The random priming method was done by partially cleaving 1 μg pAGP663 with <u>Hae</u> III restriction enzyme for 5 minutes to become relaxed or linearized DNA, then boiled for 3 minutes in the presence of 12mer oligonucleotides with random sequences (Biosearch Inc., San Rafael, CA). This procedure shall destroy hydrogen bindings and form 12mer-plasmid hybrid during subsequent freezing in dry ice and gradual warming in ice and 55°C waterbath. The labeling was done by mixing randomly primed oligomer-plasmid complex with [α-^{32}P] dCTP, [α-^{32}P] dATP as radioactive substrates and dGTP, dTTP as cold substrates. The reaction lasted 20 minutes in a volume of 50 μl with buffer constituents identical to <u>Sac</u> I restriction digestion (New England Biolab.).

Hybridization reaction was carried out at 68°C in a straight hybridization solution instead of 50% formamide in order to save time. Before hybridization the baked nitrocellulose filters were gently rinsed in 6xSSC (Appendix A) until they became thoroughly wet then they were transferred to a circular plastic box for prewashing to remove absorbed media, agarose and bacterial debris. The prewashing solution consists of 50 mM Tris • Cl (pH 8.0), 1.0 M NaCl, 1 mM EDTA, 0.1% SDS. After 2 hours prewashing at 50°C, the solution was discarded and replaced with 300 ml of prehybridization solution (5x Denhart's solution [Appendix A], 5xSSPE, 0.1% SDS, 100 μg/ml denatured [by boiling for 5 minutes] salmon sperm DNA and incubated at 68°C for 5 hours). ^{32}P-Labeled DNA probe was denatured by heating in boiling water for 5 minutes and then added in the prehybridization solution. Hybridization was continued in a tightly sealed circular box with constant shaking at 68°C for 48 hours. After the hybridization was completed, the hybridization solution was saved and stored in -20°C, the filters were washed 5 times, for 10 minutes each in 300 ml of 2xSSC and 0.1% SDS at room temperature, then 3 times at 60°C for 5 minutes in each wash. The radioactivity of filters were closely monitored by Geiger counter. The filters were dried in air at room temperature and taped on 3M paper marked with radioactive ink for precise alignment between the autoradiographs and hybridization filters. The 3M paper and filters were covered with Saran wrap and applied to X-ray film (Kodak XR) for 24 hours at -70°C with intensifing screen. Under these conditions, the background was usually high enough to give a vague image of whole filter but low enough to see distinct signals of positive clones on x-ray films, particularly convenient to locate the positive clones on agarose plates.

After screening, positive clones were identified, the one with strongest hybridization signal was selected and purified by low density plating and

transferred to nitrocellulose filters and hybridized with the stored solution containing pAGP663 probe as described previously until it became monoclonal.

2.2.4 Preparation of bacteriophage lambda DNA

The finally purified bacteriophage lambda clone plug was removed with a Pasteur pipet and stored in 1 ml SM buffer (Appendix A) and vortexed. After storing in 4°C overnight, most phage was diffused completely and 300 µl were mixed with 300 µl saturated DP50 E. coli broth and incubated in 37°C shaker for 15 min. The mixture was then diluted to 40 ml of NZYDT broth (Appendix A) and incubated at 37°C with shaking (200 rpm) for 10 hours. At the 10th hour, most bacteria lysed and became a translucent solution. The 40 ml lysate was transferred to a 40 ml-Oakridge tube and spun at 10,000 rpm for 10 minutes to precipitate bacterial debris. The supernatant was transferred to another tube and digested with 12 µl DNase (10 mg/ml) and 3 µl RNase (10 mg/ml) at 37°C for 45 minutes to hydrolyze bacterial DNA and RNA. 5.5 ml 20% PEG (Mwt. 8000) and 2.5 M NaCl were added to lysate, mixed well and incubated in 0°C for 15 minutes, then spun at 15,000 rpm for 20 min. The supernatant was discarded and drained out from the tube completely. The pelleted phage particles were resuspended in 500 ml proteinase K buffer (Appendix A), transferred to another Eppendorf tube and spun again in bench top fuge to precipitate additional bacterial debris. The supernatant was transferred to another Eppendorf tube and treated with 2 µl proteinase K (10 mg/ml) at 37°C for 45 minutes. Phage DNA liberated from protein coat was extracted twice with 20 µl phenol and once with 250 µl chloroform. Aqueous layer was removed to another Eppendorf tube and mixed with 2.5 volumes of 100% ethanol and 0.1 volume of 3 M sodium acetate. At this moment, bacteriophage DNA coagulated into a cottony ball. This ball-like phage DNA was lifted by using a metal needle and transferred to another tube containing 70% ethanol and spun for 3 minutes to precipitate the DNA

pellet. After removing 70% ethanol, the pellet was washed with another 500 µl 70% ETOH and spun down, then dried in a speed-vac evaporator (Savant Instruments, Inc. Farmingdale NY). Dried phage DNA was resuspended in 200 µl TE with heating at 55°C. The O.D. (optical density) of the mixture was determined by spectrometer. A simplified outline of preparing bacteriophage DNA is also listed in Appendix B.

2.2.5 Restriction mapping and Southern Blot analysis

0.5 µg of prepared bacteriophage lambda DNA were cleaved by single and double restriction digest of Eco RI, Sac I, Bam HI, and Sal I endonucleases. The digested DNA was then electrophoresed in 0.75% agarose gel and transferred to a nitrocellulose filter. The nibrocellulose paper was hybridzed to a ^{32}P-labeled 792 bp Pst I/Pst I fragment of pAGP663 plasmid and exposed for autoradiography as modified from Maniatis et al. (1982). In the modified procedure, the agarose gel was stained with ethidium bromide and photographed along a ruler after electrophoresis (Fig. 2.1). The stained DNA pattern was traced and copied on Saran plastic covering the gel for exact alignment and site identification (Fig. 2.2). The agarose gel was then left in UV light for 10 minutes to cause DNA breakage. The UV exposed agarose gel was transferred in a container of 1.5 M NaCl and 0.5 M NaOH for 40 minutes at room temperature with constant shaking in order to denature the DNA. The denatured gel was neutralized by soaking it in the solution of 1.0 M Tris • Cl (pH 8.0) and 1.5 M NaCl for 1 hours at room temperature with gentle shaking. The neutralized gel was inverted, placed against a damp 3M paper saturated with 10 x SSC by extending itself into reservoir of gel box. On the surface of the agarose gel a piece of wet nitrocellulose filter (wetted in 2XSSC for 5 minutes) of identical gel size was placed evenly without air bubble. Two sheets of 3M paper followed by a stack (10 cm high) of Teri

Fig. 2.1. An agarose gel electrophoresis of restriction endonuclease-digested DNAs of both rat AGP lambda 40 and rat liver. In each restriction digestion, 1 µg of lambda AGP lambda 40 and 20 µg of rat liver was used. The digested DNAs were electrophoresed in 0.75% agarose gel with 1 x TBE buffer. The voltage was 20 mA and run for overnight. The gel was then stained with ethedium bromide and photographed.

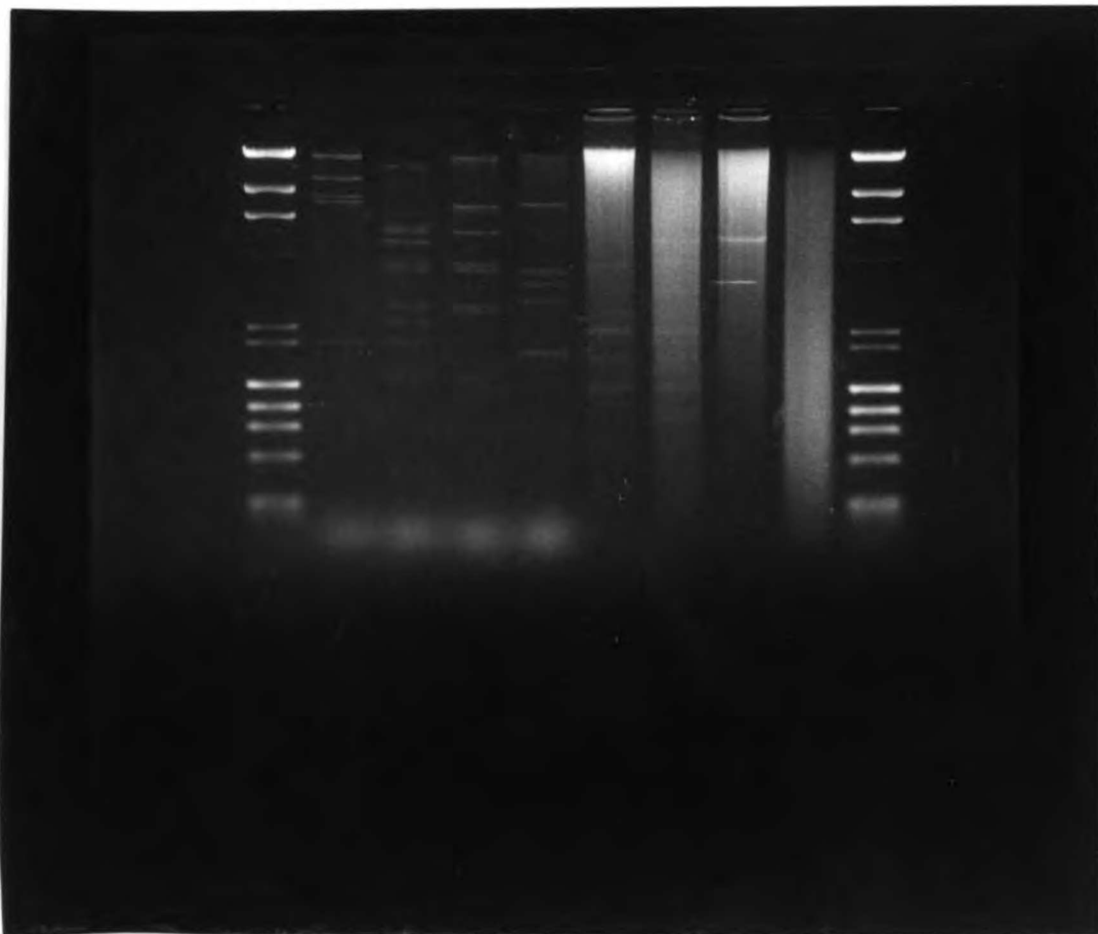
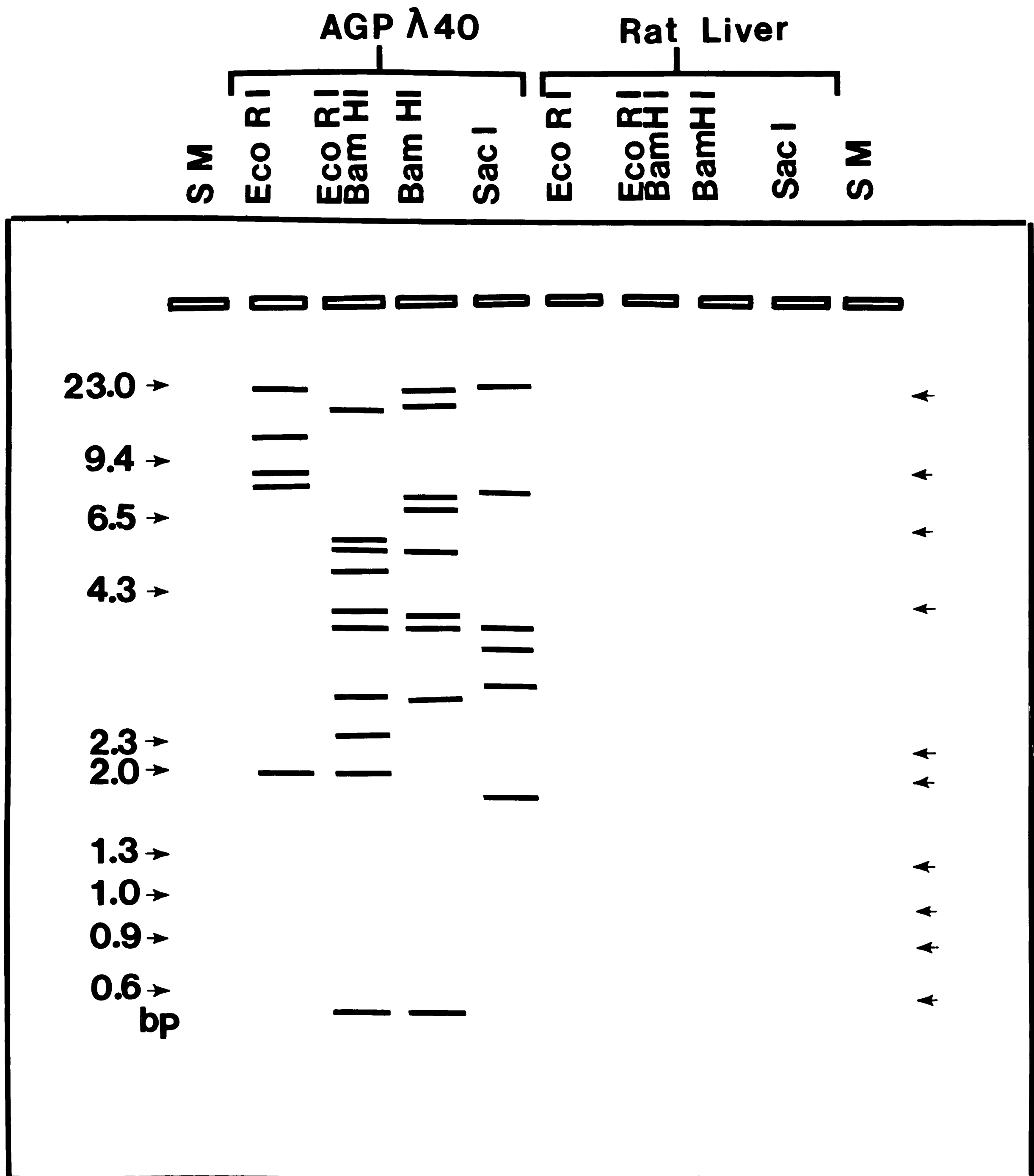

AGP λ 40
Rat Liver
SM
Eco RI
Eco RI
Bam HI
Bam HI
Sac I
Eco RI
Eco RI
Bam HI
Bam HI
Sac I
SM

Fig. 2.2. An agarose gel replica of Fig. 2.1. In order to precisely measure sizes of restriction fragments for making an accurate map of the rat α_1-AGP gene, a plastic replica was derived by directly tracing the electrophoresis patterns from the previous agarose gel (Fig. 2.1). The sizes of DNA fragments were latter determined by log phase plotting. The estimated sizes were listed on Fig. 2.4A.



AGP λ40
Rat Liver
S M
Eco R I
Eco R I
Bam HI
Bam HI
Sac I
Eco R I
Eco R I
BamH I
BamH I
Sac I
S M
23.0
9.4
6.5
4.3
2.3
2.0
1.3
1.0
0.9
0.6
bp

paper towel (Kimberly-Clark Corp, Roswell, GA) with similar size to nitrocellulose filter was again placed on top of the wet nitrocellulose paper and pressed down by an inflexible metal or glass plate and a 1 kg lead block. The transfer of DNA proceeded for 24 hours and the towels were replaced twice. Finally, the nitrocellulose filter was peeled off from the dehydrated gel, and soaked in 6XSSTC solution at room temperature for 5 minutes, then dired in air. The dried nitrocellulose papers were sandwiched between 2 pieces of 3M paper and baked for 2 hours at 80°C under vacuum. The baked filter was stored at room temperature. In order to confirm the copy number and absence of sequence rearrangements of the α_1-AGP gene during recombinant DNA manipulation, Southern hybridization using particular restriction enzymes on cellular DNA was also performed. Each 20 μg of total mammalian DNA was digested in the same manner as bacteriophage DNA and applied to a single gel slot for electrophoresis. The resulting autoradiograph of Southern hybridization on both cellular and bacteriophage recombinant DNA is shown in Fig. 2.3.

2.3 Results

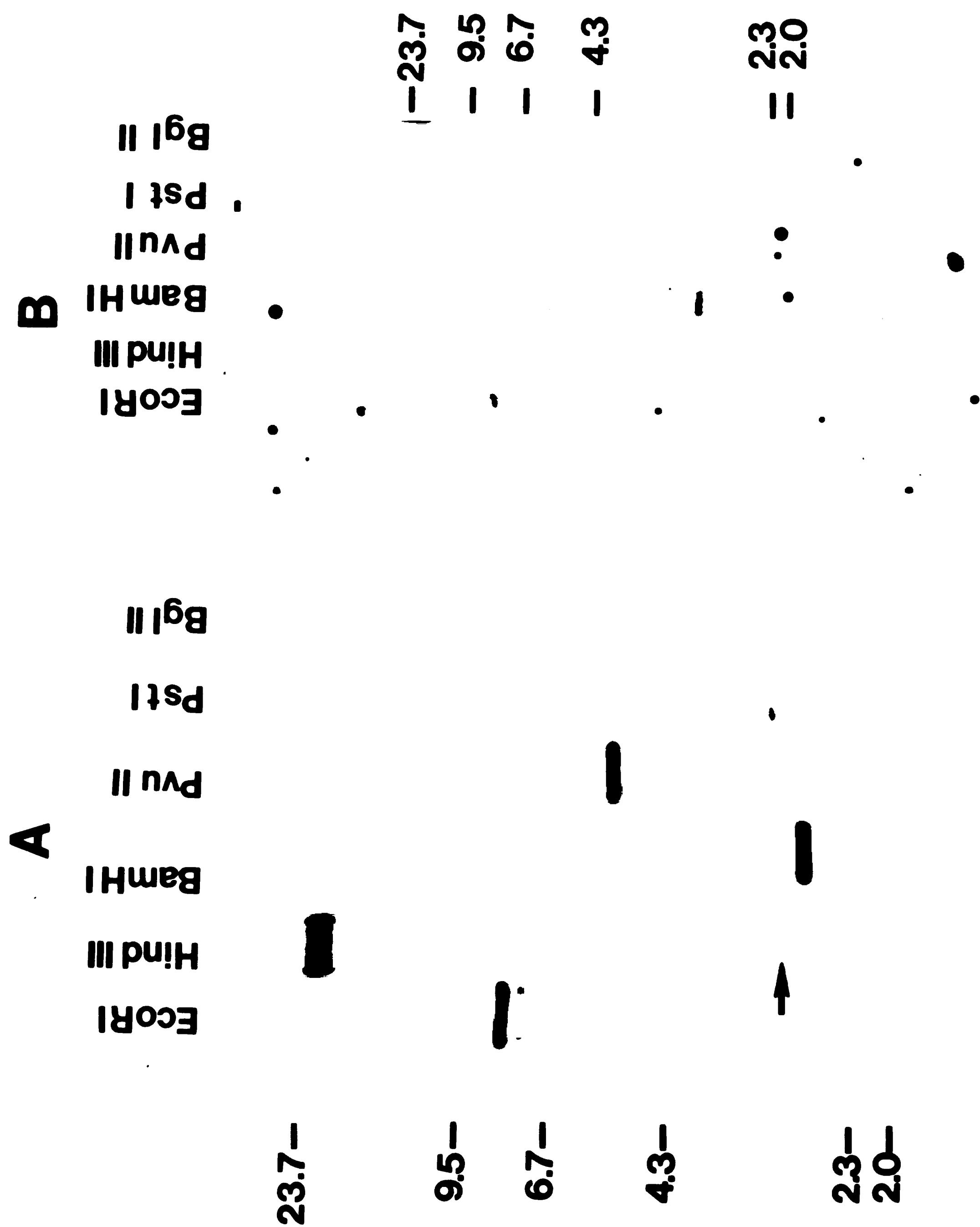
The results derived from these restriction digestions and Southern blotting analysis made it possible to construct a physical map of the genomic clone lambda AGP40 as shown in Fig. 2.2, and indicate that lambda AGP40 clone contains a 17.5 kb insert of rat genomic DNA with two internal and two flanking Eco RI restriction sites. This covers the complete coding regions of the α_1-AGP gene by matching the patterns of Southern blotting of rat and bacteriophage DNA. After analyzing the information derived from the sizes and patterns of single and double-restriction digest, it is clear that there are four Eco RI restriction sites which divide the total inserted DNA into three fragments arranged as 8.5 kb, 7.5 kb, 2 kb from 5' toward 3' end respectively. This mapping is

possible due to the existing map of lambda Charon 4A. Based on the intensities of hybridization signals on autoradiograph, the majority (80%) of α_1-AGP coding sequences are located within the 8.5 kb Eco RI fragment, 20% is in the 7.5 kb Eco RI fragment and none was found in the 2 kb Eco RI fragment (Fig. 2.3).

Fig. 2.3. Southern analysis of recombinant lambda (AGP lambda 40) and rat cellular DNA (rat hepatoma cell).

(A) Southern blotting of AGP lambda 40 DNA one μg of lambda DNA were used in each lane. The number on the left indicates the size marker with length unit of Kilobase-pair. The arrow indicates the existing band of Hind III fragment after extended exposure. The majroity of high molecular weight DNA in Hind III hybridization indicates incomplete digestion of Hind III endonuclease.

(B) Southern blotting of restriction endonuclease-digested hepatoma DNA. The experimental condition is the same as Fig. 2.3A, except that the agarose gel is 1% and the amount of DNA used is 20 μg.


B
EcoRI
Hind III
BamHI
Pvu II
Pst I
Bgl II
23.7
9.5
6.7
4.3
2.3
2.0
A
EcoRI
Hind III
BamHI
Pvu II
Pst I
Bgl II
23.7
9.5
6.7
4.3
2.3
2.0

2.4 Discussion

In conclusion, the lambda rAGP40 clone isolated and selected among other eight candidates from the rat genomic library contains a naturally intact and continuous DNA fragment covering the α_1-AGP gene(s). This is particularly convenient, since it would avoid additional work of screening new clones if the isolated clone was found to be incomplete. Furthermore, a majority of its protein coding sequences are located within the 8.5 kb Eco RI fragment. This information in combination to the known restriction sites of cDNA made it possible to draw an accurate physical map of this gene and to formulate an efficient sequencing strategy.

The restriction map of the rat α_1-AGP were determined based on various informations derived from restriction mapping of rAGP lambda 40, Southern hybridization, cDNA sequences, and complete map of lambda Charon 4A vector. Out of totally 6 potential models (Fig. 2.4B), one was finally determined to be the correct one (Fig. 2.4A).

The discrepancies between data of Vannice et al. and what is presented in this dissertation are most certainly due to their use of DNA prepared from a modified hepatoma cell line which obtained in artifical Eco RI site during mutagenesis manipulations. These facts indicate that the rat α_1-AGP gene isolated, characterized and eventually sequenced in this dissertation present a naturally intact and continuous gene copy.

These results do not agree with the published data of Vannice and Ringold et al. (1984). In their paper, a photograph of Southern blotting was presented, although major restrictive fragments of Hind III (3.0 kb), Bam HI (2.5 kb), Pvu II (5.0 kb), Pst I (3.0 kb), Bgl II (8.0 kb) are found, there were many additional fragments. Most importantly, the major 8.5 kb Eco RI band was missing and replaced by a prominant 5.0 kb fragment which was not seen on my Southern

blottings. This is a significant discrepancy, since the Eco RI restriction site was used during genomic cloning and is a main mark characterizing the gene structure. It suggests a potential artifact during recombinant DNA manipulations. Less likely possibilities include intraspecies polymorphism in the α_1-AGP gene and the nature of DNA used. The first possibility is unlikely, since the α_1-AGP gene is highly conserved among species and it is difficult to envision that such divergence has occurred within an extensively inbred Spague Dawley rat. It has been observed on the other hand that different results in restriction enzyme digestion may be encountered when using DNA prepared from livers of different rats in different experiments. Eventually a parallel matching of Southern analysis between rat and recombinant bacteriophage DNA was obtained by using DNA derived from rat hepatoma cell culture. Presumably, DNA in adult rat livers may be very much methylated or modified in different degrees. The DNA used by Vannice and Ringold et al. was also derived from rat hepatoma cell, however, a particular hepatoma cell line designated as JZ.1 was used. JZ.1 cell line was dervied from rat hepatoma cell which has been widely used elsewhere and also in this dissertation. JZ.1 cell line was established by first infecting and then selecting the rat hepatoma cells to contain one single provival copy of MMTV, then heavily mutgenizing with ethyl-methane sulfonate in order to generate a mutant having no steroid responses (Grove PNAS 78:4349-4353, 1981). It is possible that numerous mutations were established, since a total spectrum of glucocorticoid responses involving all related gene expressions including MMTV and α_1-AGP were abolished. It is most reasonable to conclude that the mutation resides on a glucocorticoid receptor gene or on loci of some other steroid hormone response because of the widespread abolishing affects. Although simultaneous damage to all glucocorticoid inducible genes is less likely, an incidental mutation

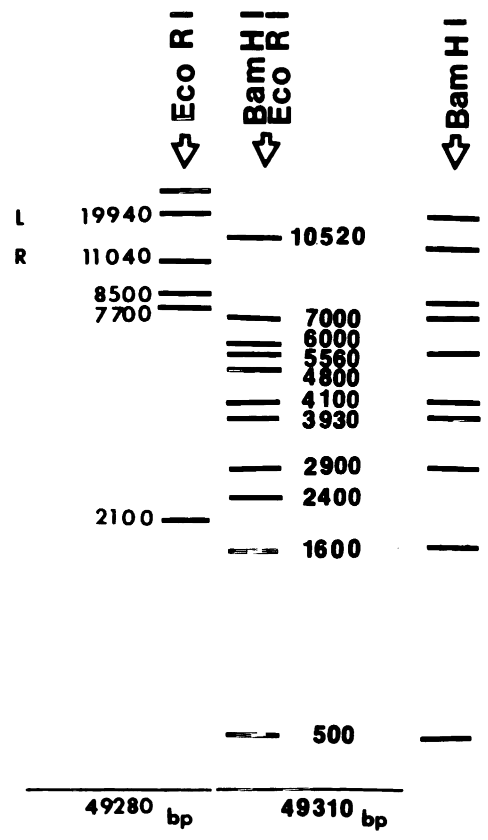
may occur on the α_1-AGP gene to creates an additional Eco Ri restriction site. This artificial Eco RI site must be within the 8.5 kb Eco RI natural fragment and must split it into 5.0 kb and 3.5 kb subfragments. It is predicted that this new Eco RI site exists on the 5' and generated a 3.5 kb fragment which contains no coding areas of α_1-AGP gene message, as a consequence it was undetectable cDNA probe during Southern hybridization (Fig. 2.3). Repeat of Southern analysis by using DNA of the original rat hepatoma cells generated identical Southern patterns to those derived from DNA of lambda AGP 40. Because DNA used in making the rat genomic library include the lambda AGP 40 clone extracted from normal rat tissue, and because its restriction patterns are identical to rat hepatomas, it can be concluded that there is no structural change on the α_1-AGP gene between normal rat cell and rat hepatoma cell.

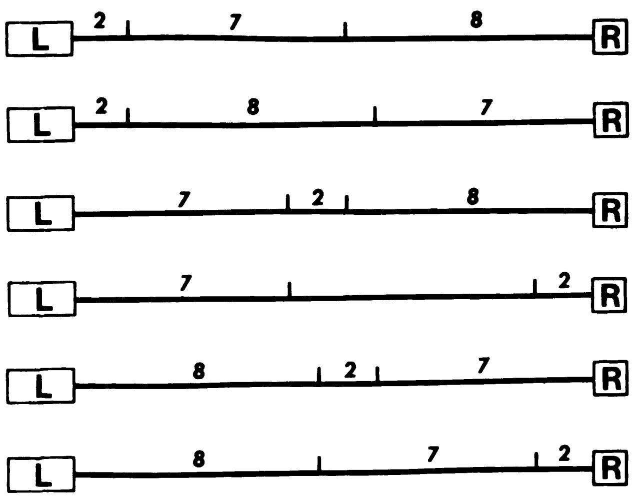
Fig. 2.4. The mapping of the rat α_1-AGP gene.

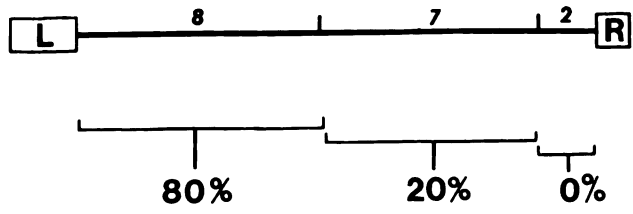
(A) An estimation of subfragment sizes derived from restriction digestion of Eco RI, Bam HI and combination of both. The sizes were extrapolated from log phase plotting. L box represents left arm, R box represents right arm of bacteriophage lambda vector. The numbers represent sizes of fragments in base-pair unit.

(B) Six possible models were listed for the map of the rat α_1-AGP gene based on the number and size of Eco RI digestion. L and R boxes represent left and right arms of bacteriophage lambda vector, respectively. The sizes of lambda arms are not on scale in this figure. The numbers 8, 7 and 2 represent the Eco RI fragments of 8.5, 7.7 and 2.1 kb respectively.

(C) Only one model is possible, after applying the information derived from restriction analysis of Bam HI as well as combination of Bam HI and Eco RI endonuclease digestions. The intensities of radiograph further indicates that 80% of coding region is located within the 8.5 kb Eco RI fragment.


A
Eco RI
Bam HI Eco RI
Bam HI
L
R
19940
11040
8500
7700
2100
10520
7000
6000
5560
4800
4100
3930
2900
2400
1600
500
49280 bp
49310 bp


B
L 2 7 8 R
L 2 8 7 R
L 7 2 8 R
L 7 2 R
L 8 2 7 R
L 8 7 2 R


C
L 8 7 2 R
80%
20%
0%

CHAPTER 3: NUCLEOTIDE SEQUENCING OF THE RAT α_1-AGP GENE

3.1 Introduction

The α_1-AGP gene is central to our study of RNA splicing, maturation, m-RNA stabilities, and nucleocytoplasmic transportations. In order to determine in more detail the molecular anatomy of the rat α_1-AGP gene for intron-exon assignment, it is necessary to determine the complete nucleotide of a gene which can be used as an inducible de novo system. In addition, the DNA sequence gave us an opportunity to identify the existence and location of the hexanucleotide (TGTTCT) which was postulated to be involved in glucocorticoid receptor binding and subsequent induction of gene expression. This chapter describes the strategy, planning, methods and results of the sequencing of the rat α_1-AGP gene.

3.2 Methods and materials

3.2.1 Materials and sequencing techniques

Using new DNA sequence analysis methods has changed nucleotide sequence determination from state of art to routine in the field of molecular biology. Unfortunately, the commonly practiced technique, although not too difficult to learn, is rarely performed without problems. It is obvious that DNA sequencing will continue to a unique field to be performed by specialized persons in handling a project of a large scale.

Currently, two DNA sequencing methods are available: (1) the chemical degradation method of Maxam and Gilbert (Maxam et al., 1977) and, (2) the chain termination method of Sanger, Nicklen and Coulson (Sanger et al., 1977). Until recently, the chemical degradation method has been the method of choice for most researchers because of its straight-forward simplicity of planning and ready availability of the common chemical reagents and laboratory facilities.

This method, however, needs a lengthy preparation of DNA fragments and multi-step degradation procedures. It is used mostly in sequencing short DNA fragments.

Chain termination, using dideoxy analog, is a version of the primed synthesis method. It is a popular method for large projects because of the advancement of M-13 bacteriophage cloning. As a result, the chain termination method has a better overall-rate of generating information. However, this method requires frequent subcloning of desirable DNA fragments and relies heavily on the understanding of molecular biology in order to solve current problems of various kinds.

The reliability of chain termination sequencing method has also suffered in the past from its dependence upon chemical enzymes of the highest quality. This problem has gradually disappeared with the improvement of product quality due to recent commerical competition. The chain termination or Sanger's method was used in sequencing the rat α_1-AGP gene.

Sanger's method involves utilizing a short primer molecule (usually 15 to 17 oligonucleotides) complementary to the DNA lying to the 3' end side of the target DNA region of a single-strand DNA template (Fig. 3.1). After the priming hybridization of oligonucleotide to the SS template, the Klenow fragment of DNA polymerase I will extend the primer according to its complementary template by incorporating the four deoxynucleotide triphosphates (dNTPs) added in the reaction buffer. The addition of 2',3'-dideoxy analogues (ddNTP) which lack the 3' hydroxyl group for continuous chain extension (by esterification) will block the elongation process once they are incorporated. Therefore, it is possible to stop the elongating DNA chains at specific nucleotide regions by adding their respective dideoxy analogues (ddNTP). Length-wise, chain termination will occur at each position occupied by the specific ddNTP analogue according to its complementary template.

By performing a set of four reactions, each containing a different specific ddNTP, a specific pattern representing specifically terminated DNA chains can be produced in each of the four reactions. Technically, due to the simple kinetic competition of substrate and analogue for Klenow fragment of DNA polymerase I, the length of synthetized DNA chains can be adjusted according to the concentration and ratio of the specific ddNTP to the corresponding dNTP in order to set the best reading result.

In the chain elongation and termination reaction, one of the dNTPs is radioactively labelled, therefore, the newly synthesized and specifically terminated DNA molecules are also radioactive. When the reaction is complete, the four fragment sets can be denatured by formamide, heated and size-fractionated by electrophoresis on a high resolution (1-nucleotide-difference) polyacrylamide gel, and autoradiographed. A pattern produced by radioacitve bands of DNA fragments on the x-ray film will represent the difference in length of one nucleotide in each (G,A,T or C) of the reaction set. The ladder-like patterns will indicate the nucleotide position of a DNA molecule at each reaction set. As a result, the pattern combination of successive bands reflecting the nucleotide sequences started from the oligonucleotide priming site can be visualized (Fig. 3.1).

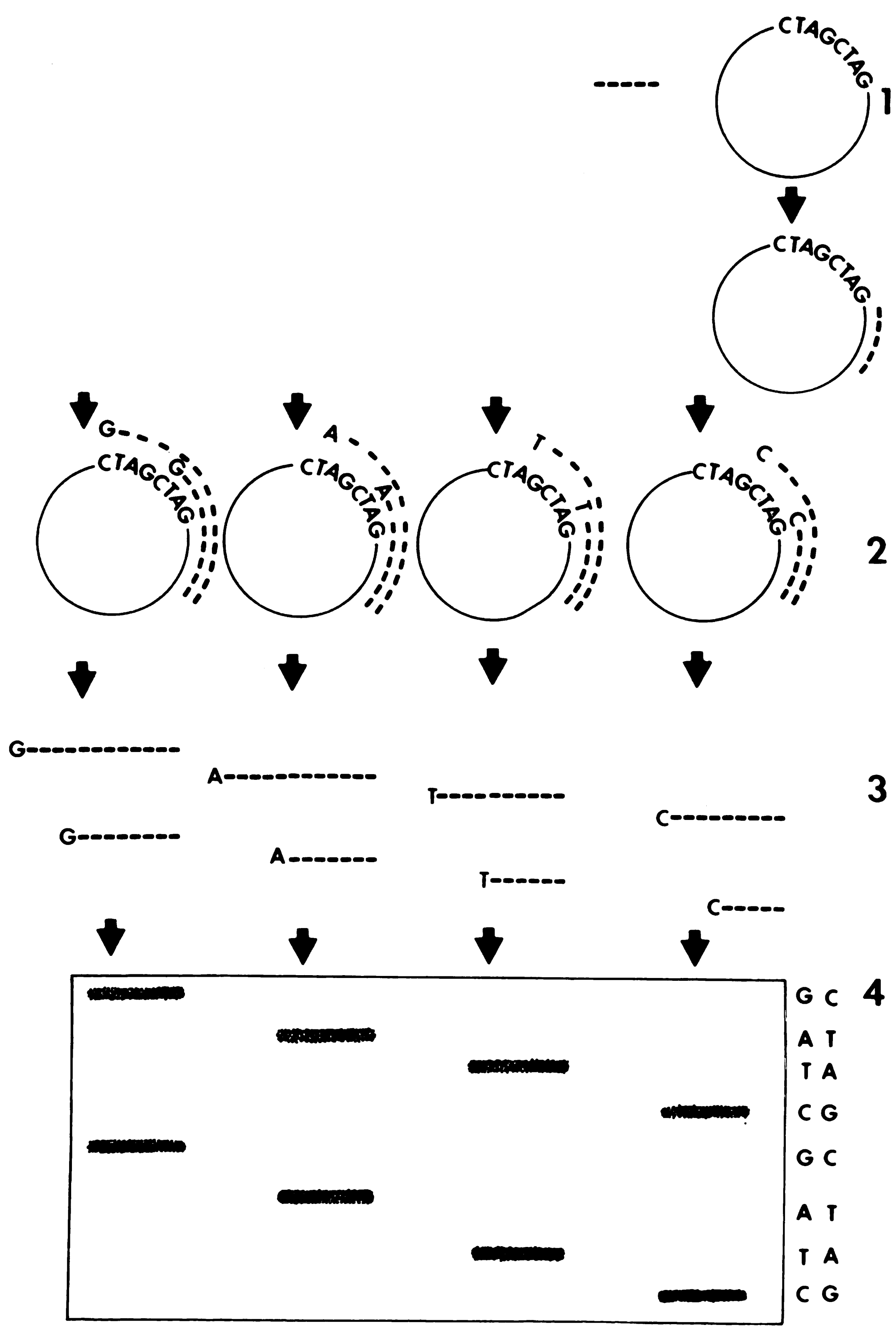
Fig. 3.1. A schematic illustration of M-13 sequencing method.

1. Five-piece-broken-line stands for oligonucleotide primer. The circle represents a single strand M-13 bacteriophage DNA a hypothetical sequence of GATCGATC. In this step, primer is annealed with its complementary sequence at the 3' end of target sequence GATCGATC.

2. During the chain extension and termination reaction, 4 tubes contain designated dideoxynucleotides, and necessary components of salts and buffer are prepared. They are from left to right containing dideoxy-GTP, GATP, GTTP, GCTP, respectively. As a result, the elongating chain would be terminated always at the designated nucleotide of dGTP, dATP, dTTP or dCTP from left to right, respectively.

3. The newly synthesized DNA chains are shown with exact relative lengths in order to emphasize the terminating position at specific nucleotides.

4. After denaturing the radiolabelled DNA chains, they can be fractionated on acrylamide gel and visualized by radiography. By reading the allocated bands of x-ray film, the complementary sequences can be deciphered; in this case CTAGCTAG or GATCGATC complementarily.


CTAGCTAG
1
2
3
4
G C
A T
T A
C G
G C
A T
T A
C G

Originally, the requirement of a single strand DNA template made by this method was of limited usefulness. The problem was later solved by the reconstruction and engineering of an E. coli bacteriophage M-13 (Messing et al., 1977, 1982) which has a single-strand DNA stage in its life cycle (Fig. 3.2). Later, a supercoil DNA sequencing method which involves denaturing and dissociating the double strand DNA by alkaline agents before primer hybridization (Chen, E., et al., 1985) was also developed to sequence double strand DNA.

The M-13 cloning procedure involves first inserting the DNA fragments of interest into the engineering polylinker region of M-13 vector through restriction digestion and ligation. Then the ligated recombinant molecules can be transformed into E. coli host for amplification. M-13 phage particles extruded from infected E. coli cells contains ssDNA strands of M-13 genomes and inserted DNA sequence can be readily purified from its protein coat through simple phenol extraction. The poly-linker or poly-cloning region contains several unique restriction sites, engineered to locate within a segment of a M-13 bacteriophage gene coding the α-peptide of β-galactosidase (Fig. 3.3). A universal primer of 15 to 17 nucleotide-length and complementarily corresponding to a sequence at the 3' side of poly-linker region is chemically synthesized and commercially available. The priming and subsequent sequencing allows us to determine the nucleotide sequence of the cloned DNA fragment. After ligation and transformation, if there is no insertion of foreign DNA into the M-13 vector, the reading frame of α-peptide of β-galactosidase gene of M-13 bacteriophage remains intact and produces a normal β-galactosidase after infecting an E. coli F' strain which has a defective or non-functional β-galactosidase gene. As a result, the E. coli F' strain acquires a complementary gain of a functional β-galactosidase. In the presence of the

Fig. 3.2. Representation of the life cycle of E. coli bacteriophage M-13 (modified from Messing, 1983). The flow of steps:

(A) The rod-shaped virus penetrates the F Pilus and its major protein coat is stripped off on cell membrane. The remaining protein coat (gene III product) considered as "pilot" protein (shaded dots, 3 of them) seems to guide the virus for infecting penetration. The naked ds or ss DNA of M-13 bacteriophage can also be introduced into bacteria through chemical means (white arrows). After getting into bacterial cell they replicate identically as normal bacteriophage.

(B) The viral ssDNA (+ strand) is converted into a double strand circular form or replicative form (RF) DNA by host machinery to synthesis the complementary strand, the (-) strand.

(C) The (-) strand serves as the template for synthesis of progenyphage DNA, the (+) strand.

(D) The gene V product (not shown) binds to the newly synthesized (+) strand and forms a protein ssDNA complex then moves to the periplasmic space. The gene V product then is replaced by coat proteins.

(E) The mature phage particle then is released from the bacterial wall.

(F) A M-13 bacteriophage map. The arrows indicate the origin and direction of replication for (+) or (-) DNA strand. The Roman numerals indicate the number assignment of bacteriophage genes. The exact functions of gene products are still unclear, except for genes III, VIII, and V.

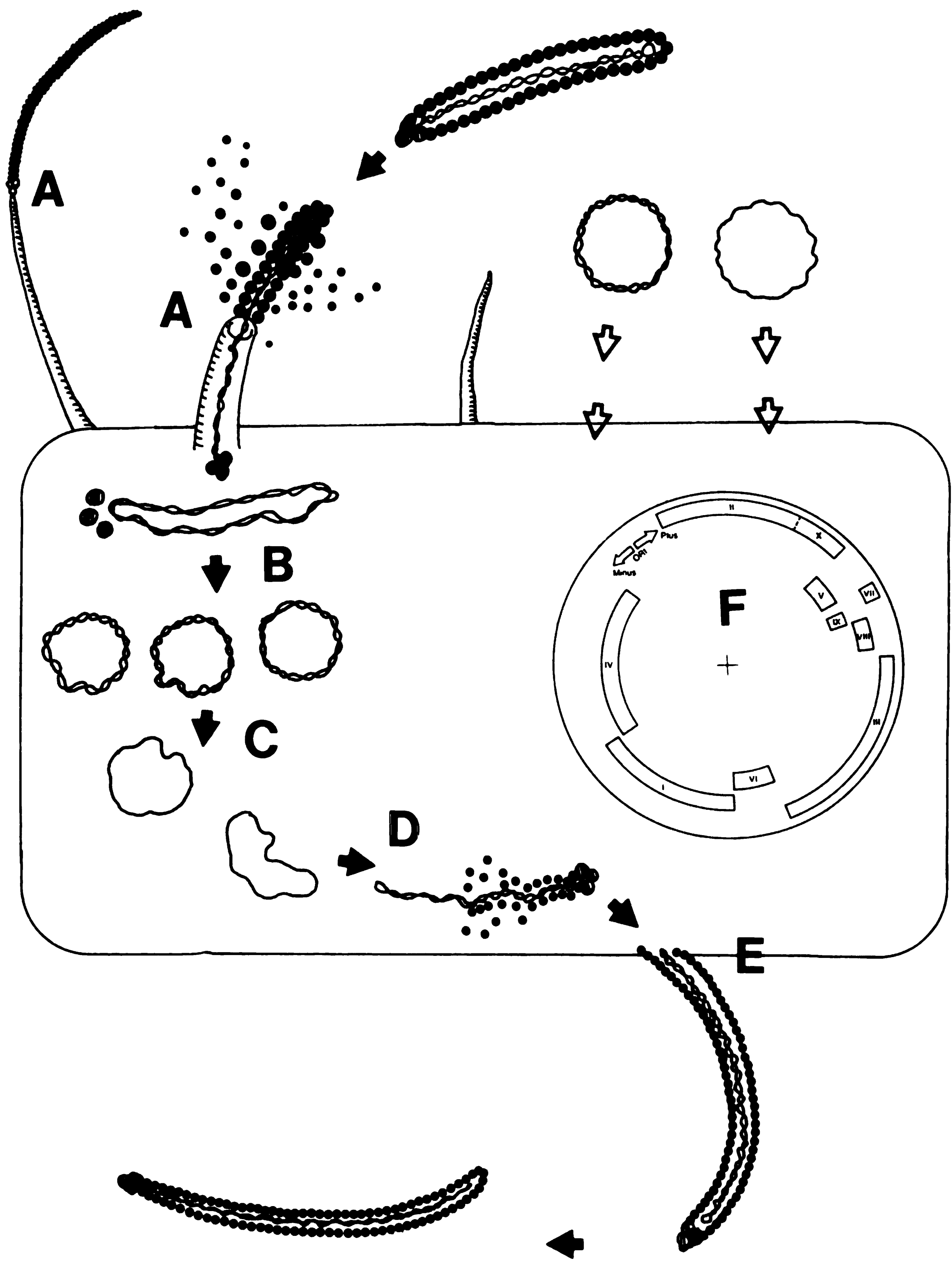
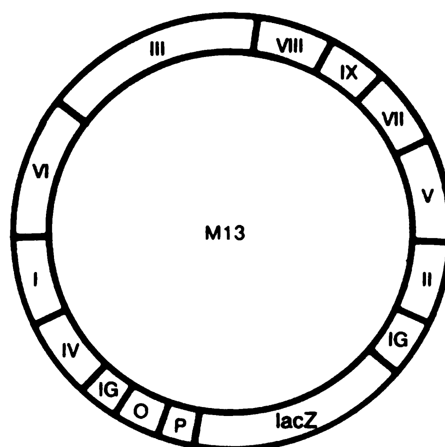

A
A
B
C
D
E
F
II
X
Plus
ORI
Minus
V
VII
IX
VIII
IV
III
I
VI

Fig. 3.3. Diagrammatic representation of the M-13 bacteriophage vector.

(A) The physical restriction map of whole vector.

(B) The nucleotide sequences of polycloning sites of different M-13 versions.



mp18

```
GAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGG
EcoRI SacI   KpnI     BamHI XbaI   SalI  PstI  SphI   HindIII
                 XmaI              AccI
                 SmaI              HincII
```

mp19

```
AAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTC
HindIII      PstI  SalI  XbaI   BamHI    KpnI  SacI   EcoRI
      SphI         AccI             XmaI
                   HincII           SmaI
```

mp10

```
GAATTCGAGCTCGCCCGGGGATCCTCTAGAGTCGACCTGCAGCCCAAGCTT
EcoRI SacI    XmaI BamHI XbaI   SalI  PstI     HindIII
              SmaI              AccI
                                HincII
```

mp11

```
CCAAGCTTGGGCTGCAGGTCGACTCTAGAGGATCCCCGGGCGAGCTCGAATTC
HindIII     PstI  SalI  XbaI   BamHI       SacI   EcoRI
                  AccI              XmaI
                  HincII            SmaI
```

mp 8

```
GAATTCCGGGGATCCGTCGACCTGCAGCCAAGCTT
EcoRI    BamHI SalI  PstI     HindIII
     SmaI      AccI
     XmaI      HincII
```

mp 9

```
AAGCTTGGCTGCAGGTCGACGGATCCCCGGGAATTC
HindIII  PstI  SalI  BamHI    EcoRI
               AccI      SmaI
               HincII    XmaI
```

inducer isopropyl-β-D-thio-galactopyronoside (IPTG), the β-galactosidase of the transformant is able to hydrolyze a chromogenic substrate 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (x-gal) to bromo-chloroindole which is blue in color. A blue plaque on a bacterial lawn of an agar plate, therefore, indicates a colony containing infected clones. If a DNA insertion occurs on any of the polycloning sites so that the reading frame of β-galactosidase gene is interrupted, the synthesis of β-galactosidase will be either terminated early or produce a lengthy but non-functional fusion protein. As a consequence, the infected plaque is colorless and can be selected (Fig. 3.4). Rarely, but possibly, the insertion of cloned DNA remains in frame with the triplet translations and produces a fused or recombinant β-galactosidase-like peptide and contains no early termination codon. In this case, the β-galactosidase-like peptide may be functional and a blue plaque is produced. Most of the time, the colony can be distinguished from the others by its less intensive color due to the weaker activity of chromogenic hydrolysis by the fusion protein.

Fig. 3.4. A schematic representation of cloning a DNA fragment into M-13 bacteriophage.

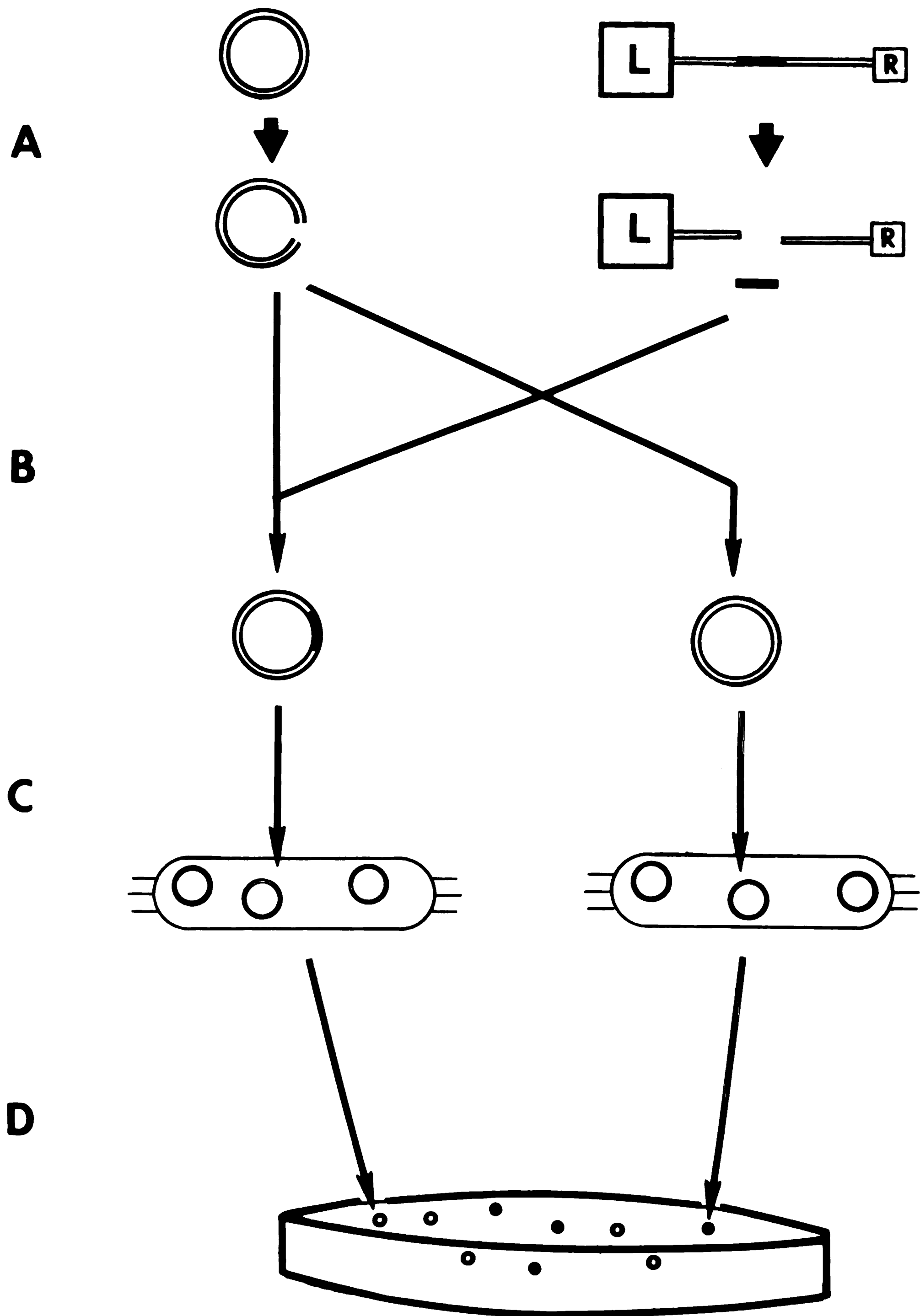
(A) A circular M-13 vector is cleaved open at a specific restriction site.

(B) Target DNA is also cleaved open by the same restriction enzyme.

(C) Vector is ligated with restriction endonuclease-digested target DNA.

(D) Ligated DNA recombinant is transformed into E. coli host cells.

(E) Transformed bacteria are plated out on an agar plate. The colonies containing the M-13 vector without DNA insertion produces translucent plaques (white dot); the colonies absorbing the M-13 vector with DNA insertion produce blue-colored plaques (black dot).


A
B
C
D
L
R
L
R

By using the common restriction sites found on both M-13 vectors and target DNA, the most direct and simplest sequencing strategy is to clone overlapping DNA fragments derived from these sites to generate overlapping recombinants for sequencing reactions. It is possible to accumulate the sequences by overlapping fragmental data to determine the continous sequences of an entire DNA molecule.

In reality, considerable problems have been encountered by this simple strategy. This is due to the non-random distribution of desirable cloning sites on target DNA, the tendency of vector ligation with smaller restriction fragments, and the stability of the cloned DNA in recombinant M-13 phage. Although viral DNA up to 6 times the length of M-13 DNA has been packaged (Messing, unpublished), deletion may occur within the cloned DNA.

In order to solve these problems, fragmenting DNA by DNase I, Bal 31 nuclease, exonuclease III, T4 DNA polymerase and sonication have been employed to produce a series of recombinants in which the inserted DNA fragments have either random (by DNase, sonication) or progressive deletions from one end (exonuclease III, T4 DNA polymerase) or both ends (Bal 31 exonuclease). In the project of sequencing the rat α_1-AGP gene, none of these back-up methods was necessary. The problem of instability of foreign DNA (especially of mammalian genes) in M-13 vector is partially solved by using the pUC vector which was engineered to contain identical poly-linker and priming sites but multiplies in E. coli as a double strand plasmid.

In order to improve the resolution and clarity of sequence autoradiograms, radioactive label of $[^{35}S]$-α-thio dATP and a wedged ultra-thin (0.175 mm) acrylamide gel were used. The $[^{35}P]$-α-dATP was occasionally employed in cases of solving sequence compression in combination with inosine mixture and higher reaction temperatures. The technical details of M-13 cloning and sequencing

are listed in Appendix B, classified into sequential sections for easy identification and for following instructions.

Enzymes. Restriction endonucleases were purchased from New England Biolabs and Bethesda Research Laboratories, except high concentration Eco Ri, Sma I and Sal I which were obtained from Boehringer-Mannheim. The large Klenow fragment of DNA polymerase I was also purchased from Boehringer-Mannheim. The dideoxy mix was purchased from Bethesda Research Laboratories. The 17mer priming oligonucleotide was obtained from Biosearch and the Isotope was obtained from Amersham.

In Appendix B, the complete description includes preparation of template DNA, chemical reaction of chain termination and manual information for handling the gel and autoradiograph.

A general description of cloning target DNA into M-13 vector, and its transformation and selection is also described.

3.2.2 Sequencing strategies

Usually, in order to sequence a DNA fragment larger than 2 kb, it is best to subclone it into pUC plasmid for amplification before the actual sequencing takes place. The purpose is to ensure a sufficient quantity of material for M-13 subcloning, to have the confidence of dealing with a specific fragment of DNA, and to avoid the instability problem of large inserts. In the case of the rat α_1-AGP gene, although the two fragments harboring this gene are large (8.5 and 7.7 kb respectively), it was decided to subclone them directly into M-13 vectors bypassing the pUC requirement. The decision was made with the following reasons: 1) subcloning into pUC takes additional time, 2) the coding regions and the lambda AGP40 clone are well characterized, 3) there is only one gene in rat, 4) there are unique cloning sites in cDNA, and 5) the cDNA sequences of

α_1-AGP are known and computer programs are available for quick analysis. Based on these advantages, lambda DNA were restriction-digested and directly subcloned into unique and specific restriction sites of M-13 vectors. This procedure is termed "shot-gun cloning".

A sequencing strategy was formulated by taking advantage of the existing Sal I and Bam HI sites in the α_1-AGP cDNA. Recombinant bacteriophage DNA was digested with Sal I, Bam HI, and Eco RI, then directly subcloned into M-13 bacteriophage polylinker sites. Because of the rarity of the Sal I site in the mammalian genome and lambda bacteriophage, I decided to clone the reading frame on the first trial. The sequence was then expanded from the Sal I site toward both the 5' and 3' ends of the gene (Fig. 3.5). Further subcloning was undertaken using identified restriction sites of the new sequence and convenient Sac I sites found on positively hybridized fragments on Southern blot analysis (Fig. 2.3). The total sequencing was completed by overlapping sequences of M-13 subclones and extended about 1 kb beyond the TATA box and polyadenylylation signal AATAAA at the 5' and 3' ends, respectively. The 5' to 3' end orientation of the α_1-AGP gene in recombinant lambda bacteriophage was determined from the nucleotide sequences of exons and the relative locations of restriction sites, including Eco RI, Sac I, Bam HI and Sal I, on Southern hybridization.

3.3 Results

As a result of specific cloning by the shot-gun approach, about 100 different M-13 subclones were obtained by using various restriction sites for ligations. The identification of cDNA sequences based on Sal I site toward both 5' and 3' end proved very useful. The M-13 clone, mp 8 Sal I → Bam HI 2 and mp 18 Sal I → Eco RI 1 coding the reading frame of the rat α_1-AGP cDNA were among the first to be sequenced. In order to prove that the Sal I is a natural restriction

site instead of being a result of RNA splicing, a M-13 subclone read from Xmn I toward Eco RI was generated later to overlap the Sal I sequence. Among the approximate 100 or more M-13 subclones, a majority of them were found coding the non-rat α_1-AGP areas and were screened out during preliminary c-trackings. Totally, 18 M-13 subclones were found containing the target gene by overlapping of extended sequences. These 18 M-13 subclones cover the entire rat α_1-AGP gene and are listed in Appendix CI according to their vector, cloning sites, lengths of insertion and directions of sequence readings.

The entire nucleotide sequences of the rat α_1-AGP gene were put together by overlapping the sequences of M-13 subclones and is shown in Appendix CII. Computerized and complete restriction sites and their nucleotide locations are listed in Appendix CIII, and future applications including engineering and subconstruction for expressions or intron probes are noted.

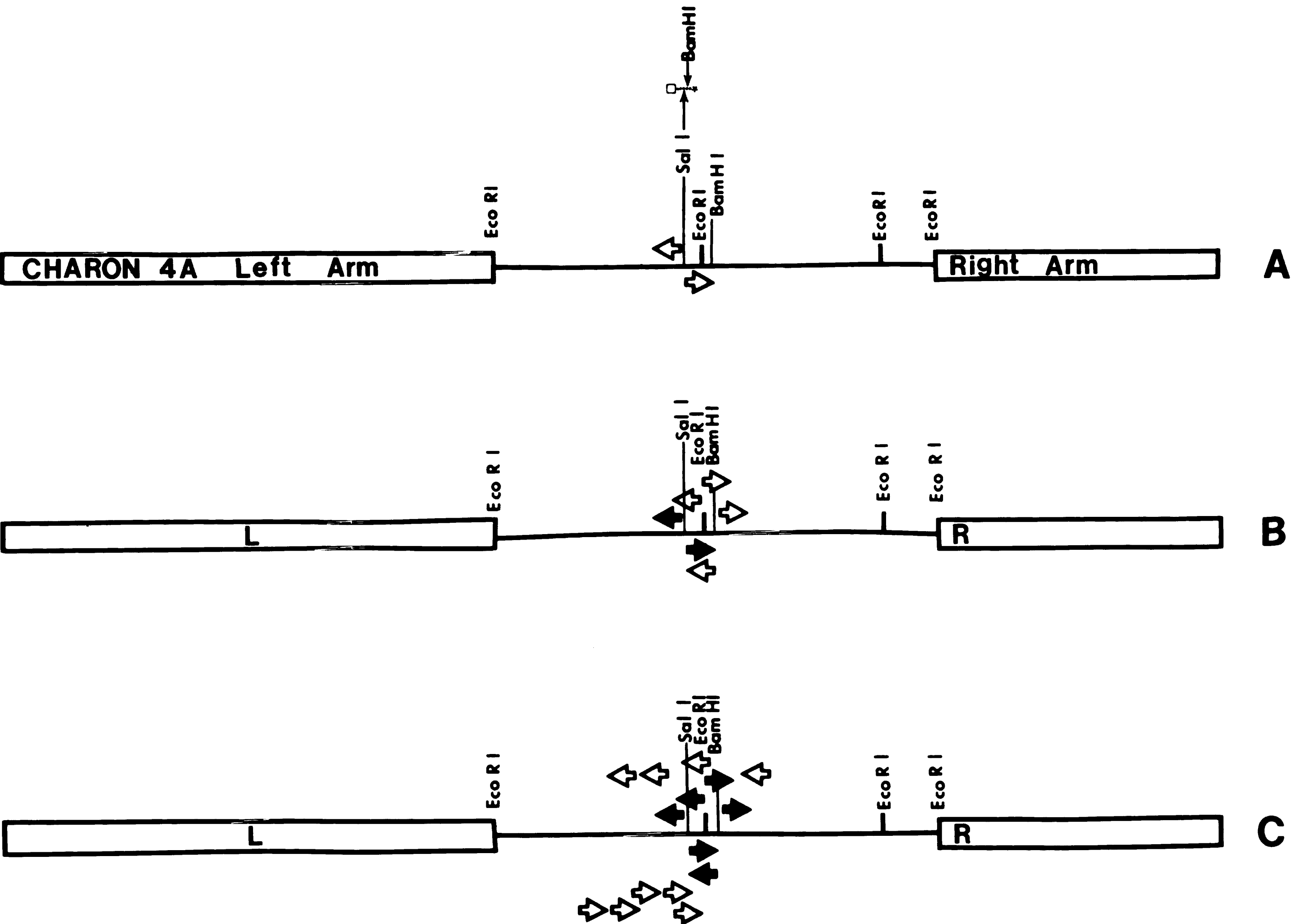
Fig. 3.5. The sequencing strategy was organized into three separate steps:

(A) Using the existing restriction sites of Sal I and Bam HI and the genomic boundary of Eco RI, two clones going from Sal I toward Eco RI and Sal I toward Bam HI were obtained. The empty arrows indicate the planned direction of cloning. The square box, dots and asterisk represent the cap, the coding sequence and the poly(A) tail of α_1-AGP mRNA respectively.

(B) The black arrows represent resulting clones derived from the previous step. In this step, the discovered restriction sites, Eco RI and Bam HI, were used in order to generate clones starting from Eco RI toward Sal I (5' end), Bam HI (3' end) and from Bam HI toward Sal I (5' end), Eco RI (3' end).

(C) Totally, 6 clones derived from the planning of Steps A and B were established (black arrows). Based on the discovered restriction sites from sequencings and restriction mappings, nine additional subclonings either centering in, expanding from, or overlapping with the Sal I, Eco RI, Bam HI sites were undertaken. About one hundred M-13 clones were derived during the whole process, essentially 18 of them cover the need of sequencing the entire rat α_1-AGP gene.

Strategies of Sequencing The Rat Alpha 1-AGP Gene



3.4 Discussion

The complete nucleotide sequence of the rat α_1-acid glycoprotein gene has been determined from an isolated lambda recombinant bacteriophage. Southern blot analysis and DNA sequencing indicate that there is only one gene per genome; it contains six exons and is located within a 3200-base-pair fragment starting from a "TATA" box and extending to the polyadenylylation signal AATAAA. Transcription starts 37 base pairs upstream from the beginning of translation codon ATG (Ricca et al., 1981). The TATA box (TATAAA) lies 26 base pairs upstream from this site. The α_1-acid glycoprotein gene is known to be regulated by glucocorticoid hormones. Although this regulation is thought to be an indirect hormonal action, the gene contains several potential glucocorticoid receptor binding sites, both inside and outside the structural gene.

The organization of AGP gene as shown in Fig. 3.6 consists of six exons and five introns and strictly observes Chambon's rule; all introns begin with dinucleotide G-T and end with A-G. The more detailed consensus sequences of donor [G-T-$\frac{A}{G}$-T-T-G-T] and the acceptor [$(\frac{T}{C})_{11}$-N-$\frac{C}{T}$-A-G] proposed by Breathnach and Chambon (Breathnach et al., 1981) are reasonably matched by the splicing junctions of the introns and exons in the rat α_1-AGP gene.

Comparison of exons of the AGP gene with previously published cDNA sequences shows that there is only one base difference: G instead of A at the fifth nucleotide after the translation stop codon TAG on the 3' end. The difference could be due to intraspecies polymorphism or replication error of lambda bacteriophage. It has no effect on amino acid coding sequences.

At 5' End Region, the transcriptional initiation site was determined previously by reverse transcriptase catalyzed cDNA synthesis in the presence of a 67-bp Dde I cDNA fragment as primer and chain-terminating dideoxynucleotide

triphosphates (Ricca et al., 1981). The nucleotide residue "A," 27 bases downstream from the TATA box, is the transcriptional start.

The Goldberg-Hogness box ("TATAAAA") is located between 21 and 27 bases upstream from the transcriptional start. There is no apparent consensus sequence of a "CAAT" box in its anticipated (50 bp) upstream position; however, there may be a mixed "TATA" and "CAAT" sequence between 262 and 272 bases upstream from the transcriptional start. Its distant location indicates that it may not be a CAAT box; however, shortly repeated GA and CT, which resembles Z-DNA structure, on its flanking areas suggests a potential involvment in expressional regulation. A hexanucleotide sequence, T-G-T-T-C-T, which has been postulated to be a binding site for the glucocorticoid receptor complex in MMTV (Scheideveit et al., 1983), is found 431 bp upstream from the transcriptional start. The hexanucleotide sequence was also found previously 451 bp upstream from the transcriptional start of the human renin gene; hence, the renin gene was speculated to be under steroid hormone regulation (Miyazaki et al., 1984).

Intron Regions. The potential glucocorticoid receptor binding sequence of T-G-T-T-C-T is also present three times, within introns located at positions 1511, 3281 and 3553, respectively (Fig. 3.7). A survey of intron regions shows unusual patterns of nucleotide sequences, including the existence of an alternating purine-pyrimidine of Z-DNA-like structure, followed by 14 Gs at location 3383 (Fig. 3.7). The neighborhoods of the glucocorticoid binding hexanucleotide consist of partially symmetrical $G(T)_{2-5}$ repeats GTTTTT(TGTTCT)GTTTTGTTTGTTGTTTT GTTTTT, located at 3547 bp (Fig. 3.7). This is followed twice by a sequence of ACCCTGCA 7 bp and 24 bp downstream. Intron 4 has only 130 bp and contains sequences of $T(C)_5A(C)_3ACA$ and CCCATT in the middle, with their symmetrical counterparts at the acceptor side of the splicing junction (Fig. 3.7).

At the 3' End Region, the polyadenylylation signal (AATAAA) is found 18 nucleotides upstream from the beginning of the poly(A) tail formation site. A decanucleotide sequence of TCCT$^{C}_{T}$CT$^{CCC}_{TTT}$ is repeated four times, overlapping one another within a 30-bp region located 15 bp downstream from the poly(A) addition signal and immediately followed by a complementary GGAGAGGAGG. These particular sequences created five Mnl I restriction sites within 40 bp's length, and produce a highly symmetrical region in centering at the second Mnl I. Its significance is unknown, although it may play a role in the disassociation of RNA polymerase II from the DNA reading frame.

The several-hundred-fold induction of AGP mRNA observed in glucocorticoid-treated HTC rat hepatoma cell indicates a close relationship between this gene and steroid hormones. Steroid hormones were believed to act by binding to a soluble cytoplasmic receptor protein, which in turn enters the nucleus, binds to specific DNA sequences adjacent to the regulated genes, and activates transcription (Jensen E.V. et al., 1968; Gorski, J., 1968). The induction of α_1-AGP mRNA synthesis in contrast to that of MMTV RNA appears to depend on the ongoing protein synthesis (Vannice, J.L. et al., 1983; Feinberg, R.F. et al., 1983; Bauman, H. et al., 1983). Our findings indicate that there is a sequence consistent with a glucocorticoid receptor binding site upstream of the α_1-AGP mRNA initiation site and three others within the structural gene. Glucocorticoid and progesterone receptors have been found to bind to overlapping regions of the MMTV and lysozyme DNA (Hamada, H. et al., 1979; Nordheim, A. et al., 1983). Moreover, findings indicate that MMTV, which classically responds only to glucocorticoids, also responds to androgens after it has been experimentally transferred to cells containing androgen receptors (King, R.J.B., 1984). Hence, glucocorticoids and androgens may act on the same DNA binding site. Recently, the identical hexanucleotide sequence has also been found in the human renin

gene (Scheidereit, C., 1984). The hexanucleotide T-G-T-T-C-T in the human renin gene is located almost identically with that in the rat α_1-AGP gene at 450 and 430 bp upstream from mRNA initiation site, respectively. Since this area of DNA has been postulated to be the site of glucocorticoid receptor complex binding, the two genes may be under similar glucocorticoid control.

The second T-G-T-T-C-T sequence of the rat α_1-AGP gene is located within intron 2. The third and fourth sequences are in intron 5. Although the functional significance of additional binding sites is unknown, such sites have been found in the human growth hormone genes (Moore, D.D. et al., 1985). Furthermore, Charney et al. (1984) and Wright et al. (1984) have demonstrated that gene-specific controlling elements of human α- and β-globin gene can be located within the structural gene instead of in the promotor region.

It should be noted that the fourth hexanucleotide is surrounded by a $G(T)_{2-5}$ element of $G(T)_5$[TGTTCT]$G(T)_4G(T)_3G(T)_2G(T)_4G(T)_5$. A Z-DNA-like structure of seven continuous dinucleotide G-Ts, followed by $(G)_{10}A(G)_3$, at position 3382 (Fig. 3.7) is also located in this intron, 139 bp upstream from the fourth glucocorticoid binding sequence. $(GT)_n$ oligomers have the potential for forming a Z-DNA structure under particular conditions, including ionic strength and superhelical density, and are widely distributed throughout eukaryotic genomes (Hamada, H. et al., 1979; Nordheim, A. et al., 1980). It has been suggested that long $(GT)_n$ oligomers ($n<20$), which do not form a typical nucleosomal structure, may act as chromatin structural element only and have no regulatory function. However, smaller Z-DNA sequences ($n<10$), such as those found in rat α_1-AGP gene, have been found in a number of enhancer and transcriptional control sequences from DNA and RNA viruses.

Since the rat α_1-AGP gene does contain potential glucocorticoid receptor binding sites, the gene expression may not be entirely due to secondary action

of the hormone, as previously suggested (Vannice, J.L. et al., 1983; Baumann, H. et al., 1983). We propose that the rat α_1-AGP gene may be directly regulated by glucocorticoid receptor complex, but that a permissive labile factor supplied by continuous protein synthesis in particular cell types must be present in order to sustain high levels of stable mature α_1-AGP mRNA. However, this sequence may not be a bona fide glucocorticoid receptor binding site. This possibility is supported by two facts: 1) the homology of these putative binding sites does not extend further into regions of homology among human growth hormone, human metallothionein II, MMTV genes (Moore, D.D. et al, 1985); and, 2) the transcription rate of the rat α_1-AGP gene is only marginally increased by the addition of glucocorticoid as measured by in vitro transcription in isolated nuclei (Vannice, J.L. et al., 1984). A DNA foot-printing experiment with glucocorticoid receptor binding protection shall resolve this issue.

CHAPTER 4: SEQUENCE ANALYSIS

4.1 Introduction

Originally, when methods of determining the amino acid sequences became available, protein chemists were excited because of the potential to characterize and predict the function of proteins by the composition of amino acid residues. Very soon, scientists realized that the comparison of amino acids of the same analogous proteins between different species revealed insights into evolution at the molecular level. This has been referred to as a protein "clock" and it can be studied in the living species. The vast number of species and numerous proteins provide virtually unlimited living fossils. Differences of amino acids among the same proteins of different species reveals information similar to gross morphological variations and can now be used to construct a phylogenic tree of molecular sense.

The technical development of protein electrophoresis and amino acid sequencing became available in the 1960s. It was immediately discovered that there was extensive protein polymorphism in the natural population, and the rate of amino acid substitution is approximately constant in evolution. Since sequence evolution is divergent, it is reasonable to construct phylogenetic trees from the available sequence data. The tree constructed with this manner may depict the approximate order of divergence of the lineages which in turn may lead to the identification of species originating from a common ancestor. To date, phylogenetic analysis of amino acid data has made a major contribution to verifying the validity of phylogenetic analysis based on morphology and our understanding of evolutionary mechanism on the molecular level.

The information derived from protein phylogenetics also aids in determining the time of divergence between related species. This approach is to use the

known absolute time of one branching event as reference to estimate the approximate time of another branching event in the phylogenetic tree. By knowing the average rate of single amino acid substitution among related species, and the total number of amino acid substitutions on an analogous protein, we can predict the time of divergence between species. For example, the branching event separating the human lineage from that of apes is estimated at 5 million years, which is compatible with fossil evidence. Other studies indicate that the combination of molecular approach and fossil records is useful in establishing a more detailed and accurate time record of evolutional events and is extremely useful in the fields of paleontology, anthropology and systemic biology.

In the late 1970s, a new technology based mainly on restriction enzymes, DNA modification, DNA reconstruction and sequencing has led modern biology into a new era. The rapid accumulation of information on the structure, organization of genes and their evolution and comparison of nucleotide sequencing of the same gene from different types of organisms or different genes in the same organisms revealed unexpected variations of nucleotide changes at different areas of genes, and led to another set of evolutionary hypotheses, including concerted evolution and horizontal gene transfer. The average value of nucleotide substitution derived from a large pool of genes and species samples was discovered and remains relatively constant (Hayashida et al., 1983). This result suggested the establishment of a DNA "clock" which has provided even more insights into the study of evolution than the protein clock.

The advantage of using nucleotide sequences is that we can distinguish nucleotide changes that cause amino acid replacement (nonsynonymous changes) from those that do not (including synonymous changes in protein coding regions and changes in non-coding regions). This distinction is essential because the nucleotide changes demonstrate generation time effect (Laird, C.D., 1969),

whereas amino acid changes do not (absolute time only). The nucleotide changes in the introns and gene flanking regions further reveal the accuracy of time effect which is not represented on amino acid sequences.

The combined uses of DNA and protein clocks have provoked a new surge of theories on the molecular mechanisms of evolution and have generated many controversies. Among these hypotheses, Kimura's neutral theory, which emphasized that amino acid substitutions, and hence protein polymorphisms are mostly a consequence of neutral mutations and random genetic drift instead of pure Darwanian selection have gained major support from accumulated evidence (Kimura, 1983).

FUNCTION

The studies of α_1-AGP mimic the general trend of scientific developments as mentioned above, however, it has not been applied to phylogenetic study. The amino acid residues of human α_1-AGP were determined in 1973 by using cyanogen bromide fragmentation (Schmid et al., 1973). The rat α_1-AGP cDNA was cloned and its amino acid (deduced from cDNA sequences) and nucleotide residues were determined in 1981 (Ricca et al., 1980, 1981). In 1985, the human α_1-AGP cDNA was cloned and sequenced (Luciana et al., 1985). Meanwhile, the same laboratory also isolated and sequenced the human α_1-AGP gene (unpublished). The rat α_1-AGP gene was separately and independently isolated and sequenced by Reinke et al. (1985) and Liao et al. (1985). Despite the extensive studies over more than a decade, the definite function of α_1-AGP remains unknown. Although it is desirable to subject the human and rat α_1-AGP for detailed phylogenetic study in order to identify its function by linking to related protein including using its amino acids, nucleotide sequences of both cDNA and gene, it is probably not suitable to follow it into detailed, sophisticated and complex statistical, and mathematical computations at this stage. This is due to the

existence of certain important discrepancies in evolution at sequence and organismal levels. Although the proteins of two groups of organisms such as mammals and amphibians have similar rates of sequence changes, it has been discovered that the mammalian morphologies have evolved a great deal while the amphibians remain essentially unchanged during the same span of time. It has not been determined if the difference can be solely explained by the selection pressure from the varieties of land terrains for mammals and the relative simplicity of aquatic environments for amphibians. Similarly, controversies have also been found among different classes of proteins which evolve in drastically different speeds even in the same species. This chapter intends to analyze the available amino acids and nucleotide sequences for pursuing the extrapolation of the α_1-AGP's biochemical characteristics and its possible functions. The relationship on species of the evolution of α_1-AGP is also discussed at a conservative level without trespassing into phylogenetic biology.

4.2 Methods and materials

In order to examine and analyze sequences of approximate 800 base pairs cDNA and 200 amino acid residues of both human and rat α_1-AGP, it is more efficient to use various computing softwares. In certain cases, visual identifications and comparisons are obvious and easy, however, graphs, plots, homology matches, etc., allow us to visualize the significances with distinction. For these reasons, a variety of figures are employed in this chapter.

4.3 Results

The results derived from sequence analysis can easily become awkward descriptions. In order to depict the results in a more objective and simple manner, the results are organized into systematic illustrations of figures. In

principle, the results of analyses and comparisons are arranged as of rat vs. rat, human vs. human and rat vs. human, at levels of amino acid residues first, then nucleotide sequences. In this section, therefore, descriptions are limited to the figure legends. The order of figures are also organized to coincide with the discussions in Section 4.4.

Fig. 4.1. A homology plot of the rat α_1-AGP vs. itself. The x and y axis represent the amino acid sequence of the rat pre-α_1-AGP which is the primary translate without leader sequence cleavage. The amino acid sequences started from position 0 as the first amino acid of amino terminal toward positions 206 which represents their final carboxyl terminals. The rat pre-α_1-AGP has 205 amino acid residues including a leader signal of 20 amino acids. The number 206 on the plot is due to the additional counting of the stop codon. In this particular plot, matching condition is set at the window of 7 amino acid and a match of 4 amino acid, a most commonly used condition. Under this condition, every matching of 4 uninterrupted amino acids within a 7 amino acid window span, generates a scoring point represented by a dot on the plot. A self-plotting, therefore, ensure a 100% linearity. The drifted dots represent the presence of sequence repeats; its significance is discussed in Section 4.4.

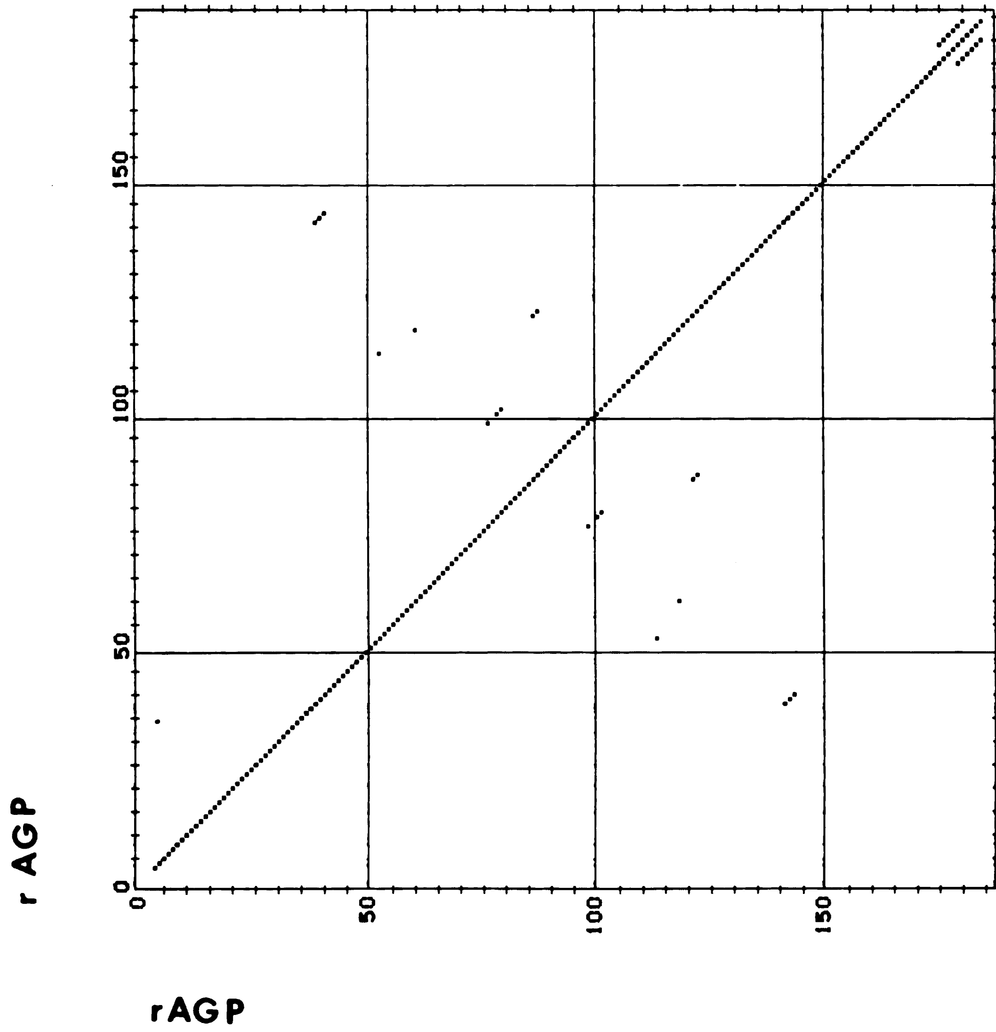

4 : 7
r AGP
rAGP
0
50
100
150
0
50
100
150

Fig. 4.2. A homology plot of the rat pre-α_1-AGP vs. itself. This figure is identical to Fig. 4.1, except that leader signal is included. The purpose is to show the existence of repeated amino-acid residue in leader signal.

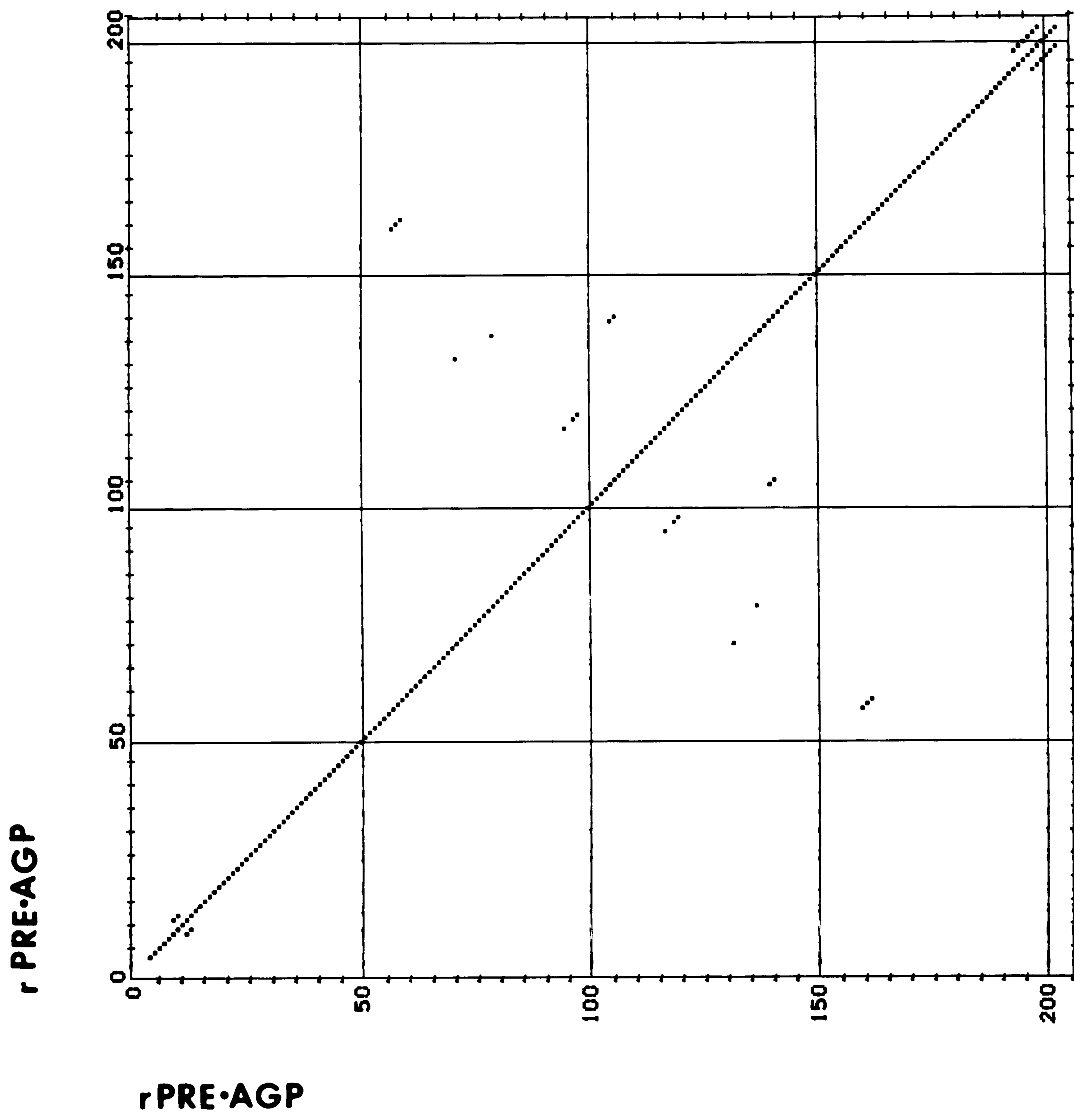
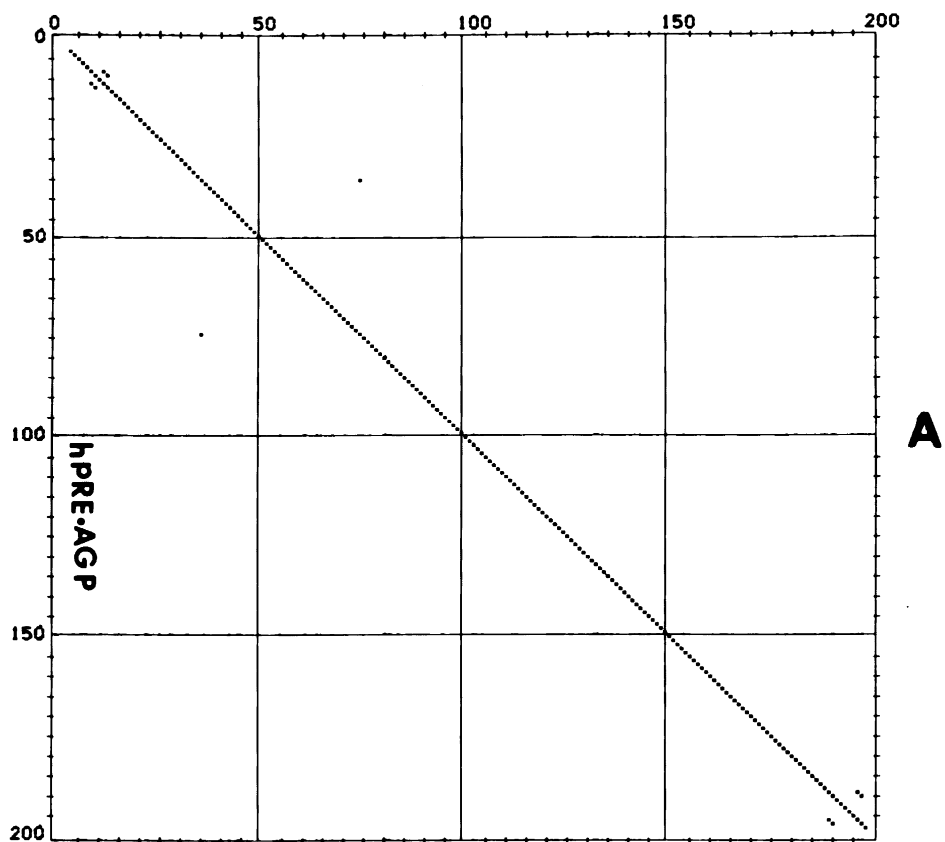

4 : 7
r PRE·AGP
r PRE·AGP
0
50
100
150
200
0
50
100
150
200

Fig. 4.3. A. A self-homology plot of the human pre-α_1-AGP amino acid sequences. Both x and y axis stand for the human pre-α_1-AGP. The description is essentially the same as the rat version, except the length of human α_1-AGP is slightly shorter than the rat's. The unprocessed human α_1-AGP has 202 amino acids. After the cleavage of leader sequence of 19-amino-acid peptide, the remaining protein has 183 amino acid residues. The result of this plot is discussed in Section 4.4.

B. A self-homology plot of the mature rat α_1-AGP amino acid sequences. The purpose is to show the disappearance of sequence repeat at the amino terminal after taking away the leader signal.



7 : 4

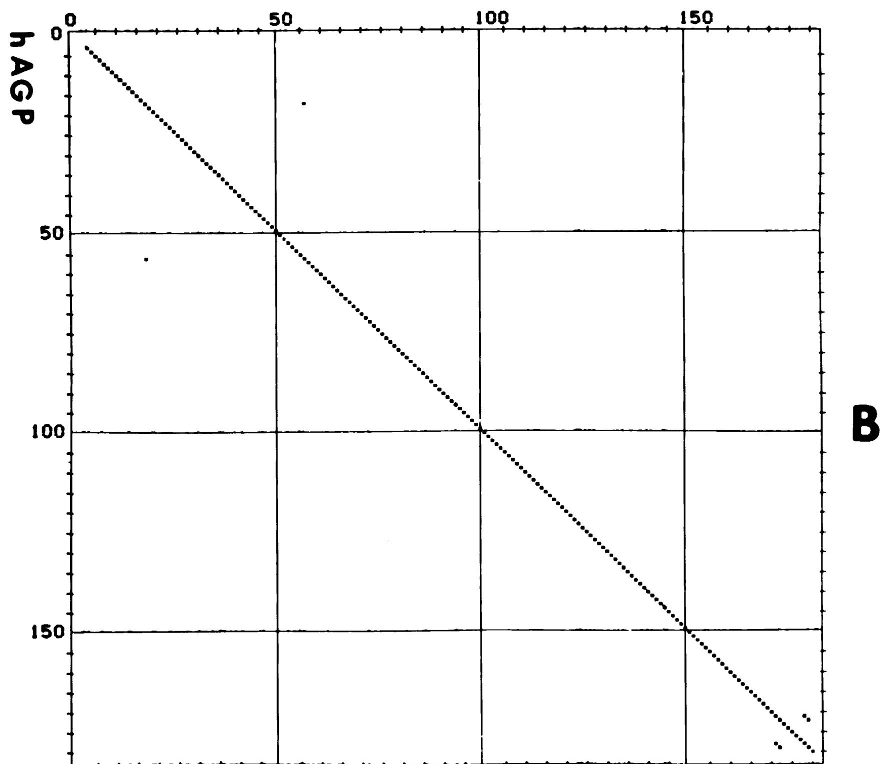


Fig. 4.4. This figure illustrates both the cDNA and amino acid sequence of the rat α_1-AGP. The triple dots on codon ATG emphasize the point of translation start, the triple broken line on codon CAG of glutamine indicates the first amino acid of mature rat α_1-AGP in circulation. The triplet codons, their corresponding amino acids and locations on this figure are discussed and referred in the Discussions.

```
length: 780

                                         •••
  1 GCTCTTCCTGGGCCGGTGCCTCTGAGTGTCTTCGGC    ATG GCG CTG CAC ATG GTT CTT GTC GTT TTG AGC CTC CTG
  1                                        Met Ala Leu His Met Val Leu Val Val Leu Ser Leu Leu

                         ___
 76 CCC TTG TTG GAA GCT CAG AAC CCA GAA CCT GCC AAC ATC ACC CTA GGC ATA CCT ATT ACC AAT GAG ACC
 14 Pro Leu Leu Glu Ala Gln Asn Pro Glu Pro Ala Asn Ile Thr Leu Gly Ile Pro Ile Thr Asn Glu Thr

145 CTG AAA TGG CTC TCA GAC AAA TGG TTT TAC ATG GGA GCA GCT TTC CGA GAC CCC GTG TTC AAG CAG GCA
 37 Leu Lys Trp Leu Ser Asp Lys Trp Phe Tyr Met Gly Ala Ala Phe Arg Asp Pro Val Phe Lys Gln Ala

214 GTT CAA ACG ATA CAG ACG GAA TAT TTT TAC CTT ACC CCC AAC TTG ATA AAC GAC ACA ATT GAA CTT CGG
 60 Val Gln Thr Ile Gln Thr Glu Tyr Phe Tyr Leu Thr Pro Asn Leu Ile Asn Asp Thr Ile Glu Leu Arg

283 GAG TTT CAG ACC ACA GAC GAC CAG TGT GTC TAT AAC TTC ACC CAT CTA GGA GTC CAG AGA GAG AAT GGG
 83 Glu Phe Gln Thr Thr Asp Asp Gln Cys Val Tyr Asn Phe Thr His Leu Gly Val Gln Arg Glu Asn Gly

352 ACC TTA TCC AAG TGT GCA GGA GCA GTG AAA ATC TTT GCC CAT TTG ATA GTG CTG AAG AAA CAT GGG ACC
106 Thr Leu Ser Lys Cys Ala Gly Ala Val Lys Ile Phe Ala His Leu Ile Val Leu Lys Lys His Gly Thr

421 TTC ATG CTT GCC TTT AAC CTG ACA GAT GAG AAC CGG GGG CTG TCC TTC TAC GCT AAA AAG CCA GAC TTG
129 Phe Met Leu Ala Phe Asn Leu Thr Asp Glu Asn Arg Gly Leu Ser Phe Tyr Ala Lys Lys Pro Asp Leu

490 TCC CCA GAG CTG TGG AAA ATA TTC CAG CAG GCT GTC AAA GAT GTG GGC ATG GAT GAA TCA GAA ATC GTA
152 Ser Pro Glu Leu Trp Lys Ile Phe Gln Gln Ala Val Lys Asp Val Gly Met Asp Glu Ser Glu Ile Val

559 TTT GTC GAC TGG ACA AAG GAT AAG TGC AGT GAG CAG CAG AAG CAG CAG CTG GAG CTG GAG AAG GAG ACT
175 Phe Val Asp Trp Thr Lys Asp Lys Cys Ser Glu Gln Gln Lys Gln Gln Leu Glu Leu Glu Lys Glu Thr

628 AAG AAG GAG ACC AAG AAG GAT CCT TAG GCCAAGCATGAACTCAGCTCTCTGAACTCCGGGGACTGTCCCCATGCCCACTCTA
198 Lys Lys Glu Thr Lys Lys Asp Pro AM*

    CCCCCACTCCTGTGCACCTCGATTCTATTTTCCACAATAATAAAGGTTTGGTTAAACAATCAAAAAAAAAAAAAAAAAAAAAAAAAA 3'
```

Fig. 4.5. A. A summarizing figure representing the counting of translational triplet codes for the rat pre-α_1-AGP. The number indicates the times of codon (sit next to it) usage in coding the corresponding amino acid (next to the codon) in the rat pre-α_1-AGP. The asterisks represent the stop codons.

B. The rat pre-α_1-AGP represented by single letter amino acid abbreviation, followed with computed number of residues, number of stop codon and moleculer weight of dalton unit.

Information represented in this figure is referred and discussed in Section 4.4.

A

6/UUU/Phe	Ø/UCU/Ser	2/UAU/Tyr	2/UGU/Cys
6/UUC/Phe	3/UCC/Ser	3/UAC/Tyr	1/UGC/Cys
1/UUA/Leu	2/UCA/Ser	Ø/UAA/OC*	Ø/UGA/OP*
3/UUG/Leu	Ø/UCG/Ser	1/UAG/AM*	4/UGG/Trp
3/CUU/Leu	3/CCU/Pro	3/CAU/His	Ø/CGU/Arg
1/CUC/Leu	2/CCC/Pro	Ø/CAC/His	Ø/CGC/Arg
2/CUA/Leu	3/CCA/Pro	1/CAA/Gln	1/CGA/Arg
7/CUG/Leu	Ø/CCG/Pro	12/CAG/Gln	2/CGG/Arg
2/AUU/Ile	1/ACU/Thr	2/AAU/Asn	1/AGU/Ser
3/AUC/Ile	9/ACC/Thr	7/AAC/Asn	Ø/AGC/Ser
5/AUA/Ile	4/ACA/Thr	7/AAA/Lys	1/AGA/Arg
3/AUG/Met	2/ACG/Thr	12/AAG/Lys	Ø/AGG/Arg
1/GUU/Val	3/GCU/Ala	5/GAU/Asp	Ø/GGU/Gly
4/GUC/Val	3/GCC/Ala	7/GAC/Asp	2/GGC/Gly
1/GUA/Val	4/GCA/Ala	5/GAA/Glu	3/GGA/Gly
4/GUG/Val	Ø/GCG/Ala	1Ø/GAG/Glu	3/GGG/Gly

B

MALHMVLVVLSLLPLLEAQNPEPANITLGIPITNETLKWLSDKWFYMGAAFRDPVFKQAV
QTIQTEYFYLTPNLINDTIELREFQTTDDQCVYNFTHLGVQRENGTLSKCAGAVKIFAHL
IVLKKHGTFMLAFNLTDENRGLSFYAKKPDLSPELWKIFQQAVKDVGMDESEIVFVDWTK
DKCSEQQKQQLELEKETKKETKKDPO

<2Ø5 residues, 1 stop; molecular weight: 236Ø8.49

Fig. 4.6. A. Essentially as described in Fig. 4.4A, except that this figure represents the rat α_1-AGP of circulating form.

B. An entire amino acid sequence of the rat α_1-AGP in abbreviations.

The purpose of Figs. 4.4 and 4.5 are to illustrate the point of codon degeneracy and neutral mutation which is discussed in Section 4.4. They also show the kind of amino acids used in leader signals.

6/UUU/Phe	Ø/UCU/Ser	2/UAU/Tyr	2/UGU/Cys
6/UUC/Phe	3/UCC/Ser	3/UAC/Tyr	1/UGC/Cys
1/UUA/Leu	2/UCA/Ser	Ø/UAA/OC*	Ø/UGA/OP*
6/UUG/Leu	Ø/UCG/Ser	1/UAG/AM*	4/UGG/Trp
4/CUU/Leu	3/CCU/Pro	3/CAU/His	Ø/CGU/Arg
2/CUC/Leu	3/CCC/Pro	1/CAC/His	Ø/CGC/Arg
2/CUA/Leu	3/CCA/Pro	1/CAA/Gln	1/CGA/Arg
9/CUG/Leu	Ø/CCG/Pro	12/CAG/Gln	2/CGG/Arg
2/AUU/Ile	1/ACU/Thr	2/AAU/Asn	1/AGU/Ser
3/AUC/Ile	9/ACC/Thr	7/AAC/Asn	1/AGC/Ser
5/AUA/Ile	4/ACA/Thr	7/AAA/Lys	1/AGA/Arg
5/AUG/Met	2/ACG/Thr	12/AAG/Lys	Ø/AGG/Arg
3/GUU/Val	4/GCU/Ala	5/GAU/Asp	Ø/GGU/Gly
5/GUC/Val	3/GCC/Ala	7/GAC/Asp	2/GGC/Gly
1/GUA/Val	4/GCA/Ala	6/GAA/Glu	3/GGA/Gly
4/GUG/Val	1/GCG/Ala	1Ø/GAG/Glu	3/GGG/Gly

A

QNPEPANITLGIPITNETLKWLSDKWFYMGAAFRDPVFKQAVQTIQTEYFYLTPNLINDT
IELREFQTTDDQCVYNFTHLGVQRENGTLSKCAGAVKIFAHLIVLKKHGTFMLAFNLTDE
NRGLSFYAKKPDLSPELWKIFQQAVKDVGMDESEIVFVDWTKDKCSEQQKQQLELEKETK
KETKKDPO

<187 residues, 1 stop; molecular weight: 21663.65

B

Fig. 4.7. Nucleotide and amino acid sequences of the human pre-α_1-AGP cDNA and protein. The first ATG coding methionine is the translational start. The last TAG coding for stop codon is the end of coding message. Information presented in the figure are referred and discussed in the text.

```
length: 808

  1 GCCCATAGTTTATTATAAAGGTGACTGCACCCTGCAGCCACCAGCACTGCCTGGCTCCACGTGCCTCCTGGTCTCAGT  ATG GCG CTG TCC
  1                                                                                 Met Ala Leu Ser

 91 TGG GTT CTT ACA GTC CTG AGC CTC CTA CCT CTG CTG GAA GCC CAG ATC CCA TTG TGT GCC AAC CTA GTA CCG
  5 Trp Val Leu Thr Val Leu Ser Leu Leu Pro Leu Leu Glu Ala Gln Ile Pro Leu Cys Ala Asn Leu Val Pro

163 GTG CCC ATC ACC AAC GCC ACC CTG GAC CAG ATC ACT GGC AAG TGG TTT TAT ATC GCA TCG GCC TTT CGA AAC
 29 Val Pro Ile Thr Asn Ala Thr Leu Asp Gln Ile Thr Gly Lys Trp Phe Tyr Ile Ala Ser Ala Phe Arg Asn

235 GAG GAG TAC AAT AAG TCG GTT CAG GAG ATC CAA GCA ACC TTC TTT TAC TTC ACC CCC AAC AAG ACA GAG GAC
 53 Glu Glu Tyr Asn Lys Ser Val Gln Glu Ile Gln Ala Thr Phe Phe Tyr Phe Thr Pro Asn Lys Thr Glu Asp

307 ACG ATC TTT CTC AGA GAG TAC CAG ACC CGA CAG GAC CAG TGC ATC TAT AAC ACC ACC TAC CTG AAT GTC CAG
 77 Thr Ile Phe Leu Arg Glu Tyr Gln Thr Arg Gln Asp Gln Cys Ile Tyr Asn Thr Thr Tyr Leu Asn Val Gln

379 CGG GAA AAT GGG ACC ATC TCC AGA TAC GTG GGA GGC CAA GAG CAT TTC GCT CAC TTG CTG ATC CTC AGG GAC
101 Arg Glu Asn Gly Thr Ile Ser Arg Tyr Val Gly Gly Gln Glu His Phe Ala His Leu Leu Ile Leu Arg Asp

451 ACC AAG ACC TAC ATG CTT GCT TTT GAC GTG AAC GAT GAG AAG AAC TGG GGG CTG TCT GTC TAT GCT GAC AAG
125 Thr Lys Thr Tyr Met Leu Ala Phe Asp Val Asn Asp Glu Lys Asn Trp Gly Leu Ser Val Tyr Ala Asp Lys

523 CCA GAG ACG ACC AAG GAG CAA CTG GGA GAG TTC TAC GAA GCT CTC GAC TGC TTG CGC ATT CCC AAG TCA GAT
149 Pro Glu Thr Thr Lys Glu Gln Leu Gly Glu Phe Tyr Glu Ala Leu Asp Cys Leu Arg Ile Pro Lys Ser Asp

595 GTC GTG TAC ACC GAT TGG AAA AAG GAT AAG TGT GAG CCA CTG GAG AAG CAG CAC GAG AAG GAG AGG AAA CAG
173 Val Val Tyr Thr Asp Trp Lys Lys Asp Lys Cys Glu Pro Leu Glu Lys Gln His Glu Lys Glu Arg Lys Gln

667 GAG GAG GGG GAA TCC TAG CAGGACACAGCCTTGGATCAGGACAGAGACTTGGGGCCATCCTGCCCCTCCAACCCGACATGTGTACCTCA
197 Glu Glu Gly Glu Ser AM*

GCTTTTTCCCTCACTTGCATCAATAAAGCTTCTGTGTTTGGAACAGCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3'
```

Fig. 4.8. A. Translation count of human pre-α_1-AGP. The figure description is essentially the same as Fig. 4.5.

B. Abbreviated amino acid sequences of the human pre-α_1-AGP.

5/UUU/Phe	1/UCU/Ser	3/UAU/Tyr	2/UGU/Cys
4/UUC/Phe	3/UCC/Ser	8/UAC/Tyr	2/UGC/Cys
Ø/UUA/Leu	1/UCA/Ser	Ø/UAA/OC*	Ø/UGA/OP*
3/UUG/Leu	2/UCG/Ser	1/UAG/AM*	4/UGG/Trp
2/CUU/Leu	1/CCU/Pro	1/CAU/His	Ø/CGU/Arg
4/CUC/Leu	3/CCC/Pro	2/CAC/His	1/CGC/Arg
2/CUA/Leu	3/CCA/Pro	3/CAA/Gln	2/CGA/Arg
1Ø/CUG/Leu	1/CCG/Pro	9/CAG/Gln	1/CGG/Arg
1/AUU/Ile	1/ACU/Thr	3/AAU/Asn	Ø/AGU/Ser
9/AUC/Ile	12/ACC/Thr	7/AAC/Asn	1/AGC/Ser
Ø/AUA/Ile	2/ACA/Thr	2/AAA/Lys	2/AGA/Arg
2/AUG/Met	2/ACG/Thr	12/AAG/Lys	2/AGG/Arg
2/GUU/Val	4/GCU/Ala	4/GAU/Asp	Ø/GGU/Gly
4/GUC/Val	4/GCC/Ala	7/GAC/Asp	2/GGC/Gly
1/GUA/Val	2/GCA/Ala	4/GAA/Glu	2/GGA/Gly
4/GUG/Val	1/GCG/Ala	16/GAG/Glu	3/GGG/Gly

A

MALSWVLTVLSLLPLLEAQIPLCANLVPVPITNATLDQITGKWFYIASAFRNEEYNKSVQ
EIQATFFYFTPNKTEDTIFLREYQTRQDQCIYNTTYLNVQRENGTISRYVGGQEHFAHLL
ILRDTKTYMLAFDVNDEKNWGLSVVYADKPETTKEQLGEFYEALDCLRIPKSDVVYTDWKK
DKCEPLEKQHEKERKQEEGES

<2Ø1 residues, Ø stop; molecular weight: 23514.75

B

Fig. 4.9. A. Translation count of mature human α_1-AGP. The figure description is essentially the same as Fig. 4.5.

B. Abbreviated amino acid sequences of the mature human AGP.

2/UUU/Phe	6/UCU/Ser	Ø/UAU/Tyr	3/UGU/Cys
4/UUC/Phe	4/UCC/Ser	1/UAC/Tyr	4/UGC/Cys
Ø/UUA/Leu	5/UCA/Ser	Ø/UAA/OC*	3/UGA/OP*
2/UUG/Leu	4/UCG/Ser	1/UAG/AM*	4/UGG/Trp
1/CUU/Leu	5/CCU/Pro	2/CAU/His	Ø/CGU/Arg
2/CUC/Leu	3/CCC/Pro	1/CAC/His	Ø/CGC/Arg
Ø/CUA/Leu	11/CCA/Pro	1/CAA/Gln	2/CGA/Arg
3/CUG/Leu	1/CCG/Pro	3/CAG/Gln	3/CGG/Arg
2/AUU/Ile	4/ACU/Thr	1/AAU/Asn	8/AGU/Ser
Ø/AUC/Ile	5/ACC/Thr	2/AAC/Asn	7/AGC/Ser
4/AUA/Ile	8/ACA/Thr	2/AAA/Lys	12/AGA/Arg
5/AUG/Met	7/ACG/Thr	3/AAG/Lys	1Ø/AGG/Arg
Ø/GUU/Val	1/GCU/Ala	1/GAU/Asp	1/GGU/Gly
Ø/GUC/Val	1/GCC/Ala	1/GAC/Asp	1/GGC/Gly
Ø/GUA/Val	3/GCA/Ala	1/GAA/Glu	3/GGA/Gly
2/GUG/Val	Ø/GCG/Ala	3/GAG/Glu	4/GGG/Gly

A

QIPLCANLVPVPITNATLDQITGKWFYIASAFRNEEYNKSVQEIQATFFYFTPNKTEDTI
FLREYQTRQDQCIYNTTYLNVQRENGTISRYVGGQEHFAHLLILRDTKTYMLAFDVNDEK
NWGLSVYADKPETTKEQLGEFYEALDCLRIPKSDVVYTDWKKDKCEPLEKQHEKERKQEE
GESO

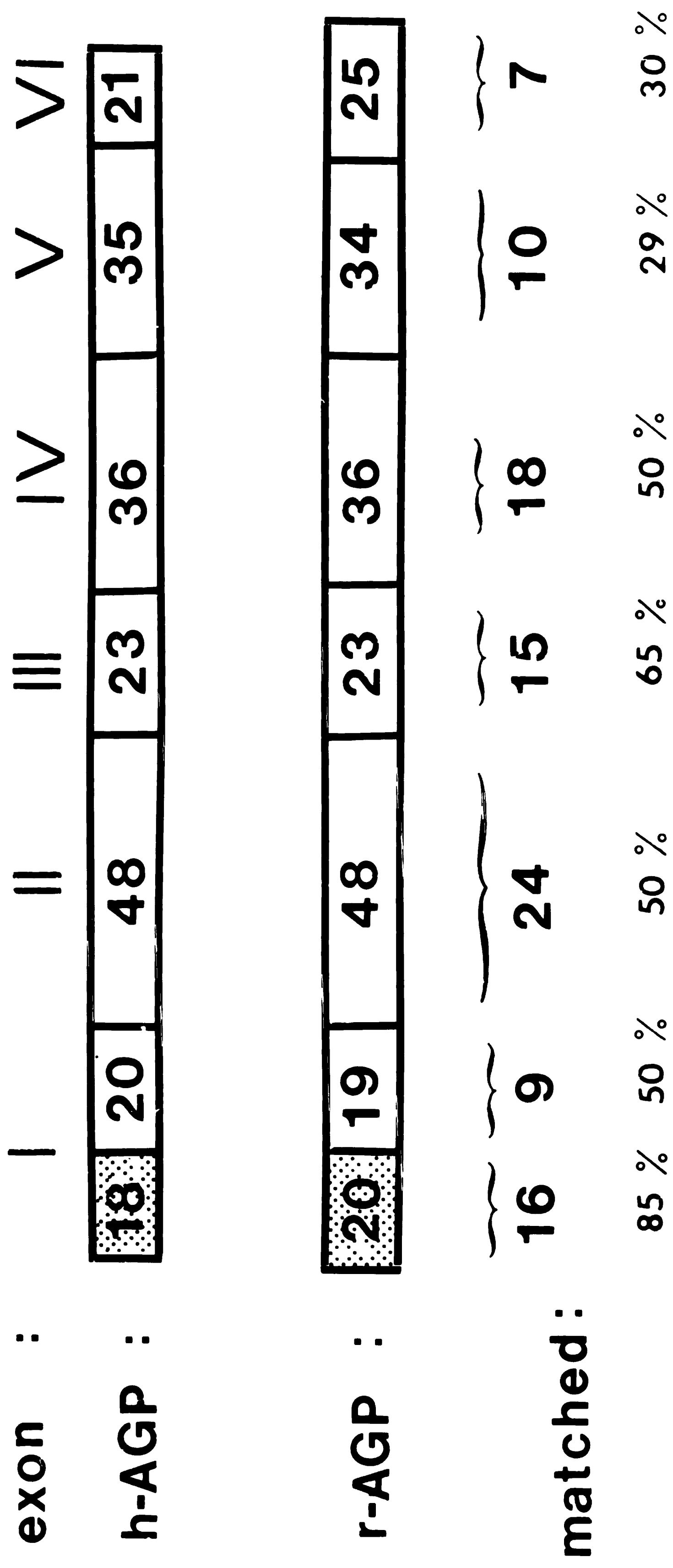
<183 residues, 1 stop; molecular weight: 21563.Ø1

B

Fig. 4.10. A comparison of amino acid residues of human and rat pre-α_1-AGP. The numbers represent the amino acids in relationship to translational start of methionine which is number 1. The first line of the 3-line assembly set is the amino acid abbreviation of human protein, the second is the rat protein and the third is the concensus sequences. Unmatched amino acids are represented with a broken line. The matched amino acids are emphasized by a dark dot beneath the amino acid abbreviation. The partially boxed amino acids of the first nineteen residues are signal sequences. Discussion derived from this figure is in Section 4.4.

```
Protein
homology = 154
                   1        1Ø        2Ø        3Ø        4Ø        5Ø
ep1.hPRE-AGP    1  MALSWVLTVLSLLPLLEAQIPLCANLVPV-PITNATLDQITGKWFYIASA
ep1.rPRE-AGP    1  MALHMVLVVLSLLPLLEAQNPEPANITLGIPITNETLKWLSDKWFYMGAA
consensus       1  MAL--VL-VLSLLPLLEAQ-P--AN-----PITN-TL-----KWFY---A
                   •••  •• ••••••••••••• •  ••     •••• ••     ••••   •
                   51       6Ø        7Ø        8Ø        9Ø        1ØØ
ep1.hPRE-AGP   51  FRNEEYNKSVQEIQATFFYFTPNKTEDTIFLREYQTRQDQCIYNTTYLNV
ep1.rPRE-AGP   51  FRDPVFKQAVQTIQTEYFYLTPNLINDTIELREFQTTDDQCVYNFTHLGV
consensus      51  FR-------VQ-IQ---FY-TPN---DTI-LRE-QT--DQC-YN-T-L-V
                   ••       •• ••   •• •••   ••• ••• ••  ••• •• • • •
                   1Ø1      11Ø       12Ø       13Ø       14Ø       15Ø
ep1.hPRE-AGP  1Ø1  QRENGTISRYVGGQEHFAHLLILRDTKTYMLAFDVNDEKNWGLSVVYADKP
ep1.rPRE-AGP  1Ø1  QRENGTLSKCAGAVKIFAHLIVLKKHGTFMLAFNLTDENR-GLSFYAKKP
consensus     1Ø1  QRENGT-S---G----FAHL--L----T-MLAF---DE---GLS-YA-KP
                   •••••• •   •    ••••  •    • ••••   ••   ••• •• ••
                   151      16Ø       17Ø       18Ø       19Ø       2ØØ
ep1.hPRE-AGP  151  ETTKEQLGEFYEALDCLRIPKSDVVYTDWKKDKCEPLEKQHEKERKQEEG
ep1.rPRE-AGP  151  DLSPELWKIFQQAVKDVGMDESEIVFVDWTKDKCSEQQKQQLELEKETKK
consensus     151  ----E----F--A--------S--V--DW-KDKC----KQ-----K----
                       •    •  •        •  •  •• ••••    ••     •
                   2Ø1      21Ø       22Ø       23Ø       24Ø       25Ø
ep1.hPRE-AGP  2Ø1  ES
ep1.rPRE-AGP  2Ø1  ETKKDP
consensus     2Ø1  E-----
                   •
```

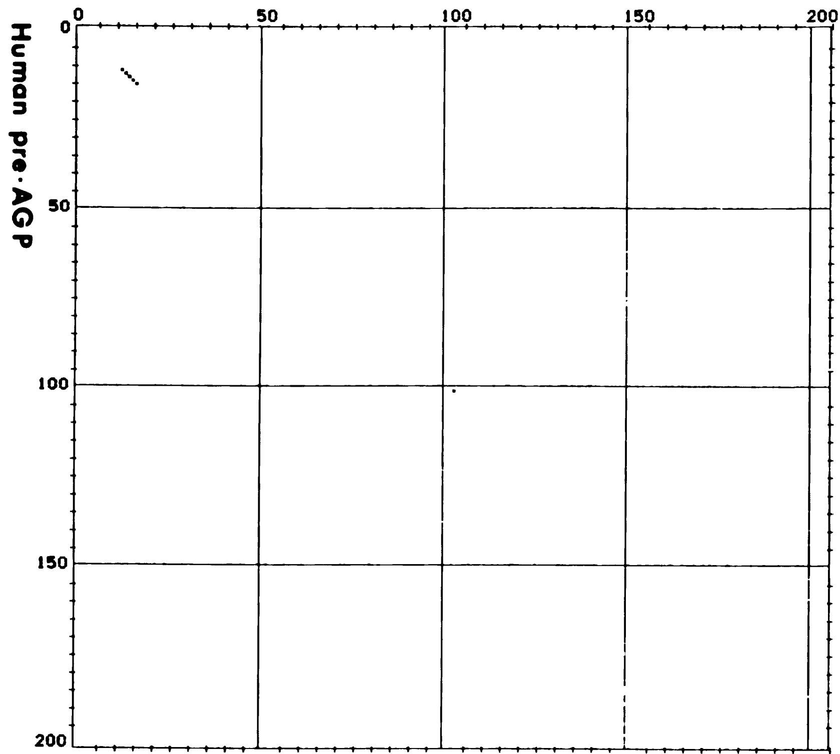
Fig. 4.11. The amino acid comparison between human and rat pre α_1-AGP classified into exonic segments. The Roman numerals represent the exon number corresponding to the underneath Arabic numerals which represent number of amino acid coded within the specific axon. The shadowed areas are leader signals. Number of matched amino acid residues are listed under each exon and the percentages of matching are listed further below. The significance of homology shown on this figure are discussed in Section 4.4.



exon : I II III IV V VI
h-AGP : 18 20 48 23 36 35 21
r-AGP : 20 19 48 23 36 34 25
matched: 16 9 24 15 18 10 7
85 % 50 % 50 % 65 % 50 % 29 % 30 %

Fig. 4.12. A. A homology plot of the amino acid residues of the rat and human pre-α_1-AGP. The x axis represent the rat protein, starting from position 1 (amino end) toward position 205 (carboxyl end). The y axis represents human protein, starting from position 1 (top right) toward position 201 (bottom left) which is carboxyl end. The plotting stringency is set at 7:7, which means window size is 7 amino acid, matching requirement is 7 amino acids. This is one of the highest stringency, under this condition; it has to have a complete and continuous match of 7 amino acids in order to generate a scoring point represented by a dot on the plotting matrix.

B. Same as described above, except plotting stringency is reduced slightly from 7:7 to 6:7. Each dot on the matrix represents a match of 6 amino acids within a stretch of 7 amino acids.


Rat Pre-AGP
7 : 7
Human pre-AGP
0
50
100
150
200
A

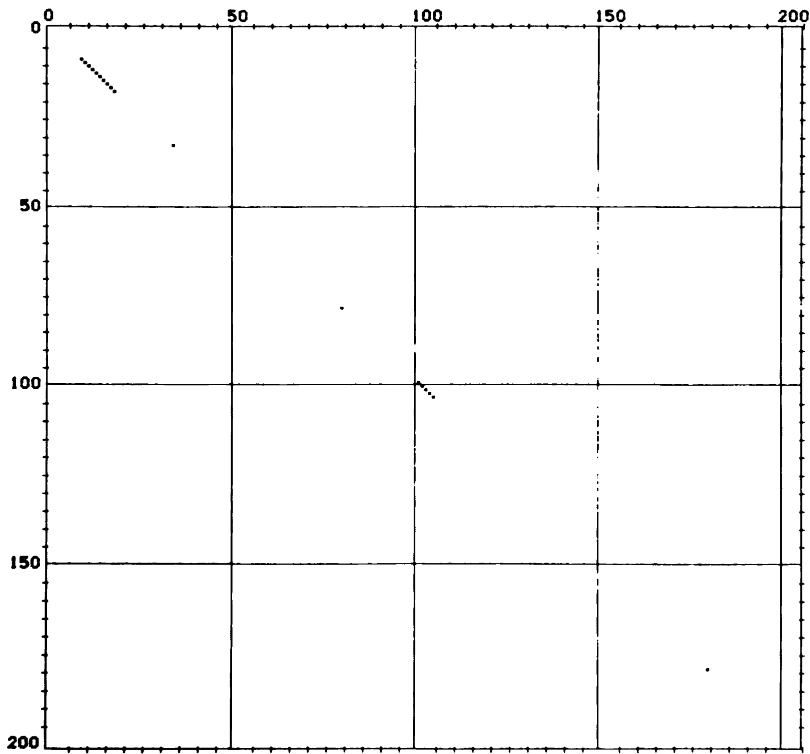
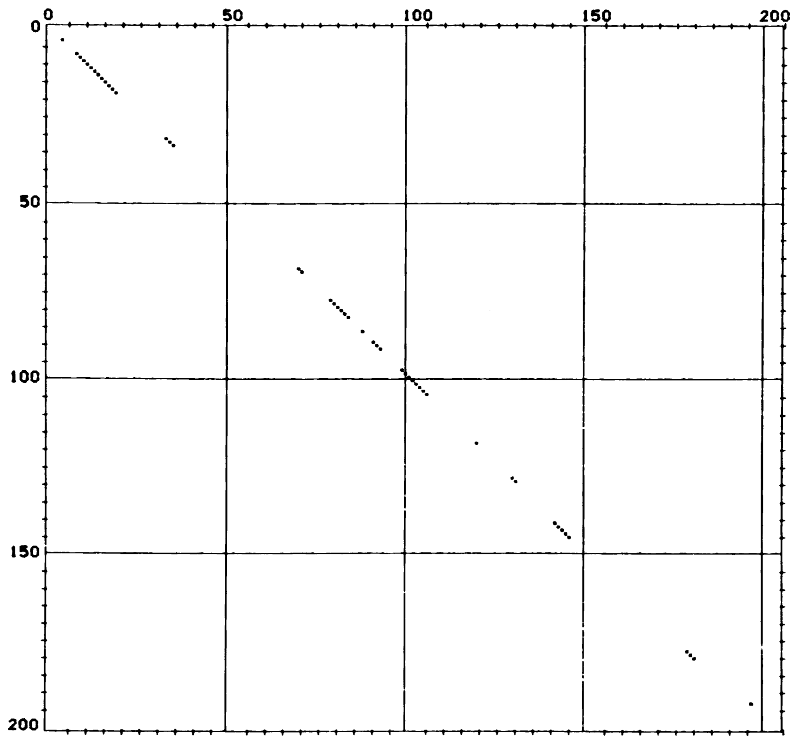
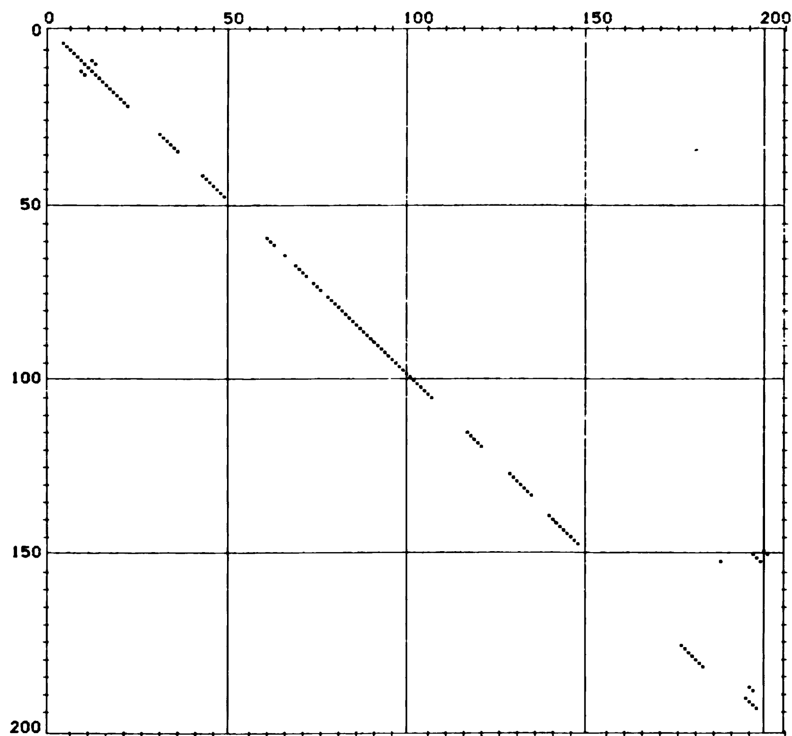

6 : 7
0
50
100
150
200
B

Fig. 4.13. C. Same as Fig. 4.8, except that the stringency is reduced to 5:7. Each scoring dot represents a continuous match of 5 amino acids within a 7 amino acid stretch.

D. Same as above, except that the plotting stringency is reduced to 4:7; four amino acids have to match between human and rat protein sequences in order to generate a printed dot on the matrix.



C

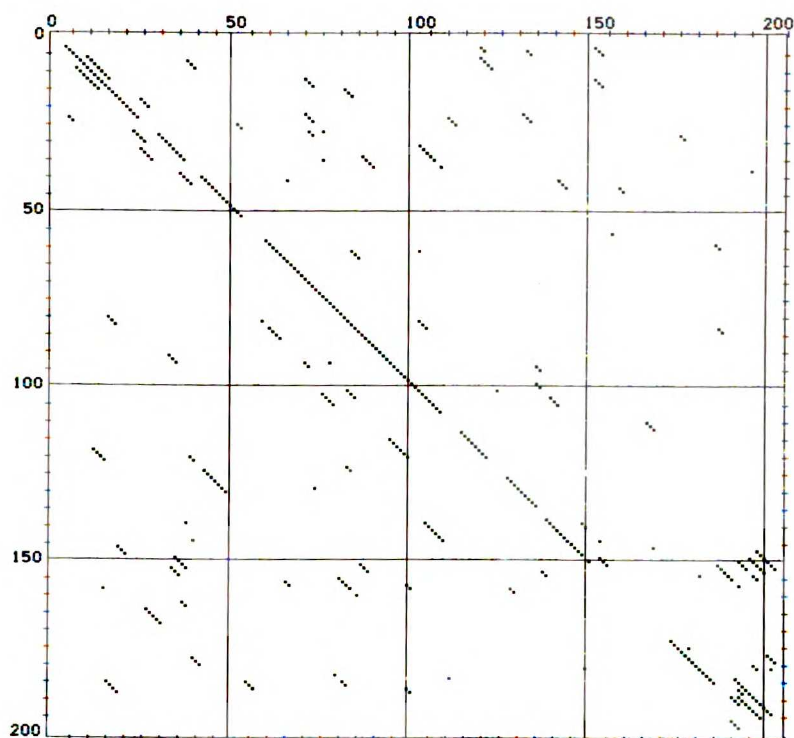


D

Fig. 4.14. E. The plotting stringency is 3:7. Under this condition, each dot represents a 3-amino-acid match within a stretch of 7 amino acids.

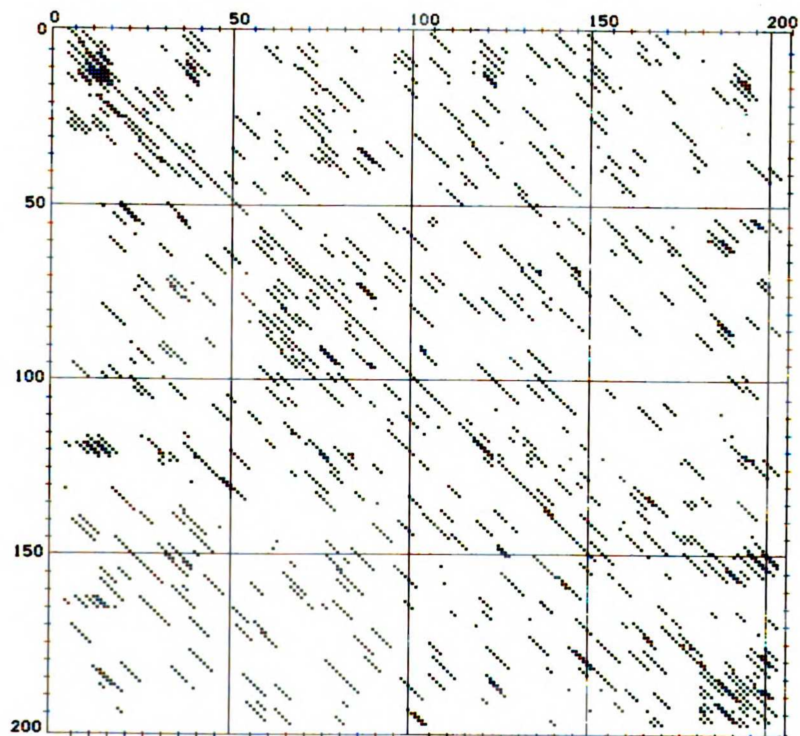
F. Same as above, except the plotting stringency is now 2:7.

3 : 7



E

2 : 7



F

Fig. 4.15. Comparisons of human (H) and rat (R) α_1-AGP glycosylation sites. The large numbers indicate the number of glycosylation sites on the protein. The small numbers indicate the positions of amino acid residues. The broken lines are variable distances between glycosylated tripeptides. The triplet codons coding asparagine which is the target residue for glycosylation were also listed for comparison.

```
     15            38             54             75             85
H. ---Asn Ala Thr---Asn Lys Ser---Asn Lys Thr---Asn Thr Thr---Asn Gly Thr ---Asp Val Asn---
      AAC           AAT            AAC            AAC            AAT            GAT

      AAC           AAT            AAC            AAC            AAT            AAC
R. ---Asn Ile Thr---Asn Glu Thr---Asn Asp Thr---Asn Phe Thr---Asn Gly Thr ---Asn Leu Thr---
      7             16             58             76             86             116
```

Fig. 4.16. Nucleotide sequence homology between the rat and human α_1-AGP cDNA. In order to achieve maximum homology, the cDNA's of both rat and human α_1-AGP are stretched out with broken lines occasionally. The human cDNA sequence is the published results of Luciana et al. (1985) and the 5' end sequences of the first 48 bases are probably derived from a genomic clone, not really belonging to the mRNA. The matched concensus sequences are listed on the third line and a heavy broken line is marked beneath it for easier visualization. The triple asterisks above codon ATG indicate the methionine as translational start, the second triple asterisks represent the coding triplet AGA for glutamine which is the first amino acid of mature α_1-AGP in both human and rat.

```
homology = 267
                      1        10        20        30        40        50
ess.hAGPcDNA      1  GCCCATAGTTTATTATAAAGGTGACTGCACCCTGCAGCCACCAGCACTGC
ess.rAGPcDNA    -47                                                5' GC
consensus         1                                                   GC
                                                         ***
                      51        60        70        80    Met  90        100
ess.hAGPcDNA     51  -CTGGCTCCA-C-GTGCCTCCTGG--TCTCAGT-ATGGCGCTGTCCTGGG
ess.rAGPcDNA      3  TCTTCCTGGGCCGGTGCCTCTGAGTGTCTTCGGCATGGCGCTGCACATGG
consensus        51  -CT--CT----C-GTGCCTC---G--TCT--G--ATGGCGCTG--C--GG
                                                          ***
                      101       110       120       130       Gln       150
ess.hAGPcDNA    101  TTCTTACAGTCCTGAGCCTCCTACCTCTGCTGGAAGCCCAGATCCCATTG
ess.rAGPcDNA     53  TTCTTGTCGTTTTGAGCCTCCTGCCCTTGTTGGAAGCTCAGAACCCAGAA
consensus       101  TTCTT---GT--TGAGCCTCCT-CC--TG-TGGAAGC-CAGA-CCCA---

                      151       160       170       180       190       200
ess.hAGPcDNA    151  TGTGCCAA---CC-TAGTACCGGTGCCCATCACCAACGCCACCCTGGACC
ess.rAGPcDNA    103  CCTGCCAACAT-CACCCTAGGCATACCTATTACCAATGAGACCCTGAAAT
consensus       151  --TGCCAA----C----TA----T-CC-AT-ACCAA-G--ACCCTG-A--

                      201       210       220       230       240       250
ess.hAGPcDNA    201  AGATCACTGGCAAGTGGTTTTATATCGCATCGGCCTTTCGAAACGAGGAG
ess.rAGPcDNA    153  GGCTCTCAGACAAATGGTTTTACATGGGAGCAGCTTTCCGAGACCCCGTG
consensus       201  -G-TC-C-G-CAA-TGGTTTTA-AT-G-A-C-GC-TT-CGA-AC---G-G

                      251       260       270       280       290       300
ess.hAGPcDNA    251  TACAATAAGTCGGTTCAGGAGATCCAAGCAACCTTCTTTTACTTCACCCC
ess.rAGPcDNA    203  TTCAAGCAGGCAGTTCAAACGATACAGACGGAATATTTTTACCTTACCCC
consensus       251  T-CAA--AG-C-GTTCA---GAT-CA--C----T--TTTTAC-T-ACCCC

                      301       310       320       330       340       350
ess.hAGPcDNA    301  CAACAAGACAGAGGACACGATCTTTCTCAGAGAGTACCAGACCCGACAGG
ess.rAGPcDNA    253  CAACTTGATAAACGACACAATTGAACTTCGGGAGTTTCAGACCACAGACG
consensus       301  CAAC--GA-A-A-GACAC-AT----CT--G-GAGT--CAGACC--A-A-G

                      351       360       370       380       390       400
ess.hAGPcDNA    351  ACCAGTGCATCTATAACACCACCTACCTGAATGTCCAGCGGGAAAATGGG
ess.rAGPcDNA    303  ACCAGTGTGTCTATAACTTCACCCATCTAGGAGTCCAGAGAGAGAATGGG
consensus       351  ACCAGTG--TCTATAAC--CACC-A-CT----GTCCAG-G-GA-AATGGG

                      401       410       420       430       440       450
ess.hAGPcDNA    401  ACCATCTCCAGATACGTGGGAGGCCAA-GAGCATTTCG-CT-CACTTGCT
ess.rAGPcDNA    353  ACCTTATCCAAGTGTGCAGGAG--CAGTGAAAATCTTTGC-CCATTTGAT
consensus       401  ACC-T-TCCA--T--G--GGAG--CA--GA--AT-T---C--CA-TTG-T

                      451       460       470       480       490       500
ess.hAGPcDNA    451  GATCCTCAGGGACACCAA--GACCTACATGCTTGCTTTTGACGTGAACGA
ess.rAGPcDNA    403  AGTGCTGAAGAAA-C-ATGGGACCTTCATGCTTGCCTTTAACCTGACAGA
consensus       451  --T-CT-A-G-A--C-A---GACCT-CATGCTTGC-TTT-AC-TGA--GA

                      501       510       520       530       540       550
ess.hAGPcDNA    501  TGAGAAGAACTGGGGGCTGTCTGTCTATGCTGACAAGCCAGAGACGACCA
ess.rAGPcDNA    453  TGAGA--ACCGGGGG-CTGTCCTTCTACGCTAAAAAGCCAGACTTGTCCC
consensus       501  TGAGA--A-C-GGGG-CTGTC--TCTA-GCT-A-AAGCCAGA---G-CC-

                      551       560       570       580       590       600
ess.hAGPcDNA    551  AGGAGCAACTGGG-AGAG-T-TCTACGAAGCTCTCGACTGCTTGCGCATT
ess.rAGPcDNA    503  CAGAGCTGTGGA-AA-ATA-TTCCAGCAGGCTGTCAAAGATGTGGGCATG
consensus       551  --GAGC----G---A-A----TC-A--A-GCT-TC-A-----TG-GCAT-

                      601       610       620       630       640       650
ess.hAGPcDNA    601  CCCAAGTCAGATGTCGTGTACACCGATTGGAAAAAGGATAAGT---GTGA
ess.rAGPcDNA    553  GATGAATCAGAAATCGTATTTGTCGACTGGACAAAGGATAAGTGCAGTGA
consensus       601  ----A-TCAGA--TCGT-T----CGA-TGGA-AAAGGATAAGT---GTGA

                      651       660       670       680       690       700
ess.hAGPcDNA    651  GCCACTGGAGAAGCAGCACGAGAAGGAGAGGAAA-CAG-G-AGGA-----
ess.rAGPcDNA    603  GCAGCAGAAGCAGCAGCTGGAGCTGGAGAAGGAGACTAAGAAGGAGACCA
consensus       651  GC--C-G-AG-AGCAGC--GAG--GGAGA-G-A--C---G-AGGA-----

                      701       710       720       730       740       750
ess.hAGPcDNA    701  GGGGGAATCC------T-AGCAGGACA-CAGCCTTGGATCAGGACAGA-G
ess.rAGPcDNA    653  AGAAGGATCCTTAGGCCAAGCATGAACTCAGC--TCTCTGAACTCCGGGG
consensus       701  -G--G-ATCC--------AGCA-GA---CAGC--T---T-A---C-G--G

                      751       760       770       780       790       800
ess.hAGPcDNA    751  ACTTGGGGCCATCCTGCCCCTCCAACCCGACATGTGTACCTCAGCTTTTT
ess.rAGPcDNA    703  ACTGTCCCC-ATGCCCACTCTACCCCCACTCCTGTGCACCTCGATTCTAT
consensus       751  ACT-----C-AT-C---C-CT-C--CC---C-TGTG-ACCTC---T-T-T

                      801       810       820       830       840
ess.hAGPcDNA    801  CCCT-CACTTGC-ATCAATAAAGCTTCTGTGTTTGGAACAGCTAAAAAAAAAAAAAAA 3'
ess.rAGPcDNA    753  TTT-CCA----CAATAA-TAAAGGTTTGGTTAAACAATCAAAAAAAAAAAAAAAAAAA 3'
consensus       801  -----CA----C-AT-A-TAAAG-TT--GT------A-CA---
```

Fig. 4.17. A homology plot of nucleotide sequences of human and rat α_1-AGP cDNA. The x axis represents the sequence of the human cDNA starting from left (5' end) toward right (3' end). The number indicates the length in base nucleotides. The y axis represents the rat cDNA starting from the top (5' end) toward bottom (3' end). Plotting stringency is 4:7, under this condition each printing dot on the matrix represents a match of four nucleotides within a measuring window of 7 nucleotides.

4 : 7
0
200
400
600
800
0
200
400
600

Fig. 4.18. Same as described in Fig. 4.12, except with increased matching stringency. The 5:7 stringency indicates that each dot on the matrix represents a 5-nucleotide match within a measuring window of 7 nucleotides.

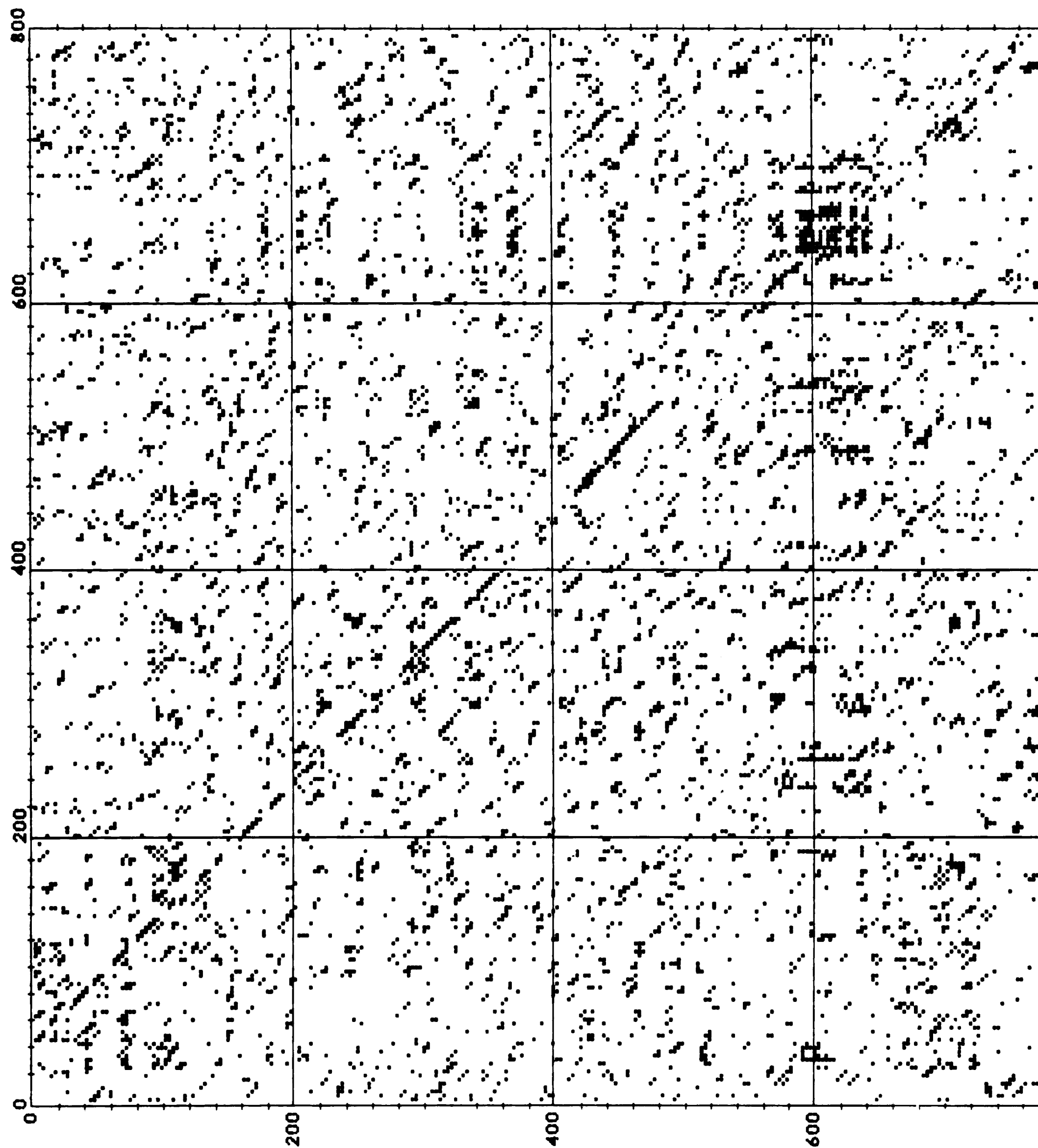

5 : 7
0
200
400
600
800
0
200
400
600

Fig. 4.19. Essentially the same as described in previous figures. The plotting requirement increases to 6:7. Under this stringency the background "noise" is reduced to a lower level. Each dot represents a match of 6 nucleotides in a 7-nucleotide stretch of comparing sequences.

6 : 7

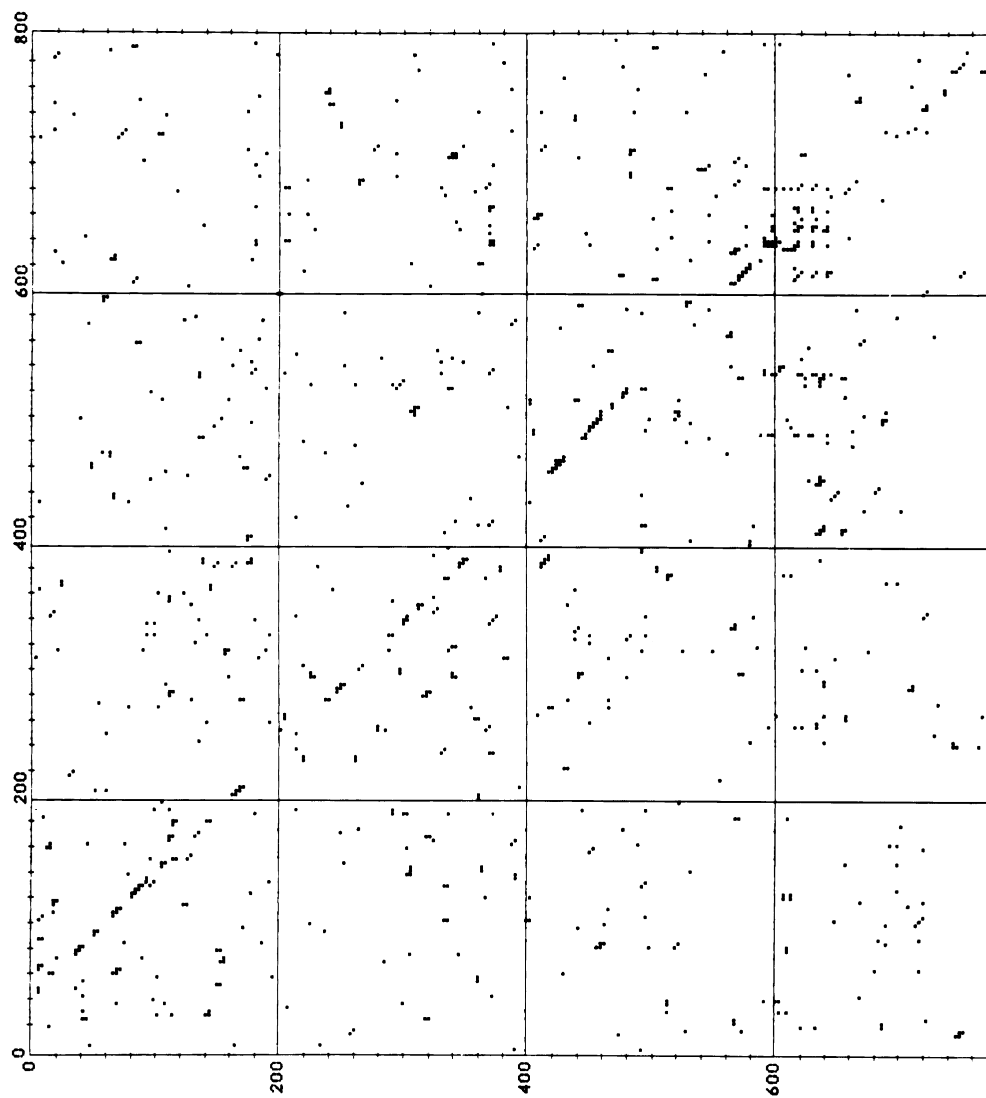


Fig. 4.20. Same as previously described, however, the stringency has been increased to the highest level. Not only the background "noise" but also the authentic matching are diminishing. It has to have a 7 out of 7 nucleotide-match in order to generate a printing dot on the homology matrix. The purpose and significance of this series of comparisons is to illustrate the concensus regions visually. The results are discussed in Section 4.4.

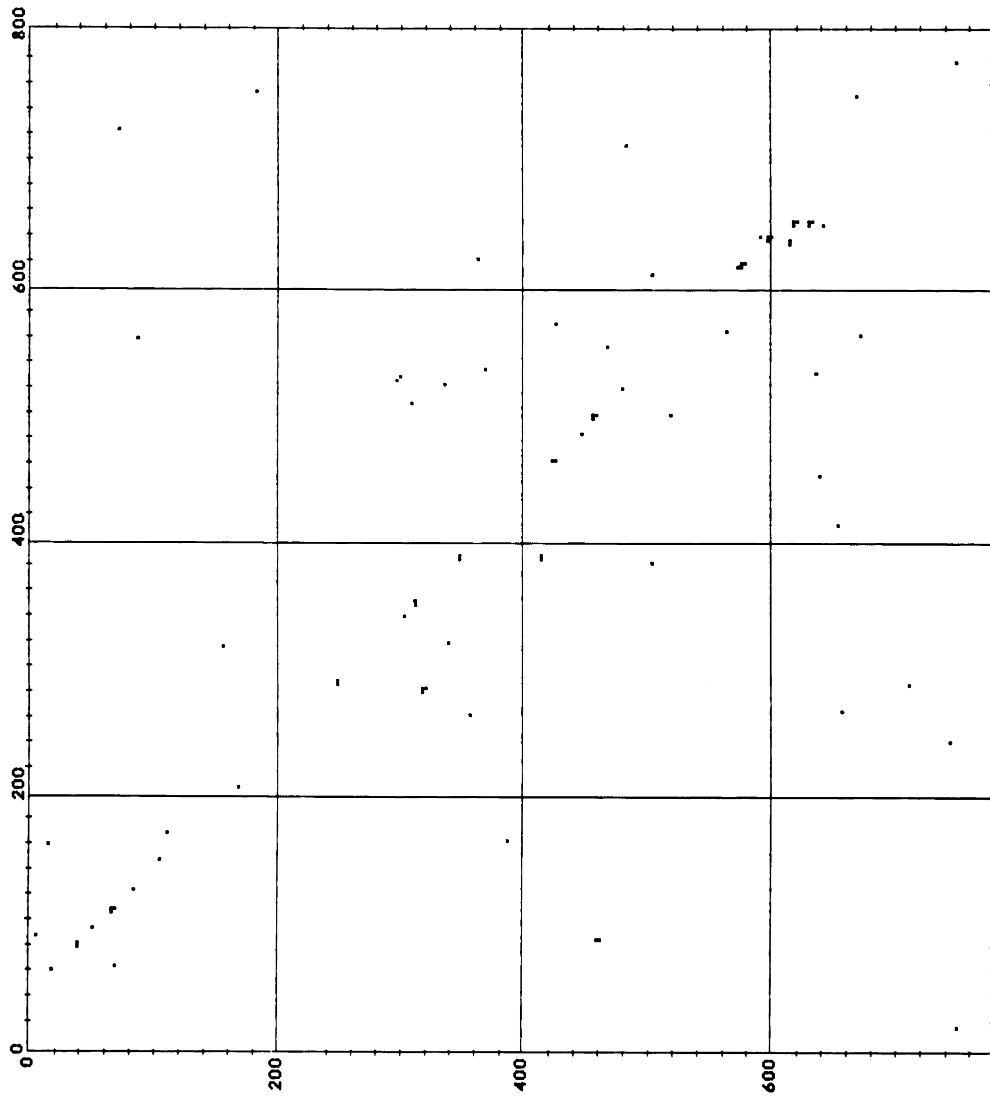

7 : 7
0
200
400
600
800
0
200
400
600

Fig. 4.21. A comparison of human and rat α_1-AGP gene based on sizes of exons and introns. The Roman numerals represent the number of exons. The larger numbers indicate the base pairs of individual introns, the smaller number represents the number of base pairs for exons.

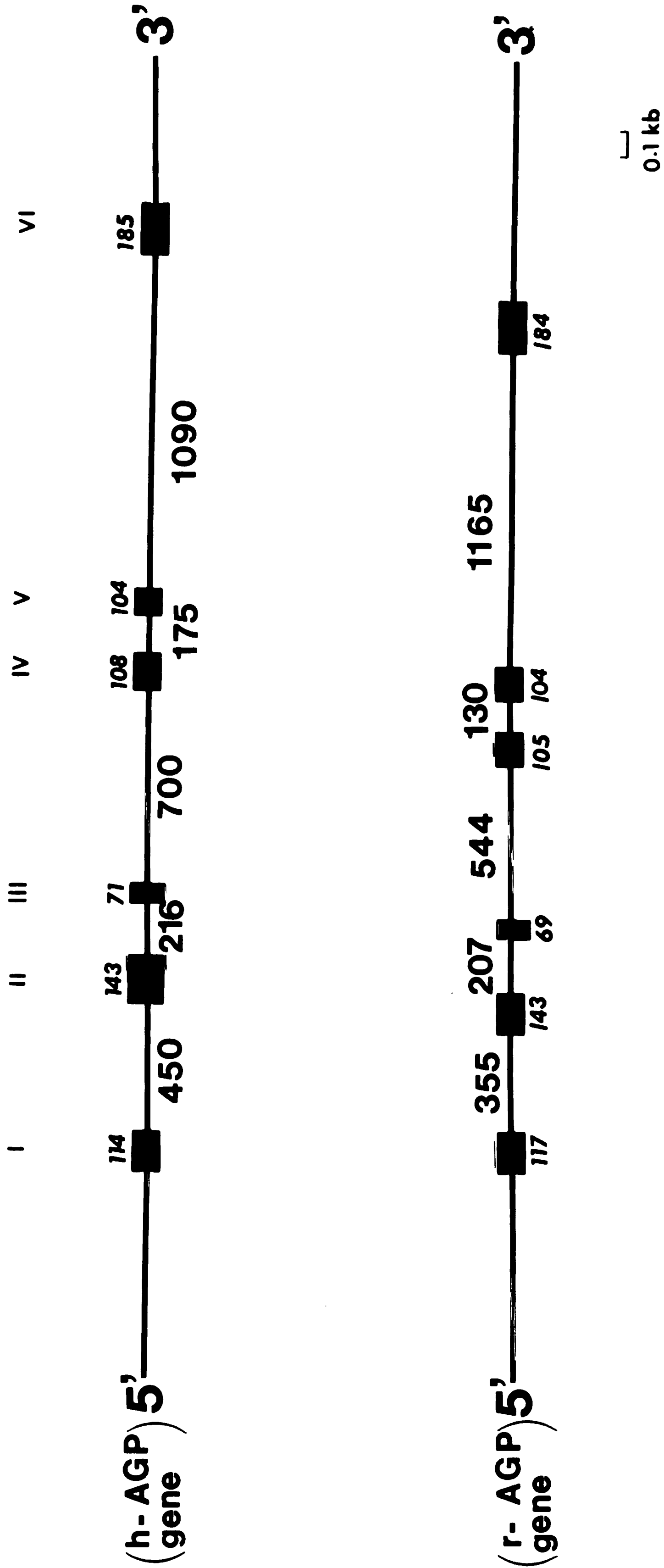

(h- AGP gene)
5'
3'
I
II
III
IV
V
VI
114
143
71
108
104
185
450
216
700
175
1090
(r- AGP gene)
5'
3'
117
143
69
105
104
184
355
207
544
130
1165
0.1 kb

Fig. 4.22. A comparison of splicing junctions of human and rat α_1-AGP gene. The Roman numerals indicate the numbers of introns. Intron I is designated to be the intron existing between exon I and exon II. Intron II is the intron located between exon II and exon III, etc. The small dots represent unspecified lengths of nucleotides connecting the donor and acceptor sides of introns. The large dots show the concensus sequences of introns between human and rat α_1-AGP gene.

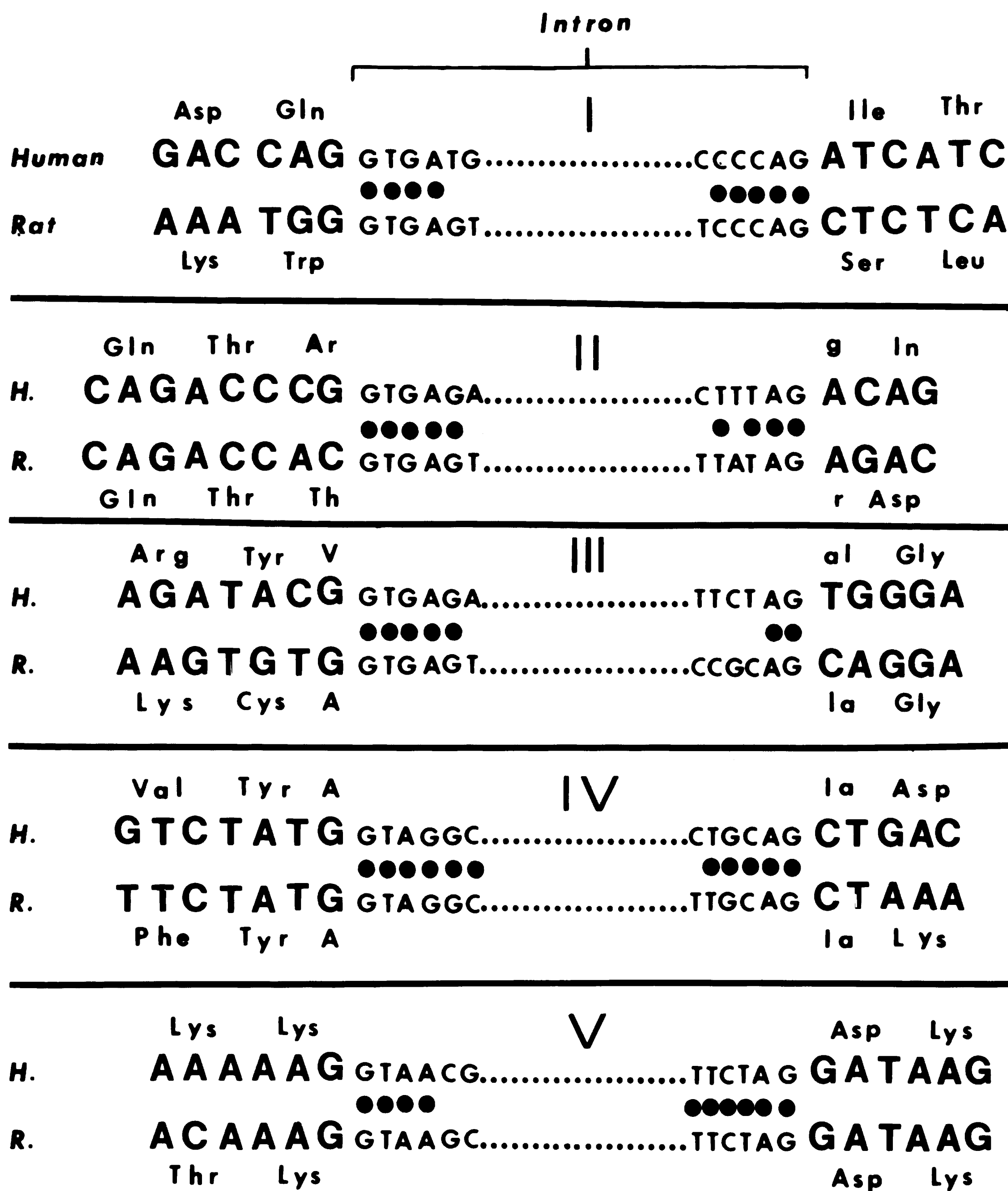

Intron
Asp Gln I Ile Thr
Human GAC CAG GTGATG....................CCCCAG ATCATC
Rat AAA TGG GTGAGT....................TCCCAG CTCTCA
Lys Trp Ser Leu
Gln Thr Ar II g ln
H. CAGACCCG GTGAGA....................CTTTAG ACAG
R. CAGACCAC GTGAGT....................TTATAG AGAC
Gln Thr Th r Asp
Arg Tyr V III al Gly
H. AGATACG GTGAGA....................TTCTAG TGGGA
R. AAGTGTG GTGAGT....................CCGCAG CAGGA
Lys Cys A la Gly
Val Tyr A IV la Asp
H. GTCTATG GTAGGC....................CTGCAG CTGAC
R. TTCTATG GTAGGC....................TTGCAG CTAAA
Phe Tyr A la Lys
Lys Lys V Asp Lys
H. AAAAAG GTAACG....................TTCTAG GATAAG
R. ACAAAG GTAAGC....................TTCTAG GATAAG
Thr Lys Asp Lys

Fig. 4.23. A hydropathy representation of the mature human α_1-AGP. The detailed software program is described by Kyte et al. (in press). In short, the number of jump means the number of amino acids used as a unit for hydropathy scanning. The number of width represents the degree of hydrophobicity (+) or hydrophilicity (-), in this case, 4 is the maximum degree of hydropathy (any scanned unit that exhibits hydropathy greater than 4 is still regarded as 4). The numbers in the far left column represent the number of amino acid residues being scanned at each jump. Abbreviated amino acid residues are shown in the second column. Degrees of hydropathy of each jump are listed in the third column. The hydropathy results of a protein molecule is also visualized by a vertical broken line of jump backbone and intensities of hydropathy (asterisks).

```
Number of amino acids: 184
Average hydropathy: -0.78

Source: Kyte,J. & Doolittle,R.F., in press
Notes:  The dotted line at -0.4 on the graph corresponds to the overall average
        hydropathy for proteins investigated by Kyte and Doolittle.
Jump: 4       Width: 4

region     seq.         hydropathy  -4   -3   -2   -1   0    1    2    3    4

   1-4     QIPL              0.78   :    :    :    :  . |****:    :    :    :
   5-8     CANL              1.13   :    :    :    :  . |******   :    :    :
   9-12    VPVP              1.30   :    :    :    :  . |*******  :    :    :
  13-16    ITNA              0.52   :    :    :    :  . |*** :    :    :    :
  17-20    TLDQ             -1.00   :    :    :    *****|    :    :    :    :
  21-24    ITGK             -0.13   :    :    :    :  .*|    :    :    :    :
  25-28    WFYI              1.25   :    :    :    :  . |******   :    :    :
  29-32    ASAF              1.35   :    :    :    :  . |*******  :    :    :
  33-36    RNEE             -3.75   :*******************|    :    :    :    :
  37-40    YNKS             -2.40   :    :  ************|    :    :    :    :
  41-44    VQEI              0.42   :    :    :    :  . |**  :    :    :    :
  45-48    QATF              0.07   :    :    :    :  . |    :    :    :    :
  49-52    FYFT              0.85   :    :    :    :  . |****:    :    :    :
  53-56    PNKT             -2.42   :    :  ************|    :    :    :    :
  57-60    EDTI             -0.80   :    :    :    :****|    :    :    :    :
  61-64    FLRE             -0.40   :    :    :    :  **|    :    :    :    :
  65-68    YQTR             -2.50   :    : *************|    :    :    :    :
  69-72    QDQC             -2.00   :    :    **********|    :    :    :    :
  73-76    IYNT             -0.25   :    :    :    :  .*|    :    :    :    :
  77-80    TYLN             -0.45   :    :    :    :  **|    :    :    :    :
  81-84    VQRE             -1.83   :    :    :*********|    :    :    :    :
  85-88    NGTI             -0.02   :    :    :    :  . |    :    :    :    :
  89-92    SRYV             -0.63   :    :    :    : ***|    :    :    :    :
  93-96    GGQE             -1.95   :    :    **********|    :    :    :    :
  97-100   HFAH             -0.48   :    :    :    :  **|    :    :    :    :
 101-104   LLIL              3.90   :    :    :    :  . |********************
 105-108   RDTK             -3.15   :   ****************|    :    :    :    :
 109-112   TYML              0.90   :    :    :    :  . |*****    :    :    :
 113-116   AFDV              1.30   :    :    :    :  . |*******  :    :    :
 117-120   NDEK             -3.60   : ******************|    :    :    :    :
 121-124   NWGL             -0.28   :    :    :    :  .*|    :    :    :    :
 125-128   SVYA              0.95   :    :    :    :  . |*****    :    :    :
 129-132   DKPE             -3.13   :   ****************|    :    :    :    :
 133-136   TTKE             -2.20   :    :   ***********|    :    :    :    :
 137-140   QLGE             -0.93   :    :    :    *****|    :    :    :    :
 141-144   FYEA             -0.07   :    :    :    :  . |    :    :    :    :
 145-148   LDCL              1.60   :    :    :    :  . |******** :    :    :
 149-152   RIPK             -1.38   :    :    :  *******|    :    :    :    :
 153-156   SDVV              1.00   :    :    :    :  . |*****    :    :    :
 157-160   YTDW             -1.60   :    :    : ********|    :    :    :    :
 161-164   KKDK             -3.80   :*******************|    :    :    :    :
 165-168   CEPL              0.27   :    :    :    :  . |*   :    :    :    :
 169-172   EKQH             -3.53   : ******************|    :    :    :    :
 173-176   EKER             -3.85   :*******************|    :    :    :    :
 177-180   KQEE             -3.60   : ******************|    :    :    :    :
 181-184   GESO             -1.20   :    :    :   ******|    :    :    :    :
```

Fig. 4.24. A hydropathy figure of the mature rat α_1-AGP. Figure legend is essentially as same as described in Fig. 4.23.

```
Number of amino acids: 188
Average hydropathy: -0.54

Source: Kyte,J. & Doolittle,R.F., in press
Notes:  The dotted line at -0.4 on the graph corresponds to the overall average
        hydropathy for proteins investigated by Kyte and Doolittle.
Jump: 4      Width: 4

region     seq.        hydropathy  -4   -3   -2   -1    0    1    2    3    4

    1-4    QNPE          -3.03     :    ***************|    :    :    :    :
    5-8    PANI           0.30     :    :    :    :  . |*   :    :    :    :
   9-12    TLGI           1.77     :    :    :    :  . |*********:    :    :
  13-16    PITN          -0.32     :    :    :    :  **|    :    :    :    :
  17-20    ETLK          -1.10     :    :    :   ******|    :    :    :    :
  21-24    WLSD          -0.40     :    :    :    :  **|    :    :    :    :
  25-28    KWFY          -0.85     :    :    :    :****|    :    :    :    :
  29-32    MGAA           1.27     :    :    :    :  . |******   :    :    :
  33-36    FRDP          -1.73     :    :    :*********|    :    :    :    :
  37-40    VFKQ          -0.13     :    :    :    :  .*|    :    :    :    :
  41-44    AVQT           0.45     :    :    :    :  . |**  :    :    :    :
  45-48    IQTE          -0.80     :    :    :    :****|    :    :    :    :
  49-52    YFYL           0.95     :    :    :    :  . |*****    :    :    :
  53-56    TPNL          -0.53     :    :    :    : ***|    :    :    :    :
  57-60    INDT          -0.80     :    :    :    :****|    :    :    :    :
  61-64    IELR           0.05     :    :    :    :  . |    :    :    :    :
  65-68    EFQT          -1.25     :    :    :   ******|    :    :    :    :
  69-72    TDDQ          -2.80     :    :**************|    :    :    :    :
  73-76    CVYN           0.48     :    :    :    :  . |**  :    :    :    :
  77-80    FTHL           0.63     :    :    :    :  . |*** :    :    :    :
  81-84    GVQR          -1.05     :    :    :    *****|    :    :    :    :
  85-88    ENGT          -2.03     :    :    **********|    :    :    :    :
  89-92    LSKC           0.35     :    :    :    :  . |**  :    :    :    :
  93-96    AGAV           1.85     :    :    :    :  . |*********:    :    :
 97-100    KIFA           1.27     :    :    :    :  . |******   :    :    :
101-104    HLIV           2.30     :    :    :    :  . |***********   :    :
105-108    LKKH          -1.83     :    :    :*********|    :    :    :    :
109-112    GTFM           0.88     :    :    :    :  . |****:    :    :    :
113-116    LAFN           1.17     :    :    :    :  . |******   :    :    :
117-120    LTDE          -1.00     :    :    :    *****|    :    :    :    :
121-124    NRGL          -1.17     :    :    :   ******|    :    :    :    :
125-128    SFYA           0.58     :    :    :    :  . |*** :    :    :    :
129-132    KKPD          -3.23     :   ****************|    :    :    :    :
133-136    LSPE          -0.57     :    :    :    : ***|    :    :    :    :
137-140    LWKI           0.85     :    :    :    :  . |****:    :    :    :
141-144    FQQA          -0.63     :    :    :    : ***|    :    :    :    :
145-148    VKDV           0.25     :    :    :    :  . |*   :    :    :    :
149-152    GMDE          -1.38     :    :    :  *******|    :    :    :    :
153-156    SEIV           1.07     :    :    :    :  . |*****    :    :    :
157-160    FVDW           0.63     :    :    :    :  . |*** :    :    :    :
161-164    TKDK          -3.00     :    ***************|    :    :    :    :
165-168    CSEQ          -1.35     :    :    :  *******|    :    :    :    :
169-172    QKQQ          -3.60     : ******************|    :    :    :    :
173-176    LELE           0.10     :    :    :    :  . |*   :    :    :    :
177-180    KETK          -3.00     :    ***************|    :    :    :    :
181-184    KETK          -3.00     :    ***************|    :    :    :    :
185-188    KDPQ          -2.25     :    :   ***********|    :    :    :    :
```

Fig. 4.25. Comparison of rat's and human α_1-AGP hydropathy. Figure legend is as same described in Fig. 4.23 except that length of jump is extended to 10 amino acid in order to do more general comparison. In each jump, only the first 4 amino acids are listed.

A. Hydropathy of human α_1-AGP.

B. Hydropathy of rat α_1-AGP.

A

region	seq.	hydropathy
1-4	QIPL	0.78
11-14	VPIT	1.60
21-24	ITGK	-0.13
31-34	AFRN	-0.88
41-44	VQEI	0.42
51-54	FTPN	-0.78
61-64	FLRE	-0.40
71-74	QCIY	0.55
81-84	VQRE	-1.83
91-94	YVGG	0.52
101-104	LLIL	3.90
111-114	MLAF	2.53
121-124	NWGL	-0.28
131-134	PETT	-1.62
141-144	FYEA	-0.07
151-154	PKSD	-2.48
161-164	KKDK	-3.80
171-174	QHEK	-3.53
181-184	GESO	-1.20

B

region	seq.	hydropathy
1-4	QNPE	-3.03
11-14	GIPI	1.75
21-24	WLSD	-0.40
31-34	AAFR	0.45
41-44	AVQT	0.45
51-54	YLTP	0.03
61-64	IELR	0.05
71-74	DQCV	-0.08
81-84	GVQR	-1.05
91-94	KCAG	-0.00
101-104	HLIV	2.30
111-114	FMLA	2.53
121-124	NRGL	-1.17
131-134	PDLS	-0.57
141-144	FQQA	-0.63
151-154	DESE	-2.85
161-164	TKDK	-3.00
171-174	QQLE	-1.70
181-184	KETK	-3.00

Fig. 4.26. A hydropathy plot of human pre-α_1-AGP. The plotting is essentially described by Kyte, J. and Doolittle, R.F. (in press). The dotted line at -0.4 on the graph corresponds to the overall average.

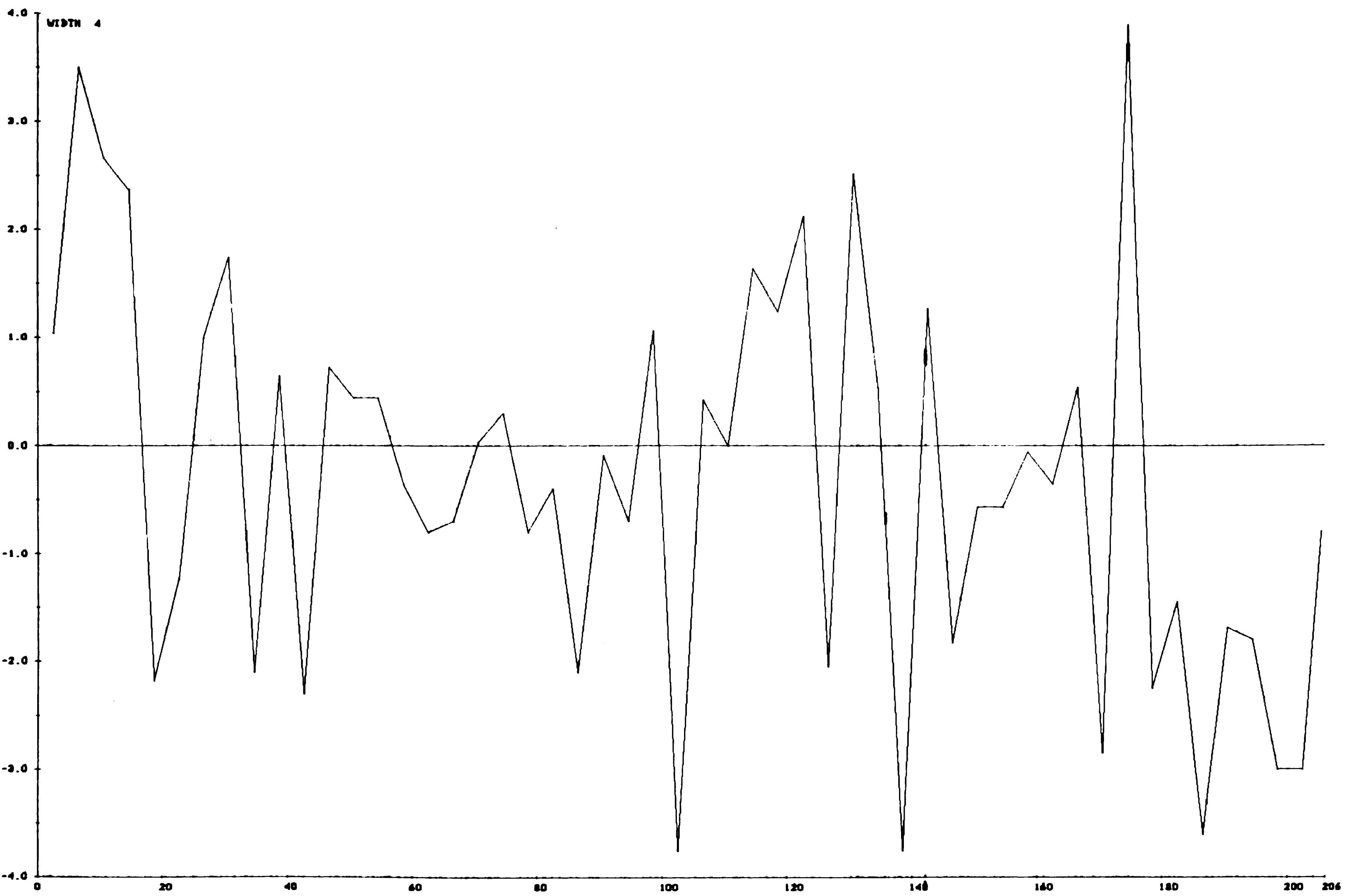
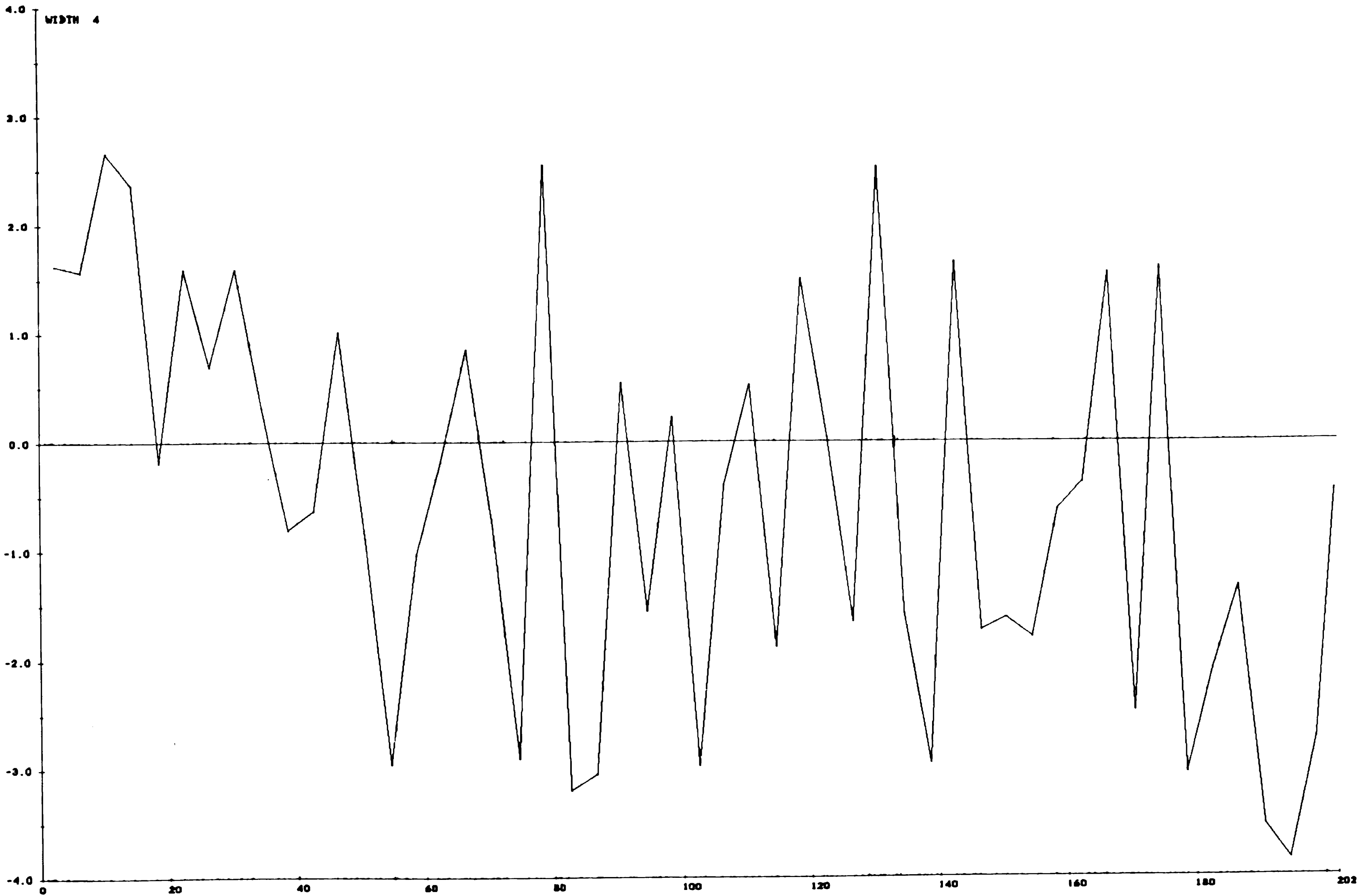

WIDTH 4
4.0
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
0
20
40
60
80
100
120
140
160
180
200
206

Fig. 4.27. A hydropathy plot of rat pre-α_1-AGP. The plotting is essentially described by Kyte, J. and Doolittle, R.F. (in press) The dotted line at -0.4 on the graph corresponds to the overall average.


WIDTH 4
4.0
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
0
20
40
60
80
100
120
140
160
180
202

4.4 Discussion

A homology plot of amino acid sequences vs. itself usually reveals certain internal information about structure and function. In the example of the rat and human pre-α_1-AGP and mature α_1-AGP, the existence of diagonal dots beside the linear line indicates several interesting points. Under the stringency of 4:7 match (Fig. 4.1 and 4.2), the drifted double dots at the amino end (Fig. 4.2) and carboxyl end indicate repeated amino acids. A closer examination at the carboxyl end of the rat α_1-AGP (Fig. 4.3) shows that the homology scores are the result of 2 Gln-Gln dipeptides, 2 Leu.Glu dipeptides and 2 Lys.Glu.Thr.Lys.Lys pentapeptide within the last 20 amino acid residues. In human α_1-AGP, the similar situation is not observed, although, there is a minor degree of repeats caused by the abundance of lysines and glutamic acids. The functional significance of the repeats in rat α_1-AGP is not clear because the human version does not contain it. It may be a consequence of genetic duplication. Both human and rat α_1-AGP do not contain many glutamic acids or lysines and, consequently, this region is highly hydrophilic. The presence or absence of the amino acid repeat at the amino terminal between the rat pre-α_1-AGP and AGP, respectively (Fig. 4.1 and 4.2), demonstrates that it has originated from the signal leader. The leader sequences of proteins include recognition sites for Golgi translocation and post-translational modification, and are cleaved before product secretion. Therefore, this region plays no functional purpose for the mature α_1-AGP. A close examination and comparison shows that they are derived from the dipeptides of Valine-Leucine dipeptides in both rat and human. The repeats are not numerous, occur only twice in both cases, and therefore, result in only two scoring dots on each plot. Like many other secreted proteins, the amino acids of leader sequences in human and rat pre-α_1-AGP are made of

hydrophobic residues and are extremely similar between these two species. The details are discussed in another paragraph.

Figures 4.4, 4.5 and 4.6 illustrate the overall characteristics of the rat pre-α_1-AGP and AGP including molecular weights, frequencies of amino acids and coding triplets. Amino acids in abbreviations are also shown. Although the information is obvious, it is necessary to list them for orderly and convenient comparsions which are frequently referred to in latter discussions. For example, a quick comparison between Fig. 4.5A and Fig. 4.6A reveals that the differences are exclusively due to 7 leucines, 3 valines, 1 proline, 2 alanines, 1 histidine, 1 glutamine, 1 serine, 2 methionines, which all contain non-polar groups and make up the signal sequences. The frequencies of codon usages will be discussed in the section of evolution.

Similarly, Fig. 4.7, 4.8, and 4.9 of human pre-α_1-AGP and AGP provide easy comparisons and technical conveniences. This information will be referred often in latter discussions.

A residue-by-residue comparison of amino acids between rat and human pre α_1-AGP is shown on Fig. 4.10. From the visual outlook, it is clear that these two proteins retain substantial homology across the species barrier. Major homology occurs within the leader signals and spreads out sporadically along the whole protein. It is reasonable to postulate that α_1-AGP plays an essential role for animal survival, and signal peptide-recognition is extremely conserved, probably due to a selection constraint of nature, i.e., any mutation at the leader sequence would probably result in an unrecognizable leader and lead to failure of α_1-AGP modificaton, maturation and secretion. Therefore, without the presence of α_1-AGP, animals would not survive. This conservation of leader sequence of α_1-AGP in man and rat leads to the consideration of the recognition procedure of peptide signals as a multi-step or high fidelity system which tolerates

little alteration on substrate and is highly specific to individual proteins. A comparison between rat α_1-AGP and albumin leader signals reveals that no homology data supports this speculation. Among a total of 18 amino acids of human rat leader peptides, only three amino acids are different. They are histidine vs. serine, methionine vs. tryptophan and threonine vs. valine at amino acid positions of 4, 5 and 8, respectively (Fig. 4.10). These different amino acids retain their common classification as non-polar. Closer check of the triplet codons-coding rat and human pre-α_1-AGP leader peptide, show that there are an additional five codon changes which result in no amino acid changes. Codon degeneration accompanied by the absence of amino acid alterations usually indicates considerable selection pressure from nature. On the contrary, other amino acids of the human and rat α_1-AGP, although maintaining a significant degree of homology, show a decrease (Fig. 4.11). The values drop from 85% of leader signal to 50%, 50%, 65%, 50%, 29% and 30% according to their coding exons I, II, III, IV, V and VI, respectively. This fact demonstrates that the amino acids in the rest of α_1-AGP probably are not as functionally critical as the signal peptide. In order to further identify the locations of amino acid homologies among the α_1-AGP, a series of homology plots between rat and human α_1-AGP with different stringencies were performed. Under the most stringent situation, that is 7 out of 7 match in amino acid sequences (Fig. 4.12A), there is no homology except the signal peptide area and a short match at position 100 of both human and rat α_1-AGP.

The short match was found to be caused by the identical sequences of a heptapeptide of valine.glutamine.arginine.glutamic acid.aspargine.glycine.threonine in human and rat α_1-AGP. Consequently, when the homology requirements are relaxed (reduced from 7:7 to 7:6, 7:5 and 7:4) the homologies of a lesser degree surface. In general, the similarities of human and rat α_1-AGP are center

in the middle region, then spread throughout the whole peptide backbones and fade at the carboxyl terminals. They seem to have no special preference for certain amino acid residues to be used as homologous points, but glutamine and lysine are the most common matches, and a majority of the rest are non-polar amino acids. It may be reasonable to say that the overall trend of conservation relates to the general shape of the molecule instead of special preservation of functional groups. At a homology stringency of 3:7, the background "noise" starts to show. Although it is not significant to extrapolate any conclusion under this condition, the drifted homologies at the far amino and carboxyl ends of human rat α_1-AGP match the gaps of linear continuation between the two. This result indicates a higher rate of amino acid degeneracy at these two areas. Functionally, it represents an area (at the carboxyl terminal) of lesser importance, and therefere, less constraint from selection. Genetically, it underlines another active reconstruction for generating diversities; and therefore, a result of either random mutation or purposeful gene reconstruction. The documented similarity between the α_1-AGP carboxyl terminal and immunoglobulin light chain and heavy chain, respectively (Schmid, 1973) further implicates the latter possibility. The discussion of this aspect is presented later when nucleotide sequences homology is considered.

Among the about 200 amino acids of human and rat pre-α_1-AGP, there are 106 changes and the overall homology is about 49%. It has been estimated that the average rate of one amino-acid change is about $7 x 10^6$ per site/year (Kimura, 1968). Based on this assumption, it will take 74.2 million years ($106 x 7 x 10^6$ years) to accumulate the differences of 106 amino acid between man and rat. The fossil records indicate the time of divergence of all mammals is at about 80 million years ago. The similar result of calculations derived independently from both morphological fossils and the molecular "fossil" of α_1-

AGP raise certain interesting points, however, great caution must be exercised to avoid extrapolations based on simple "coincidence".

It is probably safe to say that the α_1-AGP is an ancient protein. It has been moderately conserved compared to other proteins, and it plays an important role for mammalian survival. Without its presence in serum (that is the failure of secretion caused by aberrant signal peptides) mammals can not survive longer than the age of reproduction. The exact functions of α_1-AGP are still unknown, but its role involving the fighting off infections, resolving tissue damage, its abundance in serum, and its structural relationship with immunoglobins may classify it as an important member of a system defending mammals against various biological injuries.

The moderate amino acid homologies of 49% between man and rat α_1-AGP are significant but not remarkable. Other proteins such as histone and cytochrome c are almost completely conserved across not only species but also kingdom barriers. The rate of amino-acid substitution derived from comparing mammalian and avian cytochrome c, which consists of about 100 amino acids, is one change in 45×10^6 years (Margoliash et al., 1965). On the other extreme, the rate of change is about one change in 2.7×10^6 years for triosephosphate dehydrogenases of human rabbit and cattle. Averaging for many other proteins including hemoglobin, etc., provides average value is about one change of amino acid in 28×10^6 years for a polypeptide of 100 amino-acids. Assuming that the accepted 80 million years is accurate and to be the diverging time of all mammals, α_1-AGP with total changes of 106 amino acid between man and rat is estimated to have one amino acid change in 7.6×10^6 years (80×10^8 divided by 106) which is slightly higher than average value and further indicates its functional importance. Nevertheless, about 50% differences in amino acid sequences between man and rat have been accumulated and therefore, indicate the involvement of

other criteria for selection. This leads to the consideration of carbohydrate groups as other functional requirements.

The α_1-AGP is composed of more than 55% of carbohydrate groups. The unusual glycosylations and extreme acidity of this molecule probably impose certain biological functions. The α_1-AGP is one of the most acidic serum proteins in existence, and it is logical to compare its glycosylation sites at the levels of both amino acids and nucleotide sequences.

Totally, there are five N-linked, complex-type oligosaccharide chains in human α_1-AGP at positions 15, 38, 54, 75 and 85. In rat there are 6 such sites at positions 7, 16, 58, 76, 86 and 116 (Fig. 4.15). The universal glycosylation sequences of Asn.X.Thr(Ser) tripeptides are present precisely in both cases. It is important then, to pursue the functional significance of these groups on the basis of evolutionary consideration.

Although the glycosylation sites in human and rat AGP do not line up exactly in the number of amino acid residues, by examining the surrounding areas, it can be concluded that positions 15, 54, 75 and 85 of human α_1-AGP match the positions 16, 58, 76, and 86, respectively, and are probably originally from the common ancestorial sites. These observations are also supported by the fact of identical triplet codons used in the identical matching points. For example, all Asn residues are coded by AAC in both human and rat α_1-AGP at position 54 vs. 58, 75 vs. 76 and 85 vs. 86, respectively. A different triplet codon for Asn of AAT is used in position 38 vs. 16 of human and rat, respectively. Since the evolution of homologous genes between different species are mainly divergent, this finding indicates a strict requirement constraining the mutating sites among the glycosylation sites. There are only two triplet codons, AAC and AAT for Asn, therefore, any changes except AAC to AAT and vice versa will lead into a change of amino acid; as a result there will be a disappearance of

a glycosylation site. For a total of 20 amino acids found in normal proteins 61 (excluding the 3 stop codons) triplet codons, there will be a combination of 526 possible changes resulting from 1 base alteration. Among the 526 changes, 392 of the changes will result in an alteration of amino acid and 134 will be a silent mutation because of no change on amino acid phenotypes. The window allowing silent mutation on Asn is extremely narrow, only 2 out 526. The requirement of consistency of 2 neighboring amino acids for glycosylation probably is the reason for long term stability of glycosylation sites. Therefore, it may be deduced that the number of glycosylated groups is more important than the primary structure of the α_1-AGP. Besides the 4 homologous glycosylation sites, rat α_1-AGP has 2 additional glycosylation sites at positions 7 and 116; the human version has another one at position 38 which did not correspond to one in another location. It is difficult to say if these variances are due to the gain or loss in terms of their counterpart amino acid residues of different species. However, looking at the corresponding areas with amino acid, nucleotide sequences, they are all occurring at areas with homology loss of more than one residue showing a pattern of degeneracy eroding toward conserved sequences. The first glycosylation signal of Asn.Ile.Thr of rat α_1-AGP at position 7 is probably due to the original conservation, and its human counterpart of position 7 of Asn.Leu.Val which is no longer being glycosylated and is probably due to the change of codons coding Thr or Ser into Val (GTA). This change is obviously tolerated. The second glycosylation site of human α_1-AGP at position 38 is Asn.Lys.Ser and its corresponding tripeptide in rat α_1-AGP is Lys.Glu.Ala. These are groups of unrelated amino acids, and therefore, it is difficult to determine whether it is due to the loss or gain of glycosylation in rat and human. Similar situations occur at position 116 of the rat α_1-AGP. At this position, it is Asn.Leu.Thr for rat and Asp.Val.Asn for human. Although it is difficult to predict the cause-

effect, gain-loss situation, basing exclusively on the tripeptide information, extended consideration of neighboring areas tells the difference. The tripeptide differences dwell in the middle of a homology of Met.Leu.Ala.Phe.Asp.Val. Asn.Asp.Glu. and Met.Leu.Ala.Phe.Asn.Leu.Thr.Asp.Glu of human and rate α_1-AGP, respectively. It is hard to imagine that a successive mutation could achieve an additional glycosylation signal without degenerating the surrounding codons. Therefore, it is more plausible to postulate that the absence of glycosylation tripeptide in human α_1-AGP is probably due to the random mutations which accumulate at this site. As far as glycosylation is concerned, it is probably more logical to hypothesize that originally there were 5 glycosylation sites on the ancient version of α_1-AGP existing at positions 15, 54, 75, 85, 116 (postulated), and 16, 38 (postulated), 58, 76, 86, 116 of human and rat α_1-AGP, respectively. Through time, humans gain additional sites at position 38 and lose one at position 116. Rats on the other hand have not lost any, but gained an additional one at position 7. Due to the variable number of 5 and 6 carbohydrate groups in human and rat, the relative easiness of being lost (any change of the tripeptide concensus will cause it) and yet still conserved, we may predict that the importance of carbohydrate side chains are not absolute but must be extreme. The loss of glycosylation sites toward the carboxyl terminal part of the human α_1-AGP at position 116 in relationship to the rat is particularly intriguing. In this region starting from position 128 (Thr) to 150 (Pro), there is a homology consisting of 23 amino acids with 9 differences. If we isolate this fragment as an exclusive ground of evolutionary events which is representative for the whole α_1-AGP molecule we may calculate that totally $9x(7x10^6)$ years or 6.3 million years are needed to accumulate these changes. In other words, humans probably lost the 116 glycosylation sites within the last 6.3 million years - a relatively young

event. Again, this is a result derived from two bold assumptions and shall remain as a speculation.

It has not been determined whether the different numbers of carbohydrate side chains between human and rat α_1-AGP would cause any functional difference. The multi-step process for biosynthesis and secretion of glycoproteins is extremely complex and the transport of different glycoproteins is regulated according to their characteristics (Ledford et al, 1983; Lodish et al, 1983). For example, when glycosylation is completely blocked by a compound such as tunicamycin (Ledfor et al., 1983; Hickman et al., 1977), or oligosaccharide processing is inhibited by 1-deoxynojivimycin (Gross et al., 1983; Lodish et al., 1984), the transport and the secretion of some glycoproteins are stopped while some of others are not. It was found that the rate of rat α_1-AGP secretion relies heavily on the presence of its oligosaccharide chains (Pamela et al., 1985). As a result, the transportation rate are found to be proportional to the number of α_1-AGP oligosaccharide side chains. The slowest rate of secretion of rat α_1-AGP is the unglycosylated form; the fastest is the fully glycosylated form (6 oligosaccharides). Partially glycosylated (1-5 oligosaccharide chains) exhibit intermediate rates according to the abundance of glycosylations. These data raise two questions: Can we extrapolate this observation to human α_1-AGP section? Can we regard the human α_1-AGP (normally only 5 carbohydrate side chains) as an equivalent to the rat α_1-AGP with 5 carbohydrate side chains? The extrapolation seems to be unwarranted because multi-step, complex systems such as glycosylation, transportation facilities, or translational machineries usually exist as ground rules of organismal life and should be highly resistent to change. If this is true, it is a reasonable explanation for the 100-fold increase in the level of serum α_1-AGP in rat while for human α_1-AGP it is only 10-fold. A comparable difference is observed between the rat α_1-AGP with 5 and 6

oligosaccharide side chains versus that synthesized in rat hepatoma cells. It is likely then that the differences of elevation of α_1-AGP in rat and human during acute reaction is mainly due to the intrinsic difference of amino acid sequence which in turn determines the numbers of carbohydrate side chains. Simple experiments of measuring α_1-AGP mRNA and protein levels between rat and human will solve this problem; unfortunately, suitable human hepatoma cell lines are lacking for comparable in vitro measurement. It is still unknown if the 10-fold abundance of α_1-AGP in rat versus man provides any beneficial effect for survival. The rat has been well-known for their physiological superiority to face environmental challenges as a wild animal.

As mentioned previously, analysis based on nucleotide sequences has an advantage over amino acid analysis by showing generation affect instead of absolute time effect. Moreover, nucleotide changes can reflect the detailed record of mutations which does not show on amino acids if the mutation is silent (that is not altering phenotype). Fig. 4.16 demonstrates the base to base homology between human and rat pre-α_1-AGP cDNA. There is a total of 360 nucleotide differences between human and pre-α_1-AGP cDNA. Comparing the total differences of amino-acid residues (106), there are 42 nucleotide changes (360-106x3) acting as synonymous mutations. In order to visualize the nucleotide sequence homology between human and rat, a homology plot is again generated under various stringencies ranging from 4:7 (Fig. 4.17), 5:7 (Fig. 4.18), 6:7 (Fig. 4.19) and 7:7 (Fig. 4.20). The expected linear homologies are observed through different degrees of matching which is relatively consistant with amino-acid homology plot. The surprising result is the unusual cluster surrounding the near 3' end. This cluster means extensive short repeats within this area. Amino acid sequences in this area are not preserved, but nucleotide sequences are. As mentioned before, the carboxyl-terminal region of α_1-AGP shows a significant

degree of homology with the constant region of the H chain of IgG. Moreover, a high frequency of amino acid substitutions were found within this area using pooled human serum as a source which further suggested a relationship between α_1-AGP and the immunoglobin. A direct examination of nucleotide sequences in this area shows unusual high frequencies of mixed G.A codons such as GGA, GAA, AGA, AAG, AAG overlapping with one another. Since there is no significant amino acid homology between human and rat α_1-AGP within this segment, yet there are extensive A.G-rich nucleotides, it is probably due to a result of repeated breaking and rejoining at this area - a situation observed in the immunoglobulin gene. This observation leads to the conclusion that α_1-AGP is derived from an immunoglobin-like supergene, and after its evolution, its close relationship with immunoglobin might (continously) subject its gene to the diversifying mechanism of immunoglobin genes until it is "fixed" to become "independent" and selected to evolve into a useful serum protein performing selectable functions. Nevertheless, the residual "scars" linger and time is not long enough to wipe out the marks.

The nucleotide comparison can also be extended from cDNAs to the entire α_1-AGP gene of human and rat. Unfortunately the sequence of human α_1-AGP gene has not been completely published. (This comparison is therefore limited to the region of splicing junctions.) In Fig. 4.21, comparisons were organized in such a way to visualize the base by base matching of both introns and exons according to their location orders in the gene. It is interesting to see that not only are the relative lengths of introns and exons of human and rat α_1-AGP genes conserved, the splicing junctions (especially the splicing donors and acceptors) are also highly conserved despite the long period of time since human and rat divergence. In general, all introns of the rat are shorter than the human, except the 5th one.

It has been argued that generation time instead of absolute time should be used for considering molecular evolution. Currently, most workers adopt the rate-consistency hypothesis which assumes the rate of molecular evolution is proportional to absolute time. A recent study by Wu et al. (1985) demonstrates that generation time instead of absolute time is probably more relevant to estimating the changes of molecular evolution. In that study, 11 genes from rodents (mouse or rat) and man were compared with those from other mammals. It is concluded that rodents with considerably shorter generation time than human (100x shorter) evolve significantly faster than man (Wu et al., 1985) due to their higher mutation rates. This observation seems to apply well to the case of the α_1-AGP gene.

All intron sizes are reduced in the rat α_1-AGP gene except the fifth intron, which probably is exceptional and will be discussed later. Introns have been postulated to be parasitic DNA playing no known functions. Normal expression of most if not all mammalian genes do not need the existence of introns. Artifically constructed genes containing proper promotor in conjunction with its cDNA (no introns) can be expressed more efficiently in cell culture transfection. In the case of α_1-AGP, an animal (rat) with significantly faster mutation rate due to shorter generation time must evolve toward more efficient expression, a fact due to the benefit of α_1-AGP level and favorable selection pressure. As a consequence, deletions of intron sequences will be encouraged as fast as possible as long as they do not disrupt the splicing procedures. Although the intron sizes are reduced in rat α_1-AGP gene, the sequences at splicing junctions demonstrate extreme conservation. This is another important role of splicing junctions participating in the mRNA splicing action. The degree of concensus at splicing junctions between rat and human exceed that of their reading frames (Fig. 4.22). This fact indicates that the maintenance of normal

splicing is far more improtant than the intrinsic amino acid sequences, and the mRNA splicing machineries do have their fidelity in splicing to rule the outcome of RNA maturation. This result may also illustrate the highly conserved splicing mechanism among animals. The processing of mRNA precursor or nRNA has been shown to require a very large structure, which apparently contains multiple components including several protein and RNA molecules with a total mass of 60s. For comparison, a 30s ribosomal subunit already contains 21 protein molecules and a 1500-nucleotide long RNA. If all components are critical in proceeding to normal splicing, very little change would be tolerated. The high fidelity of the splicing complex between man and rat may be used as a premise to determine the sequence requirement involved in splicing. For example, the sequence concensus and differences between human and rat α_1-AGP can be used to deduce a common sequence denominator for mRNA splicings. Unfortunately, the complete nucleotide sequence for human α_1-AGP is not available.

The exceptionally long 5th intron of the rat α_1-AGP gene in contrast to other introns may not be surprising. As described in the discussion of Chapter 3, Intron V of the rat α_1-AGP gene contains not only two potential glucocorticoid receptor binding sites but also very unusual structures of GTGTGTGTGT$(G)_{10}$T$(G)_3$ and GTTTGTTTGTTGTTTTGTTTTT which connect to the TGTTCT hexanucleotide of glucocorticoid receptor binding at the upstream end (Fig. 3.7). It is not known whether the peculiar sequences of the fifth intron are involved in expression regulations, splicing or both. Deletion of this intron followed by transfection experiment may reveal this mystery. The additional sequences in the fifth intron of the rat α_1-AGP gene in comparison to man probably are not merely parasitic. They must provide certain advantages for the rat through rapid and progressive mutations.

Functional and structural analysis based on hydropathy of human and rat α_1-AGP was carried out by various methods. A software program analyzing protein 5' hydrophobic and hydrophilic natures based on amino-acid-constituents was used to determine the potential similarities between human and rat α_1-AGP in order to consider their potential functions and structures (Fig. 4.23, 4.24). With various degrees of jump (la.a. to 10 amino acids) and a width of 4 (most commonly used), general outlines of hydropathy in relationship to carbon backbone of human and rat α_1-AGP were revealed. The hydropathy features of these two molecules are strikingly similar.

It is clear that in spite of the nucleotide substitutions, the amino acid substitutions, and the general feature of the human and rat α_1-AGP remain relatively unchanged. The almost identical hydropathic patterns and the existence of highly acidic oligosaccharide chains occupying similar positions dictate and ensure the overall similarity of these molecules. The α_1-AGP, therefore, must be carrying out its mission through its abundance of carbohydrate side groups in addition to its specific morphology in the aqueous phase.

The carboxyl termini of human and rat α_1-AGP have only 29% homology on amino acid sequences, yet hydropathic features remain virtually unchanged. This is due to the selective adoption of different but always highly ionic amino acids throughout this area.

The shape similarities between rat and human α_1-AGP are further reinforced by the nearly identical disulfide bonds between cysteine residues of positions 72-164 and positions 73-165 of the human and rat, respectively. The additional disulfide bond of human α_1-AGP between position 5 and 147 (which is absent in the rat) probably is not essential, since the highly hydrophilic amino-acid backbone and exceedingly acidic carbohydrate side chains plus the previously described disulfide bond are more than enough to crush the non-polar segments

of α_1-AGP into a centrally hydrophobic domain and force the molecule to assume a common morphological feature.

CHAPTER 5: REGULATION OF THE RAT α_1-AGP GENE

5.1 Introduction

The prompt elevation of α_1-AGP in animals during acute phase reaction and the in vitro induction by glucocorticoid as previously described make the study of its gene expression an interesting and important subject. In order to study the mechanism of glucocorticoid induction, transfection of the entire gene into another host cell is one logical approach. The requirement is that the new host cell must possess all induction machineries for glucocorticoid responses in addition to the absence of the target gene. If the transfected gene contains all regulatory elements, the newly transfected cell will respond to the glucocorticoid stimulation and produce a measurable level of product derived from the newly established gene according to its characteristics. Based on this premise, the rat α_1-AGP gene was transfected into a mouse LTk(-) fibroblast cell (Reinke et al., 1985), and it was discovered that the isolated lambda recombinant contained all the regulatory sequences needed to respond to glucocorticoid induction. This study, however, could not identify the specific regions of the gene corresponding to the postulated steroid hormone receptor binding. As indicated in the sequences data of the rat α_1-AGP gene, it is possible that regulatory sequences may exist within the 5' end introns or the 3' end regions. In order to identify the exact role of specific DNA fragments of the rat α_1-AGP in regulating its expression, a more detailed and systematic reconstruction of the rat α_1-AGP gene seems to be desirable. In this chapter, a first attempt at this identification is described. During this study, the 5' end area of the rat α_1-AGP gene was isolated to fuse with another assayable gene in order to identify the extent of 5' end promotion under glucocorticoid influence.

Although RNA levels represent the most definite measurement of promotor activity, the technical requirements are great and transcription may be difficult to obtain unless a given promotor is particularly strong. It is often preferable to determine the function of a promotor by joining the promotor to a second gene segment which codes for a readily assayable enzymatic function. Therefore, instead of measuring the α_1-AGP mRNA level, an approach of fusing the 5' end DNA of the rat α_1-AGP gene with chloramphenicol acetyltransferase gene then assaying its activities in a transient assay was attempted. Quantitation of RNA is bypassed along with other tedious and inaccurate measurements. A long and unpredictable consequence of conventional transfection for establishing a permanent cell line was also avoided.

5.2 Methods and materials

Preparation of plasmid DNAs. The precedure of preparing plasmid DNAs was essentially as described by Maniatis (1983) by lysozyme-alkaline lysis and cesium chloride-ethidium bromide equilibrium gradient centrifugation. All DNA to be utilized in eukaryotic cell transfection experiment were further purified by a second round of equilibrium centrifugation steps.

Enzymes. Restriction endonucleases were obtained from New England Biolabs or Bethesda Research laboratories. DNA polymerase I were obtained from Boehringer-Mannheim, T4 polynucleotide kinase from P-L Biochemicals, bacterial alkaline phosphatase from Worthington Diagnostics, and T4 DNA ligase from New England Biolabs.

Preparation of DNA fragments. Restriction endonuclease digestions were performed according to the suppliers requirements. DNA fragments were purified by agarose gel electrophoresis and visualized by ethidium bromide staining. The

DNA fragments were eluted by electrophoresis of isolated gel in a dialysis bag then phenol-chloroform extraction and ethanol precipitation.

Bacterial transformation. Ligations were done by incubating the fragment-vector mixture with T4 DNA ligase overnight at 14°C. Frequently, incubation in room temperature for 4 to 5 hours were also done without any observable difference. Transformation of Escherichia coli HB101 was carried out according to the method of Mandel and Higa (1970). Transformed colonies were selected on plates containing ampicillin (50μg/ml).

Mammalian cell transfection. Cells were usually plated out at the density of $10^4/cm^2$ in 100 mm plates 24 hours before transfection. The media of cell culture were replaced with fresh media containing charcoal treated 10% fetal calf serum. In order to achieve consistant results, all DNA preparations were monitored by agarose gel electrophoresis to ensure that the plasmid DNA was intact. This is to avoid nicked and linearized DNA molecules which usually perform unpredictably in tissue culture transfection. It was essential to keep DNA in darkness once ethidium bromide is in contact with plasmid DNA. Calcium phosphate-DNA precipitate was prepared by the method of Graham and Van der Eb (1973).

Very fine precipitates were prepared by mixing the DNA-$CaCl_2$ and HEPES (N-2-hydroxyethylpipirazine-N'-2-ethanesulfonic acid)-buffered sodium phosphate solution under a gentle stream of nitrogen. Precipitates were allowed to stand for 30 minutes without agitation before being added to the tissue culture cells. Initially, the amount of DNA added was varied between 1 and 25 μg per plate to determine the linear range of uptake and expression of the recombinants. A comparison control between the parental PSV2-cat and the recombinant vector was performed with 10 μg of DNA per plate.

Assay of CAT activity in HTC(-) cell. Cell extracts were made 48 hours after transfection. Cells were sonicated, washed and pelleted in 100 µl of 0.25 M Tris hydrochloride (pH 7.8). After the cells were centrifuged for 15 minutes in an Eppendorf benchfuge at 4°C, the supernatants were transferred to be measured for CAT enzymatic activity. The assay volume is 180 µl containing 100 µl of 0.25 M Tris-hydrochloride (pH 7.5), 20 µl of cell extract, 1 µCi of $[^{14}C]$ chloramphenicol (50 µCi/mmol; New England Nuclear Inc.), and 20 µl of 4 mm acetyl coenzyme A. Control contained CAT (0.01 U; P.L. Biochemicals Inc.) instead of cell extract. All the reagents were preincubated together for 5 to 10 minutes at 37°C, except coenzyme A. After temperature equilibrium was reached, coenzyme A was added to start the reaction. The reaction was terminated and the chloramphenicol was extracted with 2 ml of cold ethyl acetate. The organic phase was dried, resuspended in 30 µl of ethyl acetate and spotted on silica gel thin-layer plates. The plate was run with ascending chloroform-methanol ratio of 95:5. After autoradiography of the separated acetylated chloramphenicol isoforms, spots were cut out and radioactivity was measured. All data are illustrated as the amount of chloramphenicol acetylated by 20 µl of extract from transfected cells.

5.3 Results

Construction of pAGP-CAT 1. The construction of pAGP-CAT 1 was derived from linking the 5' end area of Sac I - Hpa II fragment of the rat α_1-AGP lambda 40 DNA with a plasmid (pSVO CAT) containing CAT gene. The CAT gene was original found on E. coli transposable element Tn9 (Fong, B. et al., 1980), which delivers resistance to the antibiotic chloramphenicol, consists of a 1,102 bp CAT cistron flanked by two 768 bp Is 1 elements (Alton, N., et al., 1979). Tn9 was relocated into the E. coli plasmid pBR322 by stepwise transferring

from an R factor through bacteriophage P1 to bacteriophage lambda derivatives then to pBR322 (Scott, J., 1973).

An E. coli strain containing one of these pBR322-Tn9 constructs was established, and the plasmid was further used to isolate a CAT gene fragment by complete Tag I digestion. The 773-bp fragment deleted from CAT promotor sequence was generated by incubating the Tag I fragment which contains the CAT gene with DNA polymerase I to create blunt ends and then ligated with mixed Hind III and Bam HI synthetic oligonucleotide linker on both 5' and 3' end (Gorman, C.M. et al., 1982).

The modified 773-bp fragments was ligated with the prokaryotic/eukaryotic vector pSV2 which consists of the replication origin and ampicillin resistance gene of pBR322 and a Simian virus 40 (SV40) early transcription unit at Hind III site. The resulting plasmid pSV2-CAT, therefore, contains the entire CAT coding gene and SV40 promotor. The plasmid pSVO CAT used in this project was derived from PSV2-CAT by deleting the SV40 promotor and then resealing at Hind III site (Gorman, C.M., et al., 1982). After modifications, the pSV2-CAT plasmid, without the presence of eukaryotic promotor, contains a unique Hind III site for subsequent clonign of other eukaryotic promotors.

In this study, a 758-bp Sac I - Hpa II fragment was isolated from the rat α_1-AGP gene carrying M-13 bacteriophage and blunted at both ends by the large fragment of DNA polymerase I. The pSVO CAT plasmid was opened at the Hind III site and also blunted at both opening ends by Klenow large fragment of DNA polymerase I. The 758-bp Sac I - Hpa II fragment joined the opened pSVO plasmid by subsequent ligation, and one of the resulting plasmids was selected and verified by Bgl II and Kpn I restriction mapping to ensure the directional 5' toward 3' insertion. This plasmid is designated as pAGP-CAT 1 (Fig. 5.1).

Assay for expression of pAGP-CAT 1 in mammalian cells. CAT is able to inactivate chloramphenicol by transferring acetyl group to chloramphenicol to form non-functional mono- and diacetylated derivatives (Shaw, W., 1967). Several assays have been developed to monitor this activity. The methods of Cohen et al. (1980) and Shaw et al. (1968) were used in this study. In this experiment the acetylation of chloramphenicol is measured by silica gel thin layer chromatography which is a very sensitive and highly specific method to separate the parental, mono- and diacetylated forms of chloramphenicol.

The pAGP-CAT 1 was transfected into rat HTC cells by the method of calcium phosphate precipitation (Graham et al., 1973). At 48 hours after transfection, the cells were collected and sonicated to serve the non-induced control. A calf thymus DNA-transfected cell extract was also monitored as a control. Additional control was the extract from pSVO CAT transfected cells which express no CAT activity due to the absence of promotor.

The autoradiogram (Fig. 5.2) was developed after 24 hours in -70°C with intensifier. The quantitative activities were also monitored by counter. No observation of CAT activities was found in calf thymus DNA and pSVO CAT plasmid transfected cells. On the other hand, 1.5×10^5 cpm and 7.2×10^6 cpm of $[^{14}C]$ chloramphenicol was acetylated in the absence and presence of glucocorticoid (1×10^{-6} M), respectively.

Fig. 5.1. Construction of pAGP-CAT 1.

(A) Parental plasmid (PSVO-CAT) was digested with Hind III to become (dark arrow) an opened receptor plasmid. Meanwhile, M-13 RF (10 Bam.Sst.1.7K2) was digest with Sst I and Hpa II. The restricted Sst I/Hpa II fragment which contains the potential α_1-AGP promotor was purified by acrylamide gel electrophoresis.

(B) Both Sst I/Hpa II fragment and Hind III digested PSVO-CAT were treated with large Klenow fragment to blunt both ends.

(C) Sst I/Hpa II fragment and PSVO-CAT plasmid was ligated to form the pAGP-CAT 1.

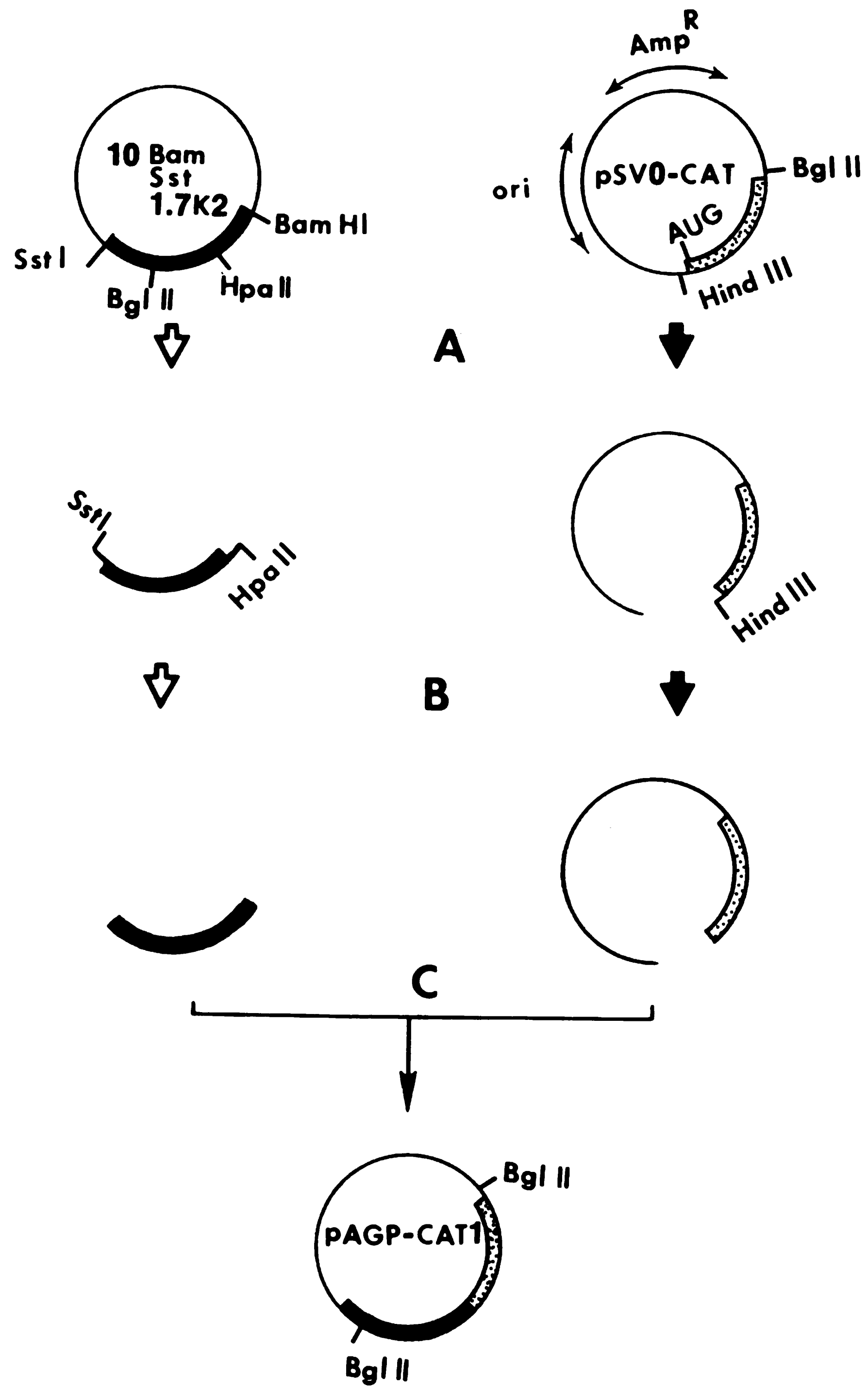
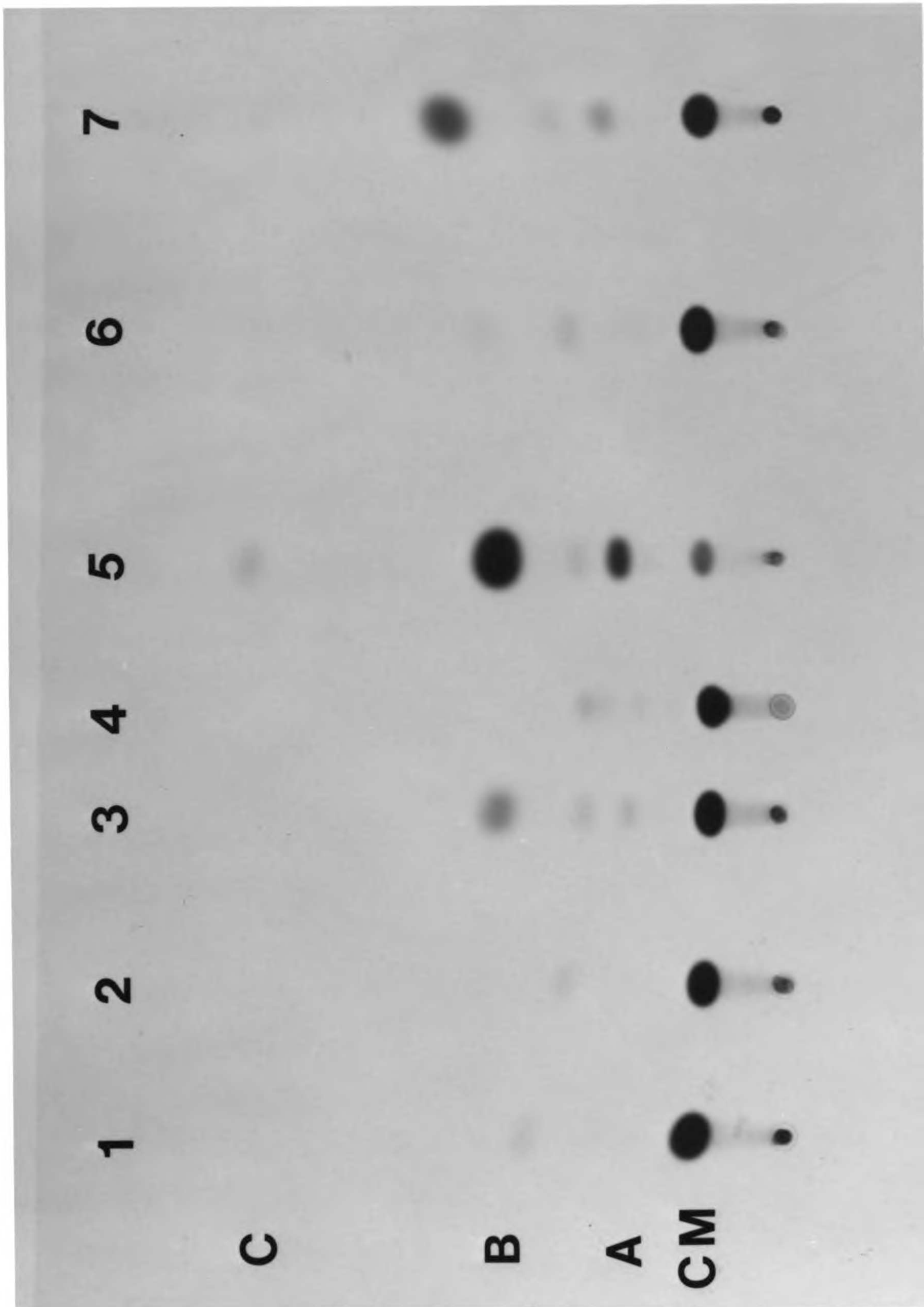

10 Bam
Sst
1.7K2
Bam HI
Sst I
Bgl II
Hpa II
Amp^R
ori
pSV0-CAT
Bgl II
AUG
Hind III
A
SstI
Hpa II
Hind III
B
C
pAGP-CAT1
Bgl II
Bgl II

Fig. 5.2. Assay of CAT activity in mammalian cells. 10 μg of pAGP-CAT 1 DNA was applied to 5 x 10^5 rat HTC cells plated 24 hours before a density of $10^4/cm^2$. After 48 hours (except lane 7 which is 24 hours), cell extracts and reaction procedures were carried out according to the text. Chloramphenicol (CM) and its acetylated form ([A] 1-acetate chloramphenicol, [B] 3-acetate chloramphenicol, and 1-3-diacetate chloramphenicol) were detected by autoradiography according to their increasing mobility due to tthe presence of acetate group.

(Lane 1) The chloramphenicol standard. The remaining lanes show the result after incubating chloramphenicol with (Lane 2) extract of rat HTC cell transfected with pSVO CAT which contains CAT gene but no promotor. (Lane 3) extract of rat HTC transfected with pAGP-CAT 1 (Lane 4) extract of rat HTC cell transfected with pSVO CAT under glucocorticoid induction. (Lane 5) extract of pAGP-CAT 1 transfected cell under glucocorticoid induction. (Lane 7) extract of pAGP-CAT transfected cell under steriod induction for only 24 hours.


1
2
3
4
5
6
7
C
B
A
CM

5.4 Discussion

In order to identify and measure the promotor activities of regulatory elements of certain genes, it is desirable to isolate and transfect genes into another cell to measure expression under induction. In the case of the rat α_1-AGP gene, this has been carried out by transfection of the whole gene into the mouse LTK(-) cell (Rosenmary et al., 1985). The result has shown that the gene is marginally inducible at least at transcription level in mouse LTK(-) cell under steroid hormone induction. The problem of this approach is that a non-parental cell has to be used in order to avoid background. The usage of a non-parental cell (a mouse fibroblast cell) may introduce a problem of incompatibility, since α_1-AGP is a serum protein and, its expression has been known to be limited in hepatocytes. Therefore, the use of fibroblasts or other cells as expression hosts raises questions, although fibroblasts do contain glucocorticoid receptors. In addition, transfection of entire gene does not identify the exact location or regulatory element. These problems are compounded by the long period of time involved in selecting and growing the correct transfectants and lead to the usage of different strategies to study the rat α_1-AGP gene.

One possible solution of the above-mentioned problem is to combine the putative regulatory region of the cloned gene of interest with a second gene which may provide an easily assayable and readily distinguished function. In this study, the putative 5' promotor region of the rat α_1-AGP gene was ligated with the CAT gene and transfected into a parental cell (rat hepatoma, HTC) to measure the strength of transcriptional promotion.

The results indicated that 5' promotor regions of the rat α_1-AGP gene contain strong promoting activity in contrast to what has been previously reported. Previous studies indicated that the expression of the rat α_1-AGP gene is only marginally increased at transcriptional levels under steroid hormone induction

(Vannice et al., 1984). It has been postulated that the induction is probably due to secondary effect of steroid hormone. This study, however, has shown that the α_1-AGP gene does contain putative steroid hormone receptor binding sequences and the 5' promotor region is playing a significant role in steroid induction.

The fact that transcriptional induction of the 5'-end-α_1-AGP-CAT hybrid was not abolished by the presence of cyclohexamide suggested that the postulated labile protein factor is not related to a transcription event. If this factor does exist, it must be involved in the maintenance of mRNA stability or other permissive events. The maintenance of stability may not be limited to mRNA but also possibly at the nuclear RNA level. The rapid turnover rate of nuclear RNA seems to provide a better site of regulatory mechanism. The previous measurement using the rat α_1-AGP cDNA probe to measure the increase of transription did not rule out the existence of nuclear RNA, since both nRNA and mRNA would hybridize positively. Furthermore, nuclear RNA with multiple exons is capable of absorbing multi-copy cDNA probes and might show false positivity at a higher degree.

The result of the 40- to 50-fold induction of CAT activity is very significant. With the existence of putative glucocorticoid receptor binding sequences, it argues strongly that the rat α_1-AGP gene probably responds to steroid directly by ligand-receptor-complex activation. However, several issues remain to be resolved. In rat liver, it has been documented that the rat α_1-AGP mRNA increases 90- to 100-fold after subcutaneous injection of turpentine, and in the hepatoma cell, the mRNA increased at least 100-fold (500-fold was recorded but unpublished by Vannice et al.) after glucocorticoid induction. The possible explanations for the differences between Vannice's 100-fold and 40- to 50-fold increases noted in our study may be due to the existence of an original

α_1-AGP gene copy which exerts an equal competition for glucocorticoid receptor binding. If they do compete equally, one half decrease (if 100-fold increase is used for comparison) seems to be an appropriate estimate. On the other hand, the CAT assay measures the translation level instead of mRNA level, therefore, if translation efficiency is not in parallel with transcription efficiency, the difference can also be explained. A measurement of CAT mRNA seems to be an alternative to verify this point.

The 5' promotor region may not contain all regulatory elements and certain other features including sequences involved in mRNA stability. Transcription promotion may also exist in other areas including introns or 3' end areas which were not involved in CAT subcloning. As mentioned previously, two additional glucocorticoid receptor binding sites were found within introns, and one of them is associated with peculiar Z-DNA-like sequences which were postulated to be involved in transcriptional activation. Attention is also drawn to the Stem-loop-structure at the 3' non-translation region. It is not known whether these sequences are involved in mRNA stability or more efficient termination of transcription which in turn led to faster turnover of RNA polymerase II. Subcloning of these interesting sequences would be necessary to answer this question.

In summary, this experiment proved that the 5' end promotor region does contain promotor activities for the expression of the rat α_1-AGP under the influence of steroid hormone, and the presence of one putative steroid receptor binding sequence further indicates that the rat α_1-AGP gene is directly regulated by steroid hormones. This conclusion, however, does not rule out the existence of other regulatory pathways. It is not uncommon that the regulation of gene expression occurs at multiple levels. The fact that cyclohexamide was able to abolish the accumulation of the rat α_1-AGP mRNA but not the MMTV mRNA (Vannice et al., 1984) indicated that the additional regulatory event exists. Also,

the 5' end promotor region may not be the only area of induction. The potential role of the glucocorticoid receptor binding hexanucleotide in addition to its close association with the unusual Z-DNA structure in the fifth intron remains to be explored.

With the data derived from this experiment, we may predict that the rat α_1-AGP gene contains a 5' end promotor region which responds to direct glucocorticoid induction. The rate of transcriptional induction is about 40-fold in comparison to the basal level and is not influenced by the presence of cyclohexamide. This result in combination with the previous observations of α_1-AGP mRNA accumulation inhibited in rat hepatoma cells which are undergoing cyclohexamide treatment indicates that certain labile protein factor(s) must be involved in post-transcriptional maintenance of the stability of α_1-AGP mRNA or nuclear RNA. It is not known how a labile protein factor exerts its vast influence on mRNA or nRNA stability, but it may act via the recognition of specific sequences of exons, 3' untranslated regions or even introns. A systematic deletion and reconstruction of new recombinants may unravel this mystery.

APPENDIX A

Solid Media

NZCYM top agar:

NZCYM broth	1 l
Bacto-agar	7 g

NZCYM plate:

NZCYM broth	1 l
Agarose (low EEO)	15 g

NZCYM top agarose:

NZCYM broth	1 l
Agarose	7 g

NZYM top agar:

NZYM broth	1 l
Bacto-agar	7 g

NZYM plate:

NZYM broth	1 l
Agarose (low EEO)	15 g

NZYM top agarose:

NZYM broth	1 l
Agarose	7 g

Liquid Media (essentially described by Maniatis et al., 1983)

NZCYM broth:

N Z amine	10 g
NaCl	5 g
Yeast extract	5 g
Casamino acids	1 g
$MgSO_4 \cdot 7 H_2O$	2 g

add H_2O to 1 l, adjust pH to 7.5 by NaOH

NZYM broth:

Identical to NZCYM except that casamino acids are omitted.

LB (Luria-Bertani) broth:

Bacto-tryptone	10 g
Bacto-yeast extract	5 g
NaCl	10 g

add H_2O to 1 l adjust pH to 7.5 with NaOH

Sequencing Gels

40% Acrylamide Mix:

Acrylamide		194 g
methylene-bis-acrylamide		6.70 g
10X TBE		100 ml
H_2O	to	500 ml

10X TBE:

Tris base	54 g
boric acid	27.5 g
0.5 M EDTA (pH 8.0)	20 ml

To make a 5% sequencing gel:

40% acrylamide mix	6.3 ml
ultrapure urea	24 g
H_2O	20.5 ml
10X TBE	5 ml
10% ammonium persulfate	0.4 ml
TEMED	20 μl

SM buffer (for storing lambda bacteriophage):

NaCl	5.8 g
$MgSO_4 \cdot 7 H_2O$	2 g
1 M Tris . Cl (pH 7.5)	50 ml
2% gelatin	5 ml

H_2O up to 1 l autoclaved

Solutions

20X SSC:

NaCl	175.3 g
Sodium citrate	88.2 g

add H_2O to 1 l

adjust pH to 7.0 by ION NaOH

20X SSPE:

NaCl	174 g
$NaH_2PO_4 \cdot H_2O$	27.6 g
EDTA	7.4 g

Adjust volume to 1 l with H_2O

Adjust pH to 7.4 with NaOH

100X Denhardt's Solution:

Ficoll	5 g
Polyvinylpyrrolidone	5 g
BSA	5 g
H_2O to	500 ml

Filtered through a disposable Nalgene filter

<u>10X TE:</u>

(pH 7.4)

100 mM Tris . Cl (pH 7.4)

10 mM EDTA (pH 8.0)

(pH 8.0)

100 mM Tris . Cl (pH 8.0)

10 mM EDTA (pH 8.0)

<u>10X STE:</u>

100 mM Tris . Cl (pH 8.0)

100 mM NaCl

1 mM EDTA (pH 8.0)

2X proteinase K buffer:

0.02 M Tris (pH 7.8)

0.01 M EDTA

1.0% SDS

APPENDIX B

I. Methods of DNA sequencing (M-13, chain termination procedure)

This procedure describes the sequencing of DNA fragment using M-13 phage from ligation through gel drying.

1.0 POLYLINKER CLONING SITES OF M-13 VECTORS

2.0 MATERIALS

2.1 10 x Reaction Buffer (10xRB)

70 mM Tris, pH 7.6

50 mM β-Mercaptoethanol

550 mM Sodium Chloride

1 mM EDTA

2.2 Dilutions

	Microliters			
	G	A	T	C
0.5 mM dGuanosine triphosphate	20	200	200	200
0.5 mM dAdenosine triphosphate	200	20	200	200
0.5 mM dThymidine triphosphate	200	200	20	200
10 x RB	100	100	100	140
dideoxy-X				
N mix	100	100	100	100
1 mM dideoxy-X	25	25	55	20
Water	75	75	45	80

*For reading bases greater than 400 from cloning site, halve ddX in mix.

**Adjust as needed for each batch of ddNTPs. This is easily done empirically, i.e., mix approximately 20 microliters, ddX mix, run reaction, adjust as necessary.

***For inosine mixture, dGuanosine triphosphate (0.5 mM) is replaced by inosine triphosphate (2.5 mm)

2.3 Isotope PB10165 from Amersham, 400 Ci/mmole, 10 M Ci/ml [$d^{32}P$] dCTP

3.0 LIGATION

3.1 For cohesive-ended fragments, molar ratio target/vector = 50/1; for blunt ends, 500/1.

3.2 Ligations are in 10 microliters:

1 microliter vector at 10 ng/microliter

1 microliter 10 x ligation buffer

1 microliter 10mM rATP

6 microliters (Target DNA and water)

1 microliter T4 DNA ligase

3.3 Allow to ligate for 2 hours or more for sticky ends, 4 hours or more for blunt ends at ambient temperature.

3.4 As a control, always do a ligation with vector alone. This gives the background.

4.0 TRANSFORMATION

4.1 Competent cells are prepared and kept in frozen state. Grow up approximately 1000 ml of JM101 to 0.4-0.6 O.D. in 2 YT Broth.

4.2 Spin down bacteria for 5 minutes at 5 K rpm. Save pellet. Resuspend in 100 ml 100 mM $CaCl_2$. Incubate in ice for 20-60 minutes.

4.3 Spin 5 minutes at 5 K rpm. Resuspend in 100 ml $CaCl_2$ then freeze in liquid nitrogen and store in -70°C until needed. When needed thaw out in room temperature for 5-10 minutes.

4.4 Place in sterile culture tube with 5 microliters of ligation mixture and 300 microliters competent cells. Incubate on ice for 40 minutes.

4.5 Heat shock transformations for 2-3 minutes at 45°C, return to ice for a minute or two. Add 10 microliters 200 mM IPTG, 50 microliters 20 mg/ml X-Gal in DMF. Place at room temperature and add 300 microliters of non-competent cells. Add 3.5 ml 2YT top agar. Vortex. Plate on LB agar plate.

4.6 When solidified, invert plates, and incubate overnight at 37°C.

5.0 TEMPLATE PREPARATION

5.1 Compare plate from ligation with appropriate control. Controls should be almost all blue plaques for single-cut vector; only a few white for double-cut. This is optimum. There should be more white plaques on the ligation than on the control.

Decide on an appropriate number of plaques to pick, i.e., if you expect one orientation of a single fragment, pick 3. If you expect two orientations of a single fragment, pick 6 and so forth. This must be adjusted according to the number of whites as compared to the control. For instance, if the ligation has 100 white plaques and 10 blue plaques, and the control has 10 white plaques and 10 blue plaques, this is ideal. If the ligation has 100 white plaques and 10 blue plaques and the control has 50 white plaques and 10 blue plaques, it may be advisable to pick more.

5.2 Touch a sterile toothpick to a plaque, then touch it to the surface of 1.5 ml of 2YT which has been innoculated with a 100 microliter log phase culture of JM101. Grow for at least 10 hours or overnight with shaking at 37°C in a culture tube.

5.3 Transfer completely to Eppendorf tube. Spin in a microcentrifuge for 5 minutes.

5.4 Pour approximately 1.2 ml of supernatant into new tube. Save the pellet and remaining supernatant to be used if it is necessary to prepare the template again.

5.5 Add 200 microliters 2.5 M NaCl, 20% PEG, mix by inversion until thoroughly mixed. Incubate at room temperature for 10-15 minutes or in ice for 5 minutes.

5.6 Spin in a microcentrifuge for 5 minutes. Aspirate off supernatant, vortex tubes and aspirate supernatant again. Resuspend in 200 microliters of TE; vortex. Heat 5 minutes at 55°C.

5.7 Add 100 microliters phenol, vortex. Add 100 microliters $CHCl_3$, vortex hard. Let sit 5 minutes. Vortex hard and spin for 4 minutes.

5.8 Pipet 100 microliters of aqueous phase to a new tube. Add 200 microliters in sodium acetate pH 4.5 and 400 microliters of ethanol, vortex and freeze 5 minutes on dry ice. Spin for 5 minutes.

5.9 Aspirate off supernatant. Add 1 microliter 70% ethanol (-20°C). DO NOT VORTEX. Spin for 5 minutes. Aspirate off supernatant and dry.

5.10 Resuspend in 15 microliters TE. This template may now be used for sequencing.

6.0 TRACKING

6.1 Make up primer mix for 1 set of 4 reactions (11 microliters), as follows:

1 microliter primer at 1 pM of 17-mer per microliter

1.5 microliters 10 x RB

1.5 microliters 35 mM $MgCl_2$

2 microliters 10165

5 microliters H_2O

6.1.1 For GATC reactions, make 11 microliters/template.

6.1.2 For C-tracks, make 11 microliters/4 templates.

6.2. Pipet 1 microliter template into reaction tube. Add 2.5 microliters primer mix. Heat tube at 55°C for 5 minutes to anneal primer and template.

6.3 Make ddC+ enzyme - 8 microliters ddC mix and 1.9 unit enzyme/4 tubes.

6.4 Spin droplets in reaction tube down. When tube has cooled slightly, add 2 microliters ddC+ enzyme and spin it down. Incubate for 15 minutes at room temperature.

6.5 Add 1 microliter chase mixture (2.5 mM of all 4 dNTP's) and spin down. Incubate 15 minutes at room temperature.

6.6 Add 6 microliters formamide/dye mix (96% formamide, 0.1% xylene cyanol, 0.1% Bromophenol Blue, 10 mM EDTA) and spin down.

7.0 FOUR-TRACK SEQUENCING

7.1 Make sufficient primer mix, 10 microliters/template (see 6.1).

7.2 Pipet 4 microliters template into annealing tube . Add 11 microliters primer mix. Incubate at 55°C for 5 minutes.

7.3 Meanwhile, set up reaction tubes. Add 2 microliters of appropriate ddX mix to each tube.

7.4 After annealing, allow primer/template mix to cool to room temperature. Add 1 microliter of 1 unit Klenow fragment.

7.5 Pipet 3.2 microliters primer/template and enzyme mixture to each reaction tube. React at room temperature for 15 minutes or 25 minutes if low ddX mixers are used. React at 37°C, if inosine mixture is used.

7.6 Add 1 microliter chase (2.5 mM all dNTP's, 0.5 units enzyme/4 tubes, 7 mM $MgCl_2$). React 15 minutes more.

7.7 Add 6 microliters formamide/dye mix.

8.0 GELS

8.1 Equipment

1 very clean backing plate (no chromorage required, 8"x 18-1/2".

1 very clean, freshly silated notched plate (clean the plate with a few microliters of $CCl_4/(CH_3)_2\ Cl_2Si$ 100/5).

2 spacers, 0.014" mylar.

Polyethylene tape, made by 3M 1" wide.

6 binder clamps, medium size (2" wide).

Combs - Shark's tooth type, with 3 or 4 mm slots, can be hand made from mylar sheet.

8.2 Clamp plates together with spacers down sides. Tape edges, making sure there are no channels for acrylamide to ooze through. Insert combs.

8.3 Clamp aluminium plates (1/8" thick) to backing plate. Avoid contact with buffer.

8.4 Heat sample at 100°C for 2 minutes and immediately place on ice before loading. If 2 loadings are needed, split each sample and keep in ice until needed. Sample will degrade slowly after heating even at -20°C. (It should not be kept for more than 3 hours).

8.5 Load gel using a 10 microliter Hamilton Syringe, with needle ground down to fit between plates. Load 1 microliter/4 mm slot GATC reactions for 1 Bromophenol Blue runs, 1 microliter/slot for longer runs. Load 2 microliters/slot for C-tracks.

8.6 Reading basics:

8.6.1 To read less than or equal to 300 bases from cloning site, run 5% gel, running Bromophenol Blue to the bottom.

8.6.2 To read 150–400 bases from cloning site, use 5% gel, running Xylene Cyanol off by 10 cm.

8.6.3 To read 350–600 bases from cloning site, use 5% gel, running Xylene Cyanol 2 times the length of the gel (reload a marker).

8.7 Dry gel onto 3M paper. Discard the top 3 cm of the gel. Dry the gel in a thermostatically controlled, Hoeffer Scientific drier which runs off a vacuum line (use a trap). The sequence gel should dry in 45 minutes.

8.8 Expose 12-16 hours, without an intensifier screen at room temperature.

8.9 This procedure is modified from Genentech's laboratory methods.

II. Procedure of small scale DNA preparation from lambda bacteriophage

1. Mix 250 microliters DP50 E. coli bacteria and 25 microliters 10 mM $MgCl_2$ and $CaCl_2$ with 100 microliters phage stock. Incubate at 37°C for 5 minutes. Add to 40 ml NZYDT broth at room temperature and incubate at 37°C for 16 hours with shaking. The culture should be just slightly turbid from lambda resistent cells, yet containing lysed cellular debris.

2. Transfer 40 ml to 40 ml-Oakridge tube and spin 10K for 10 minutes.

3. Transfer supernatant to another Oakridge tube and add 12 microliters DNase (10 mg/ml), 3 microliters RNase (10 mg/ml) then incubate at 37°C for 45 minutes.

4. Add 5.5 ml 20% PEG and 2.5 M NaCl. Mix well, stand 15 minutes at 0°C, and spin at 15 K for 20 minutes, discard and drain all supernatant.

5. Gently resuspend pellet at 500 microliters proteinase K buffer and transfer to Eppendorf tube. Spin 3 minutes in benchfuge and transfer supernatant again. Add 2 microliters of Proteinase K (10 mg/ml) and incubate at 37°C for 45 minutes.

6. Add 250 microliters each of phenol and chloroform. Vortex hard and spin 3 minutes in benchfuge. Remove the aqueous layer and repeat the extraction one more time.

7. Add full volume chloroform, vortex and spin, transfer aqueous (3x). Add 0.1 volume of 3 M sodium acetate and 2.5 volumes of ethanol. DNA should fall out of solution as a cottony ball. Using a metal needle, carefully lift this ball and transfer to another tube containing 70% ethanol. Spin down the pellet and remove ethanol.

8. Add 500 microliters 70% ethanol and spin again for 3 minutes. Remove ethanol and dry in spin-vac.

9. Resuspend in 200 microliters TE (This may require heating the tube at 55°C for a few minutes). Measure O.D.260.

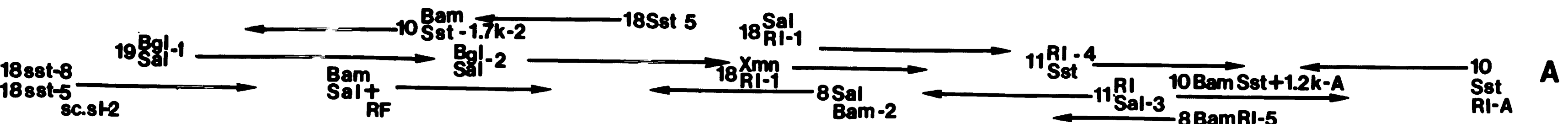
10. This procedure is modified from Genentech's laboratory methods.

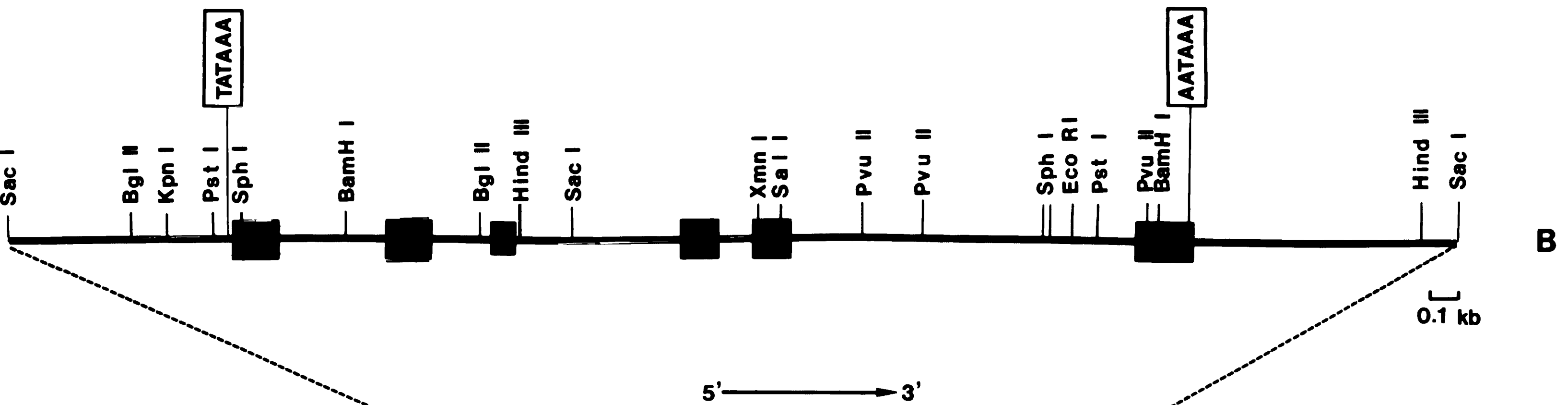
APPENDIX C

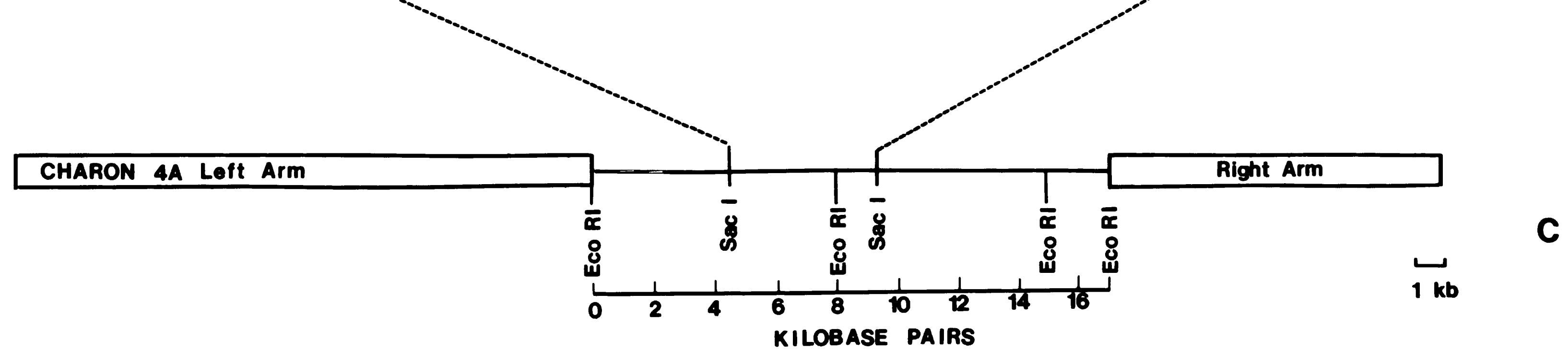
I. Map of the rat α_1-AGP gene.

(A) The strategy of sequencing. Arrows indicate the directions and lengths of M-13 subclones used in dideoxynucleotide sequencing. The names represent the M-13 clones listed at the beginning of arrows. For example, 18 Sst-8 means a clone of number 8 using mp 18 as vector cloned in Sst I site. Clone 19 Bgl.Sal-1 means that it was the clone number 1 derived from using mp 19 vector with Bgl II toward Sal I insert. Clone 18 Sst-5, sc.sl-2 is identical to clone 18 Sst-8 but generated earlier and differently. It was derived from cutting the double strand 18 Sst.5 (self-cut, sc) then ligated back (self-ligation, sl). The purpose is to generate a clone with opposite orientation for sequencing. Clone 10 Bam Sst-1.7k-2 and 10 Bam.Sst+1.2k-A are clones derived from using mp 10 vector with Bam HI and Sst I site to absorb purified DNA fragments of 1.7 kb and 1.2 kb, respectively.

Rat Alpha 1-Acid Glycoprotein Gene







II. DNA sequence of the rat α_1-AGP gene.

Exons are flanked with sign, and translational codons are printed as triplets and lined with their designated amino acids. The "TATA" box and polyadenylylation signal (AATAAA) are boxed. A single asterisk (*) indicates the beginning of transcription and triple asterisks (***) indicate the codon (TAG) of translational stop. The hexanucleotides T-G-T-T-C-T of potential glucocorticoid-receptor binding sequences are underlined with a solid line. Palindromic sequences in intron 4 are marked by arrows and dots; dinucleotide G-T repeats followed by poly(G) are marked by arrowheads and open squares. $G(T)_{2-5}$ repeats following the fourth hexanucleotides T-G-T-T-C-T are underlined with open circles and dots. The Mnl I-rich area is underlined with a broken line and its following complementary sequence is marked with a broken line at the top.

5'GAGCTCCCAGCAATGGCTCCATTCCTCAGTGGGAGTCACCATGCCTGCCCATAACAATACTTCCCGAGAGTGTTCATTCATGTGGACATAGACCAGACTGTTATCAGGGAGGAATTAGAAC 121

CAGATATGCTGGACACAGATCCTGGATTGGGCACACATTATCTTGTCGGATTCTTTCCTGCTGTAAATACTGGGAGGTTTGCTCAACTCAGATTCACCCCTCTTTGCTGGGCATTTGACTG 242

GTTTCAGGGCTGGGTAGGGGTCCCAAGGAGCATCACACTCACTTAGAAACTAAACCATCTTTGTTCTTGGCCCTCAACTTGAGCCCCTCAGTGTCTTCTAAGTTTACTAACAACCTGACCT 363

GGGACCTCTTCATAACTCCCTCAGGGGTTGATTTATGTGTAGATCTTACTCCTGCTAGGCAGCCCTCCCAGGGCTAGGAGGCAGATGAGACCTATATTACAATCTCTGACTGGTACCATGT 484

CAAGTTTCATCAGGCTAGCCTTGTACCCATCTTGGCCATGAATCAGCCACTCTGGTTTTCAAGCAGCTTATAGGGCAGGAGCCTGTGTCAGGGCTGGCTGGCTGTGGGAACAGCCCAGCCC 605

AAAGCTGGCTTGAGGGAACATTTTGTGCAAGACATTTCCCAAGTGCTGGTGAGATTGTGCCACAGCTCTGCAGCCCCTGGCTGCACGCCCTTCCCACACCTTGTTATAAAAGTCACTGCAC 726

Met Ala Leu His Met Val Leu Val Val Leu Ser Leu Leu Pro Leu Leu
TCTCCAGCCACCAGTT AGCTCTTCCTGGGCCGGTGCCTCTGAGTGTCTTCGGC ATG GCG CTG CAC ATG GTT CTT GTC GTT TTG AGC CTC CTG CCC TTG TTG 827

Glu Ala Gln Asn Pro Glu Pro Ala Asn Ile Thr Leu Gly Ile Pro Ile Thr Asn Glu Thr Leu Lys Trp
GAA GCT CAG AAC CCA GAA CCT GCC AAC ATC ACC CTA GGC ATA CCT ATT ACC AAT GAG ACC CTG AAA TGG GTGAGTGCCTGCCTGGGGCATGGACCTGA 925

CCATTGTAGGTGCCTTTTTTCCCTCTGGGCTTTCCTTTCCCTGTCTGTGTTCCTCTCTGGGCTCCCGGTGCTGCCCTTCCCTCTCCTGGGGATTGTTGTGGCATCCCCCTGCCTCCAAACC 1046

CAGAAGCATCACTCTGAGGTCTCTCGGCAGGGGACAGAGTCATAGGATGGTTGACTTCTGAGTTGTCCATGGGATCCTGGCCTGAAACTTCAGTCTTCCCATGTATGGCTCAAACGCCATC 1167

Leu Ser Asp Lys Trp Phe Tyr Met Gly
ACTCACCTGCCCCTCATTGTTGCTAACTGCCAAGCTCCATACTCCAAACCCTGTTCATCTGCGTGCCTGCTTCTTTCTTCCCAG CTC TCA GAC AAA TGG TTT TAC ATG GGA 1278

Ala Ala Phe Arg Asp Pro Val Phe Lys Gln Ala Val Gln Thr Ile Gln Thr Glu Tyr Phe Tyr Leu Thr Pro Asn Leu Ile Asn Asp Thr
GCA GCT TTC CGA GAC CCC GTG TTC AAG CAG GCA GTT CAA ACG ATA CAG ACG GAA TAT TTT TAC CTT ACC CCC AAC TTG ATA AAC GAC ACA 1368

Ile Glu Leu Arg Glu Phe Gln Thr Th
ATT GAA CTT CGG GAG TTT CAG ACC AC GTGAGTCCTTGCAGCAGCCAGCCCACCCCAGCCCTGGCCTCCAATCCCTGGGGCCTAGAGACCTGAACAAACTGCCCCTGCCTGGC 1480

CTCCCCAGCCGCCTTAGGAATCGGGAGGGCTGTTCTTGTCTACCTGCCTCTTGCCCACTCTCACCCTGTGCTGGTCAGATCTCCTCTCTCTCTGTTACCCAACTTCTCTTTGCTTTATAG 1600

r Asp Asp Gln Cys Val Tyr Asn Phe Thr His Leu Gly Val Gln Arg Glu Asn Gly Thr Leu Ser Lys Cys A
A GAC GAC CAG TGT GTC TAT AAC TTC ACC CAT CTA GGA GTC CAG AGA GAG AAT GGG ACC TTA TCC AAG TGT G GTGAGAGTGTGAACTCGAAGCTTTC 1696

TGGGGGGAGGGGGGGGTCATTGCAAGGACCTGAGGGGGAAGCAGAGCAGGGCAGTTTAGTCTGTTAGGTCAGCTCTGAGGCTGGTAAGTAGAAGCAGATTTCAAACCAGAGTCCAGCTCCC 1817

AGCCCAGTCTCGCTTTTAACCATTCTTTATGATCCTTTGTGAGCTCAGCGTGGGAACCCATGGGATGGGAAGCCAGTGACTTCAGCCACATCTCCCATGTCACCAGGGACAACACATAGGA 1938

AGGCAACCAAACACTGGTGGCCAGATCAGCCTCTCCTTAGCTTTGAATCATCCTAAGGGGATGTTTTAATGTTTAAGGGGAAGTTTTAACTGTGCCCACCACCCCACCACCAAGGGCCCTT 2059

TCCAAGGCTCCAGCCGTTCAGTATCTATAAAGAATGAGAAGTTAGGCTGGCACCCAACAGGGACTGAGCACATGGTATCCTCAGGGAGCCGCAGGCATTTGCCATGGGGACCACGGCACCA 2180

la Gly Ala Val Lys Ile Phe Ala His Leu Ile Val Leu Lys Lys His Gly Thr Phe Met Leu Ala
TTCTAGTGTCTTACAAAGCCCTTCCTTCCGCAG CA GGA GCA GTG AAA ATC TTT GCC CAT TTG ATA GTG CTG AAG AAA CAT GGG ACC TTC ATG CTT GCC 2278

Phe Asn Leu Thr Asp Glu Asn Arg Gly Leu Ser Phe Tyr A
TTT AAC CTG ACA GAT GAG AAC CGG GGG CTG TCC TTC TAT G GTAGGCTCCTCAGAACCCTGGACCCCAGTTCTGCTGGGCCTTACCCTTGGTCCCCCACCCACATCCC 2385

la Lys Lys Pro Asp Leu Ser Pro Glu Leu Arg Lys Ile Phe
TATCTGGTCTCCCATTTGCTGGACCCTTACCCAGAACACACCCCACCCCCTTCTCCCTTGCAG CT AAA AAG CCA GAC TTG TCC CCA GAG CTG CGG AAA ATA TTC 2489

Gln Gln Ala Val Lys Asp Val Gly Met Asp Glu Ser Glu Ile Val Phe Val Asp Trp Thr Lys
CAG CAG GCT GTC AAA GAT GTG GGC ATG GAT GAA TCA GAA ATC GTA TTT GTC GAC TGG ACA AAG GTAAGCGACGAGGCTGCATGATGCCACCCTGGCAGTG 2589

GTCCCGTTTTAGTGTCCCTAGAGGCTCAGAGAGGGTCAGCTCCTGGTTAAGGCAGCACAGCAAGGCGGGTATCTTGCTGTTGAGAACGCCCGTGCCTACTGTGACAGCTGAGATTCACAGA 2710

GATCTTAGATGGGCAGTGAGCCTGAGGTGGGTGGGGTTAAGTAGGAGATACTAGAAAGACTGGAGGATCAGGGGCGAGACCAGGAAGGGTGATTGAGGGAACAGCTGGTGCCAGAGTCCCA 2831

GGTTTCTCCCTGCCTTTCATTTTCACTTCCTATACAAGCCAGAGTTGTGACTAAAGTCTCCCTGTGCCAGCAAGCCCAATGGCCTTGGGGATGGGGGGGGGTGCCCACACACCAAGTCATC 2952

TTGATGCTCTGTCTAAATTTGTGCAGAGGGAAAGGCCATCAGCTTTGGCTTTACCCCCAGCTGAGGCCACAGGATCAACCAATCAGCTTTCGACTTAGTGCGAGACACACTTGATATTTTG 3073

TGACTCCATTCTCTGGGCACCTGAGCCTTTCCTGACATCACCAGCCCCCCAGGTTCCCAGGGGAAAGGTGTCTGCATACAGACACTGCCATTGAGCTGTTTTGCAACCTGCCTTCCCTGCC 3194

TAGCCAGTGGGGTCAGTAAATTCACCCTGTACCTGTAGGCAGAATTGTTCAGAAAACGCCTCCACTTGGGGGACTCTGTTCTTCCATTCATTGGAATAATAAGCTGGAACTTAGAGACCAG 3315

TGGTTCTCAAGGCTGACACTATCCTGCCCTTTCAGGTCAACCACCCCCAGAAATGTATTCATTGCTAGTGTGTGTGTGTGGGGGGGGGGTGGGCATGCCCATGCACAAGCATGCCATAAGA 3436

CGTGTGTAGATGTCAGAGGACAACTGTGGGGCGTGTCCTCTTCCGACATTCTCTCATGGGCTCTGGGGATTGAATTCAAGCTGTCCTTCAGGATTACCTACTATGCCAGGGTTTTTTGTTC 3557

TGTTTTGTTTGTTGTTTTGTTTTTGCATCTCACCCTGCAGCCCAGGTTACCCTGCAACTCATAGCAATCCTCCTGCTTCAGTATACTCCAACCCAGATGTGTGCCACCACCCCTGGCTCCG 3678

Asp Lys Cys Ser Glu Gln Gln Lys Gln Gln Leu Glu Leu Glu Lys Glu Thr Lys Lys Glu
TCACTTTGTCACTTTGCTCTCTTCCCTCGTATCTTCTAG GAT AAG TGC AGT GAG CAG CAG AAG CAG CAG CTG GAG CTG GAG AAG GAG ACT AAG AAG GAG 3777

Thr Lys Lys Asp Pro ***
ACC AAG AAG GAT CCT TAG GCCAGGCATGAACTCAGCTCTCTGAACTCCGGGGACTGTCCCCATGCCCACTCTACCCCCACTCCTGTGCACCTCGATTCTATTTTCCACAATAATA 3892

AAGGTTTGGTTAAACAATC AGTATCACTTCTTTGTTCCCTCCCTCCCTTCTTTCCTTCCTCCCTCCCTTCCTGGAGAGGAGGACTGCCATGAGTGGAGACTTAGCTAAGTCCAATAAGCC 4012

CTACCCAGGGAATCTGAGAGGGCTTCCTGGAAGGGGCATTTGAGTCAGGCTAGACGGAACTGATAGAACCTGATTATAGCTCTTGCTATAATCCCCAAACCCTAGAGGGCACATTTGGTGA 4133

AGACAAAGTAGAAGACCCAAGCAGCGGTGACTGGGAGTGGGGAGCCCAGGTGTGCCGTTCCTGTCACAGGGGCCTTCACTTCAGATGTGCCATTGGAACTCTGGAAGGGAGGTGAGTTGGG 4252

AGACAGGAAGGCGTCGATGCTAGCTATGGTCACGGTATAAAAGGCCAAGGCAAGTGGCATAGACTTAGGGCTAGACTAGAGGGGTTGGATATCCCCAGGAAGAGAACTCAGTGGTCCGTGT 4375

GGGACACCGTGAGATGACACTGCTTCCCTAGAGGCCAGTTGACATCTATTTCACAATTGTGTTCAAAGTCCTGAGGCACAGTCACTCAGTGGCAAACACAGGGTCTGAGATAATCTTTACA 4496

CTTGACTTGATTGGGCAGTCACCTAAGGGATGTGCTCGGTGTGTCTCTGTAAGGACATTTCCACAGAGAGTCACCTGCCTTGACTGTAGGCAGCGCCATCCCTTGGGCTTGGTCCCCAAT 4617

TAACAAAAGGAGAAAACTAGCTGAGCACCCACAGTTCATCTGTCTGCTAAGCTTACGACTGTCCCAGTTGGATACAGTATGACATGGGCAGGTGTCCCTCCTCACCCATGAGCCCCACTTC 4738

AATAGCTTTCGTCAACTATTTTGTCACAGAAGGGGGAAAGCAGTTCTTATGGAGCTC 3' 4795

III. The complete nucleotide sequences of the rat α_1 gene with comprehensive restriction sites.

```
     aluI
    sacI
    hgiAI
    bsp1286                   ddeI        hphI
    banII     nlaIV     mnlI      hinfI  nlaIII                    avaI             nlaIII
  1 GAGCTCCCAG CAATGGCTCC ATTCCTCAGT GGGAGTCACC ATGCCTGCCC ATAACAATAC TTCCCGAGAG TGTTCATTCA TGTGGACATA GACCAGACTG
    CTCGAGGGTC GTTACCGAGG TAAGGAGTCA CCCTCAGTGG TACGGACGGG TATTGTTATG AAGGGCTCTC ACAAGTAAGT ACACCTGTAT CTGGTCTGAC

                                                 scrFI
                                                 bstNI
                                             sau3AI
                                             dpnI
            mnlI                            xhoII        bsp1286               hinfI                      mnlI
101 TTATCAGGGA GGAATTAGAA CCAGATATGC TGGACACAGA TCCTGGATTG GGCACACATT ATCTTGTCGG ATTCTTTCCT GCTGTAAATA CTGGGAGGTT
    AATAGTCCCT CCTTAATCTT GGTCTATACG ACCTGTGTCT AGGACCTAAC CCGTGTGTAA TAGAACAGCC TAAGAAAGGA CGACATTTAT GACCCTCCAA

                                                                      sau96I
                                                                      nlaIV
                   hphI                                               avaII
            ddeI hinfI   mnlI                                       nlaIV styI   sfaNI        ddeI
201 TGCTCAACTC AGATTCACCC CTCTTTGCTG GGCATTTGAC TGGTTTCAGG GCTGGGTAGG GGTCCCAAGG AGCATCACAC TCACTTAGAA ACTAAACCAT
    ACGAGTTGAG TCTAAGTGGG GAGAAACGAC CCGTAAACTG ACCAAAGTCC CGACCCATCC CCAGGGTTCC TCGTAGTGTG AGTGAATCTT TGATTTGGTA

                                                                          sau96I
                                                                          avaII
                  mnlI       mnlI                                    nlaIV                ddeI
               sau96I      bsp1286    ddeI                    scrFI    mboII         mstII
               haeIII      banII ddeI   mboII                 bstNI  mnlI            mnlI
301 CTTTGTTCTT GGCCCTCAAC TTGAGCCCCT CAGTGTCTTC TAAGTTTACT AACAACCTGA CCTGGGACCT CTTCATAACT CCCTCAGGGG TTGATTTATG
    GAAACAAGAA CCGGGAGTTG AACTCGGGGA GTCACAGAAG ATTCAAATGA TTGTTGGACT GGACCCTGGA GAAGTATTGA GGGAGTCCCC AACTAAATAC

        sau3AI                                                                        rsaI
        dpnI                                                                         nlaIV
       xhoII               fnu4HI    scrFI                                           kpnI
       bglII               bbvI mnlI  bstNI     mnlI                                 banI nlaIII             nheI
401 TGTAGATCTT ACTCCTGCTA GGCAGCCCTC CCAGGGCTAG GAGGCAGATG AGACCTATAT TACAATCTCT GACTGGTACC ATGTCAAGTT TCATCAGGCT
    ACATCTAGAA TGAGGACGAT CCGTCGGGAG GGTCCCGATC CTCCGTCTAC TCTGGATATA ATGTTAGAGA CTGACCATGG TACAGTTCAA AGTAGTCCGA

                      haeIII                           aluI
                     haeI    hinfI                    fnu4HI
          rsaI       balI nlaIII                      bbvI             nlaIV
501 AGCCTTGTAC CCATCTTGGC CATGAATCAG CCACTCTGGT TTTCAAGCAG CTTATAGGGC AGGAGCCTGT GTCAGGGCTG GCTGGCTGTG GGAACAGCCC
    TCGGAACATG GGTAGAACCG GTACTTAGTC GGTGAGACCA AAAGTTCGTC GAATATCCCG TCCTCGGACA CAGTCCCGAC CGACCGACAC CCTTGTCGGG

                                                                                                    fnu4HI
                                                                                     fnu4HI         bbvI
                                                                                     bbvI   scrFI
           aluI      mnlI                                       hphI             aluI pstI     bstNI
601 AGCCCAAAGC TGGCTTGAGG GAACATTTTG TGCAAGACAT TTCCCAAGTG CTGGTGAGAT TGTGCCACAG CTCTGCAGCC CCTGGCTGCA CGCCCTTCCC
    TCGGGTTTCG ACCGAACTCC CTTGTAAAAC ACGTTCTGTA AAGGGTTCAC GACCACTCTA ACACGGTGTC GAGACGTCGG GGACCGACGT GCGGGAAGGG

                                                                  nlaIV
                                                                  banI                              fnu4HI
                                                              haeIII                              hinpI
                                                         scrFI  mspI                              hhaI
                                                     mboII   sau96I       ddeI                   haeII
                                                  aluI   bstNI  hpaII  mnlI       mboII  nlaIII bbvI  nlaIII
701 ACACCTTGTT ATAAAAGTCA CTGCACTCTC CAGCCACCAG TTAGCTCTTC CTGGGCCGGT GCCTCTGAGT GTCTTCGGCA TGGCGCTGCA CATGGTTCTT
    TGTGGAACAA TATTTTCAGT GACGTGAGAG GTCGGTGGTC AATCGAGAAG GACCCGGCCA CGGAGACTCA CAGAAGCCGT ACCGCGACGT GTACCAAGAA
```

```
                                                                     styI
                                                                     avrII
                                       ddeI
                 mnlI              aluI                         hphI                                        hphI
 8Ø1 GTCGTTTTGA GCCTCCTGCC CTTGTTGGAA GCTCAGAACC CAGAACCTGC CAACATCACC CTAGGCATAC CTATTACCAA TGAGACCCTG AAATGGGTGA
     CAGCAAAACT CGGAGGACGG GAACAACCTT CGAGTCTTGG GTCTTGGACG GTTGTAGTGG GATCCGTATG GATAATGGTT ACTCTGGGAC TTTACCCACT

                                                                                                      mspI
                                                                                                     scrFI
                                                                                                     ncII
                       sau96I                                                                    nlaIV
            scrFI      avaII              nlaIV                                                 bsp1286     fnu4HI
            bstNI   nlaIII                banI          mnlI                               mnlI     banII hpaII bbvI
 9Ø1 GTGCCTGCCT GGGGCATGGA CCTGACCATT GTAGGTGCCT TTTTTCCCTC TGGGCTTTCC TTTCCCTGTC TGTGTTCCTC TCTGGGCTCC CGGTGCTGCC
     CACGGACGGA CCCCGTACCT GGACTGGTAA CATCCACGGA AAAAAGGGAG ACCCGAAAGG AAAGGGACAG ACACAAGGAG AGACCCGAGG GCCACGACGG

              scrFI            fokI                                   mnlI                      hinfI
         mnlI bstNI           sfaNI      mnlI                 sfaNI   ddeI                 tth111I       fokI hincII
1ØØ1 CTTCCCTCTC CTGGGGATTG TTGTGGCATC CCCCTGCCTC CAAACCCAGA AGCATCACTC TGAGGTCTCT CGGCAGGGGA CAGAGTCATA GGATGGTTGA
     GAAGGGAGAG GACCCCTAAC AACACCGTAG GGGGACGGAG GTTTGGGTCT TCGTAGTGAG ACTCCAGAGA GCCGTCCCCT GTCTCAGTAT CCTACCAACT

                            scrFI
                        sau3AI
                        dpnI
                       xhoII
                   nlaIII   bstNI
                  styI nlaIV   haeIII
        ddeI      ncoI bamHI  haeI             mboII  nlaIII                         hphI      mnlI                  aluI
11Ø1 CTTCTGAGTT GTCCATGGGA TCCTGGCCTG AAACTTCAGT CTTCCCATGT ATGGCTCAAA CGCCATCACT CACCTGCCCC TCATTGTTGC TAACTGCCAA
     GAAGACTCAA CAGGTACCCT AGGACCGGAC TTTGAAGTCA GAAGGGTACA TACCGAGTTT GCGGTAGTGA GTGGACGGGG AGTAACAACG ATTGACGGTT

                                                                                             aluI
                                                                                          fnu4HI
                                                   mboII  aluI ddeI                    nlaIII bbvI
12Ø1 GCTCCATACT CCAAACCCTG TTCATCTGCG TGCCTGCTTC TTTCTTCCCA GCTCTCAGAC AAATGGTTTT ACATGGGAGC AGCTTTCCGA GACCCCGTGT
     CGAGGTATGA GGTTTGGGAC AAGTAGACGC ACGGACGAAG AAAGAAGGGT CGAGAGTCTG TTTACCAAAA TGTACCCTCG TCGAAAGGCT CTGGGGCACA

                                      sspI                                                                   hinfI
13Ø1 TCAAGCAGGC AGTTCAAACG ATACAGACGG AATATTTTTA CCTTACCCCC AACTTGATAA ACGACACAAT TGAACTTCGG GAGTTTCAGA CCACGTGAGT
     AGTTCGTCCG TCAAGTTTGC TATGTCTGCC TTATAAAAAT GGAATGGGGG TTGAACTATT TGCTGTGTTA ACTTGAAGCC CTCAAAGTCT GGTGCACTCA

                                                     haeIII                                 mnlI
            fnu4HI                 haeIII           sau96I                                 haeIII
            bbvI                  haeI             nlaIV                                  haeI                 ddeI
         fnu4HI                 scrFI            scrFI                                   scrFI                mstII
         bbvI                   bstNI mnlI       bstNI                                   bstNI         fnu4HI     hinfI
14Ø1 CCTTGCAGCA GCCAGCCCAC CCCAGCCCTG GCCTCCAATC CCTGGGGCCT AGAGACCTGA ACAAACTGCC CCTGCCTGGC CTCCCCAGCC GCCTTAGGAA
     GGAACGTCGT CGGTCGGGTG GGGTCGGGAC CGGAGGTTAG GGACCCCGGA TCTCTGGACT TGTTTGACGG GGACGGACCG GAGGGGTCGG CGGAATCCTT

                                                                   sau3AI
                                                                   dpnI
                                                                  xhoII
         mnlI          accI      mnlI            hphI             bglII  mnlI
15Ø1 TCGGGAGGGC TGTTCTTGTC TACCTGCCTC TTGCCCACTC TCACCCTGTG CTGGTCAGAT CTCCTCTCTC TCTGTTACCC AACTTCTCTT TGCTTTATAG
     AGCCCTCCCG ACAAGAACAG ATGGACGGAG AACGGGTGAG AGTGGGACAC GACCAGTCTA GAGGAGAGAG AGACAATGGG TTGAAGAGAA ACGAAATATC

                                                               sau96I                                 aluI
                                                               avaII                                 hindIII
                                   hphI        hinfI          nlaIV               hphI           taqI
16Ø1 AGACGACCAG TGTGTCTATA ACTTCACCCA TCTAGGAGTC CAGAGAGAGA ATGGGACCTT ATCCAAGTGT GGTGAGAGTG TGAACTCGAA GCTTTCTGGG
     TCTGCTGGTC ACACAGATAT TGAAGTGGGT AGATCCTCAG GTCTCTCTCT TACCCTGGAA TAGGTTCACA CCACTCTCAC ACTTGAGCTT CGAAAGACCC
```

```
                                ddeI
                               mstII
                            sau96I                                                  mnlI
       mnlI                 avaII mnlI                                       aluI ddeI
1701 GGGAGGGGGG GGTCATTGCA AGGACCTGAG GGGGAAGCAG AGCAGGGCAG TTTAGTCTGT TAGGTCAGCT CTGAGGCTGG TAAGTAGAAG CAGATTTCAA
     CCCTCCCCCC CCAGTAACGT TCCTGGACTC CCCCTTCGTC TCGTCCCGTC AAATCAGACA ATCCAGTCGA GACTCCGACC ATTCATCTTC GTCTAAAGTT

                                                                       ddeI
                                                                    aluI
                                                                   sacI
                                                                   hgiAI               nlaIII
                                                        sau3AI     bsp1286            styI
         hinfI  aluI                                    dpnI       banII        nlaIV ncoI fokI
1801 ACCAGAGTCC AGCTCCCAGC CCAGTCTCGC TTTTAACCAT TCTTTATGAT CCTTTGTGAG CTCAGCGTGG GAACCCATGG GATGGGAAGC CAGTGACTTC
     TGGTCTCAGG TCGAGGGTCG GGTCAGAGCG AAAATTGGTA AGAAATACTA GGAAACACTC GAGTCGCACC CTTGGGTACC CTACCCTTCG GTCACTGAAG

                         scrFI                                    haeIII
                      hphI                                       haeI   sau3AI          aluI       fokI ddeI
                  nlaIII bstNI                                   balI   dpnI  mnlI   ddeI      hinfI  mstII   fokI
1901 AGCCACATCT CCCATGTCAC CAGGGACAAC ACATAGGAAG GCAACCAAAC ACTGGTGGCC AGATCAGCCT CTCCTTAGCT TTGAATCATC CTAAGGGGAT
     TCGGTGTAGA GGGTACAGTG GTCCCTGTTG TGTATCCTTC CGTTGGTTTG TGACCACCGG TCTAGTCGGA GAGGAATCGA AACTTAGTAG GATTCCCCTA

                                                              sau96I
                                                              haeIII
                                                             sau96I
                                                             nlaIV
                                                             bsp1286
                                                             banII         nlaIV
                                    bsp1286             styI apaI      styI
2001 GTTTTAATGT TTAAGGGGAA GTTTTAACTG TGCCCACCAC CCCACCACCA AGGGCCCTTT CCAAGGCTCC AGCCGTTCAG TATCTATAAA GAATGAGAAG
     CAAAATTACA AATTCCCCTT CAAAATTGAC ACGGGTGGTG GGGTGGTGGT TCCCGGGAAA GGTTCCGAGG TCGGCAAGTC ATAGATATTT CTTACTCTTC

                                                                             sau96I
                               hgiAI          ddeI                      nlaIII
            nlaIV              bsp1286       mstII     fnu4HI          styI avaII    nlaIV
            banI             ddeI nlaIII     mnlI   nlaIV              ncoI nlaIV    banI
2101 TTAGGCTGGC ACCCAACAGG GACTGAGCAC ATGGTATCCT CAGGGAGCCG CAGGCATTTG CCATGGGGAC CACGGCACCA TTCTAGTGTC TTACAAAGCC
     AATCCGACCG TGGGTTGTCC CTGACTCGTG TACCATAGGA GTCCCTCGGC GTCCGTAAAC GGTACCCCTG GTGCCGTGGT AAGATCACAG AATGTTTCGG

                                                                        sau96I                                  scrFI
                                                                        avaII                                   nciI
              fnu4HI                                                   nlaIV                                    mspI
              bbvI                                        mboII nlaIII       nlaIII                             hpaII
2201 CTTCCTTCCG CAGCAGGAGC AGTGAAAATC TTTGCCCATT TGATAGTGCT GAAGAAACAT GGGACCTTCA TGCTTGCCTT TAACCTGACA GATGAGAACC
     GAAGGAAGGC GTCGTCCTCG TCACTTTTAG AAACGGGTAA ACTATCACGA CTTCTTTGTA CCCTGGAAGT ACGAACGGAA ATTGGACTGT CTACTCTTGG

                                             sau96I
                                             nlaIV                      sau96I
                                 ddeI        avaII                      nlaIV
                                mnlI      scrFI            haeIII       avaII
                            nlaIV    bstNI             sau96I    styI              fokI
2301 GGGGGCTGTC CTTCTATGGT AGGCTCCTCA GAACCCTGGA CCCCAGTTCT GCTGGGCCTT ACCCTTGGTC CCCCACCCAC ATCCCTATCT GGTCTCCCAT
     CCCCCGACAG GAAGATACCA TCCGAGGAGT CTTGGGACCT GGGGTCAAGA CGACCCGGAA TGGGAACCAG GGGGTGGGTG TAGGGATAGA CCAGAGGGTA

          sau96I                                       aluI                          fnu4HI
          nlaIV                                      fnu4HI                          bbvI      sspI
          avaII                                      bbvI                           aluI   xmnI
2401 TTGCTGGACC CTTACCCAGA ACACACCCCA CCCCCTTCTC CCTTGCAGCT AAAAAGCCAG ACTTGTCCCC AGAGCTGCGG AAAATATTCC AGCAGGCTGT
     AACGACCTGG GAATGGGTCT TGTGTGGGGT GGGGGAAGAG GGAACGTCGA TTTTTCGGTC TGAACAGGGG TCTCGACGCC TTTTATAAGG TCGTCCGACA
```

```
                                              taqI
                                             salI                          fnu4HI                    sau96I
                     fokI                    hincII                        bbvI    sfaNI   scrFI     nlaIV
                  nlaIII hinfI               accI                       mnlI   nlaIII bglI bstNI     avaII
2501 CAAAGATGTG GGCATGGATG AATCAGAAAT CGTATTTGTC GACTGGACAA AGGTAAGCGA CGAGGCTGCA TGATGCCACC CTGGCAGTGG TCCCGTTTTA
     GTTTCTACAC CCGTACCTAC TTAGTCTTTA GCATAAACAG CTGACCTGTT TCCATTCGCT GCTCCGACGT ACTACGGTGG GACCGTCACC AGGGCAAAAT

                                                                                                              ddeI
                                      scrFI      fnu4HI                                                     aluI
              mnlI ddeI  mnlI    aluI bstNI      bbvI                                                      pvuII
2601 GTGTCCCTAG AGGCTCAGAG AGGGTCAGCT CCTGGTTAAG GCAGCACAGC AAGGCGGGTA TCTTGCTGTT GAGAACGCCC GTGCCTACTG TGACAGCTGA
     CACAGGGATC TCCGAGTCTC TCCCAGTCGA GGACCAATTC CGTCGTGTCG TTCCGCCCAT AGAACGACAA CTCTTGCGGG CACGGATGAC ACTGTCGACT

                   ddeI
                sau3AI
                dpnI                     mnlI                                         sau3AI
              xhoII                    ddeI                                           dpnI           scrFI
     hinfI    bglII                   mstII                                      mnlI           bstNI    hphI
2701 GATTCACAGA GATCTTAGAT GGGCAGTGAG CCTGAGGTGG GTGGGGTTAA GTAGGAGATA CTAGAAAGAC TGGAGGATCA GGGGCGAGAC CAGGAAGGGT
     CTAAGTGTCT CTAGAATCTA CCCGTCACTC GGACTCCACC CACCCCAATT CATCCTCTAT GATCTTTCTG ACCTCCTAGT CCCCGCTCTG GTCCTTCCCA

                      nlaIV
                      banI         scrFI
                  aluI             bstNI
         mnlI    pvuII        hinfI
2801 GATTGAGGGA ACAGCTGGTG CCAGAGTCCC AGGTTTCTCC CTGCCTTTCA TTTTCACTTC CTATACAAGC CAGAGTTGTG ACTAAAGTCT CCCTGTGCCA
     CTAACTCCCT TGTCGACCAC GGTCTCAGGG TCCAAAGAGG GACGGAAAGT AAAAGTGAAG GATATGTTCG GTCTCAACAC TGATTTCAGA GGGACACGGT

                   styI                bsp1286
                 haeIII               nlaIV                                                 haeIII
                haeI     fokI         banI                      sfaNI                    mnlI    haeI     aluI
2901 GCAAGCCCAA TGGCCTTGGG GATGGGGGGG GGTGCCCACA CACCAAGTCA TCTTGATGCT CTGTCTAAAT TTGTGCAGAG GGAAAGGCCA TCAGCTTTGG
     CGTTCGGGTT ACCGGAACCC CTACCCCCCC CCACGGGTGT GTGGTTCAGT AGAACTACGA GACAGATTTA AACACGTCTC CCTTTCCGGT AGTCGAAACC

                      haeIII
                  ddeI                                                                               nlaIV
                aluI haeI      sau3AI                 ddeI                                           banI  ddeI
              pvuII mnlI       dpnI         aluI  taqI                                hinfI         bsp1286
3001 CTTTACCCCC AGCTGAGGCC ACAGGATCAA CCAATCAGCT TTCGACTTAG TGCGAGACAC ACTTGATATT TTGTGACTCC ATTCTCTGGG CACCTGAGCC
     GAAATGGGGG TCGACTCCGG TGTCCTAGTT GGTTAGTCGA AAGCTGAATC ACGCTCTGTG TGAACTATAA AACACTGAGG TAAGAGACCC GTGGACTCGG

                               nlaIV
                            scrFI   scrFI
                hphI        bstNI   bstNI                                aluI
3101 TTTCCTGACA TCACCAGCCC CCCAGGTTCC CAGGGGAAAG GTGTCTGCAT ACAGACACTG CCATTGAGCT GTTTTGCAAC CTGCCTTCCC TGCCTAGCCA
     AAAGGACTGT AGTGGTCGGG GGGTCCAAGG GTCCCCTTTC CACAGACGTA TGTCTGTGAC GGTAACTCGA CAAAACGTTG GACGGAAGGG ACGGATCGGT

                     hphI    rsaI                             mnlI          hinfI    mboII                  aluI
3201 GTGGGGTCAG TAAATTCACC CTGTACCTGT AGGCAGAATT GTTCAGAAAA CGCCTCCACT TGGGGGACTC TGTTCTTCCA TTCATTGGAA TAATAAGCTG
     CACCCCAGTC ATTTAAGTGG GACATGGACA TCCGTCTTAA CAAGTCTTTT GCGGAGGTGA ACCCCCTGAG ACAAGAAGGT AAGTAACCTT ATTATTCGAC

        ddeI                                                hincII
3301 GAACTTAGAG ACCAGTGGTT CTCAAGGCTG ACACTATCCT GCCCTTTCAG GTCAACCACC CCCAGAAATG TATTCATTGC TAGTGTGTGT GTGTGGGGGG
     CTTGAATCTC TGGTCACCAA GAGTTCCGAC TGTGATAGGA CGGGAAAGTC CAGTTGGTGG GGGTCTTTAC ATAAGTAACG ATCACACACA CACACCCCCC

                                                                                                           bsp1286
            sphI               nlaIII                                                 mboII                banII
        mnlI nlaIII nlaIII     sphI                          mnlI                     mnlI              nlaIII
3401 GGGGAGGGCA TGCCCATGCA CAAGCATGCC ATAAGACGTG TGTAGATGTC AGAGGACAAC TGTGGGGCGT GTCCTCTTCC GACATTCTCT CATGGGCTCT
     CCCCTCCCGT ACGGGTACGT GTTCGTACGG TATTCTGCAC ACATCTACAG TCTCCTGTTG ACACCCCGCA CAGGAGAAGG CTGTAAGAGA GTACCCGAGA
```

```
                                                                                                               fnu4HI
                                                      scrFI                                            hphI    bbvI scrFI
            ecoRI   aluI                              bstNI                                      sfaNI      pstI   bstNI
3501 GGGGATTGAA TTCAAGCTGT CCTTCAGGAT TACCTACTAT GCCAGGGTTT TTTGTTCTGT TTTGTTTGTT GTTTTGTTTT TGCATCTCAC CCTGCAGCCC
     CCCCTAACTT AAGTTCGACA GGAAGTCCTA ATGGATGATA CGGTCCCAAA AAACAAGACA AAACAAACAA CAAAACAAAA ACGTAGAGTG GGACGTCGGG

                                                                                   nlaIV
                                                                               scrFI
      bstEII                    mnlI         accI                              bstNI                           mboII
3601 AGGTTACCCT GCAACTCATA GCAATCCTCC TGCTTCAGTA TACTCCAACC CAGATGTGTG CCACCACCCC TGGCTCCGTC ACTTTGTCAC TTTGCTCTCT
     TCCAATGGGA CGTTGAGTAT CGTTAGGAGG ACGAAGTCAT ATGAGGTTGG GTCTACACAC GGTGGTGGGG ACCGAGGCAG TGAAACAGTG AAACGAGAGA

                                                                                                      mstII
                                                     pvuII                                         sau3AI
                                                    fnu4HI                                         dpnI          scrFI
                                                    bbvI                                          xhoII       haeIII
                                       fnu4HI     fnu4HI                                          nlaIV      haeI
       mnlI   mboII                    bbvI       bbvI aluI      aluI            ddeI             bamHI ddeI   bstNI
3701 TCCCTCGTAT CTTCTAGGAT AAGTGCAGTG AGCAGCAGAA GCAGCAGCTG GAGCTGGAGA AGGAGACTAA GAAGGAGACC AAGAAGGATC CTTAGGCCAG
     AGGGAGCATA GAAGATCCTA TTCACGTCAC TCGTCGTCTT CGTCGTCGAC CTCGACCTCT TCCTCTGATT CTTCCTCTGG TTCTTCCTAG GAATCCGGTC

                              scrFI
                              ncII                                          mnlI
            ddeI              mspI                                    hgiAI  taqI
      nlaIII   aluI           hpaII          nlaIII                   bsp1286   hinfI
3801 GCATGAACTC AGCTCTCTGA ACTCCGGGGA CTGTCCCCAT GCCCACTCTA CCCCCACTCC TGTGCACCTC GATTCTATTT TCCACAATAA TAAAGGTTTG
     CGTACTTGAG TCGAGAGACT TGAGGCCCCT GACAGGGGTA CGGGTGAGAT GGGGGTGAGG ACACGTGGAG CTAAGATAAA AGGTGTTATT ATTTCCAAAC

                                                                        scrFI     mnlI                        aluI
                                    mnlI mnlI            mnlI mnlI      bstNI mnlI        nlaIII           ddeI ddeI
3901 GTTAAACAAT CAGTATCACT TCTTTGTTCC CTCCCTCCCT TCTTTCCTTC CTCCCTCCCT TCCTGGAGAG GAGGACTGCC ATGAGTGGAG ACTTAGCTAA
     CAATTTGTTA GTCATAGTGA AGAAACAAGG GAGGGAGGGA AGAAAGGAAG GAGGGAGGGA AGGACCTCTC CTCCTGACGG TACTCACCTC TGAATCGATT

                      scrFI     ddeI         scrFI
                      bstNI hinfI   mnlI     bstNI             hinfI                                  aluI
4001 GTCCAATAAG CCCTACCCAG GGAATCTGAG AGGGCTTCCT GGAAGGGGCA TTTGAGTCAG GCTAGACGGA ACTGATAGAA CCTGATTATA GCTCTTGCTA
     CAGGTTATTC GGGATGGGTC CCTTAGACTC TCCCGAAGGA CCTTCCCCGT AAACTCAGTC CGATCTGCCT TGACTATCTT GGACTAATAT CGAGAACGAT

                                                                                          scrFI
                                                                                          bstNI
                                                                                      bsp1286
                        bsp1286       mboII                   fnu4HI                  banII
                      mnlI         hphI             mboII     bbvI hphI              nlaIV
4101 TAATCCCCAA ACCCTAGAGG GCACATTTGG TGAAGACAAA GTAGAAGACC CAAGCAGCGG TGACTGGGAG TGGGGAGCCC AGGTGTGCCG TTCCTGTCAC
     ATTAGGGGTT TGGGATCTCC CGTGTAAACC ACTTCTGTTT CATCTTCTGG GTTCGTCGCC ACTGACCCTC ACCCCTCGGG TCCACACGGC AAGGACAGTG

                                                                              taqI
        haeIII                                                             hgaI                                 styI
        sau96I                                         hphI               ahaII     nheI                      haeIII
      nlaIV                                          mnlI                 acyI  sfaNI  aluI                  haeI
4201 AGGGGCCTTC ACTTCAGATG TGCCATTGGA ACTCTGGAAG GGAGGTGAGT TGGGAGACAG GAAGGCGTCG ATGCTAGCTA TGGTCACGGT ATAAAAGGCC
     TCCCCGGAAG TGAAGTCTAC ACGGTAACCT TGAGACCTTC CCTCCACTCA ACCCTCTGTC CTTCCGCAGC TACGATCGAT ACCAGTGCCA TATTTTCCGG

                                                              mboII
                                                         scrFI               sau96I
                       ddeI             mnlI         ecoRV  bstNI          ddeI   avaII
4301 AAGGCAAGTG GCATAGACTT AGGGCTAGAC TAGAGGGGTT GGATATCCCC AGGAAGAGAA CTCAGTGGTC CGTGTGGGAC ACCGTGAGAT GACACTGCTT
     TTCCGTTCAC CGTATCTGAA TCCCGATCTG ATCTCCCCAA CCTATAGGGG TCCTTCTCTT GAGTCACCAG GCACACCCTG TGGCACTCTA CTGTGACGAA

            haeIII                                      mnlI
           haeI                                       ddeI
         mnlI   hincII                               mstII           ddeI                  ddeI
4401 CCCTAGAGGC CAGTTGACAT CTATTTCACA ATTGTGTTCA AAGTCCTGAG GCACAGTCAC TCAGTGGCAA ACACAGGGTC TGAGATAATC TTTACACTTG
     GGGATCTCCG GTCAACTGTA GATAAAGTGT TAACACAAGT TTCAGGACTC CGTGTCAGTG AGTCACCGTT TGTGTCCCAG ACTCTATTAG AAATGTGAAC
```

```
                                                                                                     hinpI
                                                                                                     hhaI
                        ddeI      hgiAI                                                             haeII
                       mstII      bsp1286                                    hphI                 fnu4HI      styI
                    hphI      fokI                                        hinfI                   bbvI    fokI
4501 ACTTGATTGG GCAGTCACCT AAGGGATGTG CTCGGTGTGT GTCTGTAAGG ACATTTCCAC AGAGAGTCAC CTGCCTTGAC TGTAGGCAGC GCCATCCCTT
     TGAACTAACC CGTCAGTGGA TTCCCTACAC GAGCCACACA CAGACATTCC TGTAAAGGTG TCTCTCAGTG GACGGAACTG ACATCCGTCG CGGTAGGGAA

                                               hgiAI
           sau96I                              bsp1286                       aluI
           nlaIV                             ddeI                           hindIII
           avaII                           aluI                           ddeI                                   nlaIII
4601 GGGCTTGGTC CCCCAATTAA CAAAAGGAGA AAACTAGCTG AGCACCCACA GTTCATCTGT CTGCTAAGCT TACGACTGTC CCAGTTGGAT ACAGTATGAC
     CCCGAACCAG GGGGTTAATT GTTTTCCTCT TTTGATCGAC TCGTGGGTGT CAAGTAGACA GACGATTCGA ATGCTGACAG GGTCAACCTA TGTCATACTG

                                                                                                          xhoI
                                                                                                          avaI
                                                                                                        aluI
                                                                                                      sacI taqI
                                 bsp1286                                                              hgiAI
                      mnlI         banII                                                              bsp1286
                   mnlI hphI  nlaIII              aluI   hincII                                       banII
4701 ATGGGCAGGT GTCCCTCCTC ACCCATGAGC CCCACTTCAA TAGCTTTCGT CAACTATTTT GTCACAGAAG GGGGAAAGCA GTTCTTATGG AGCTC
     TACCCGTCCA CAGGGAGGAG TGGGTACTCG GGGTGAAGTT ATCGAAAGCA GTTGATAAAA CAGTGTCTTC CCCCTTTCGT CAAGAATACC TCGAG
```

```
accI(GTMKAC):          1518 2538 3638
acyI(GRCGYC):          4264
ahaII(GRCGYC):         4264
aluI(AGCT):            2 549 608 669 743 830 1200 1250 1281 1690 1767 1811 1859 1977 2447 2473 2627
                       2695 2813 2993 3011 3037 3167 3296 3515 3746 3752 3811 3995 4090 4276 4636 4667
                       4742 4791
apaI(GGGCCC):          2052
avaI(CYCGRG):          63 4793
avaII(GGWCC):          261 365 918 1654 1722 2167 2262 2338 2367 2406 2589 4367 4607
avrII(CCTAGG):         860
balI(TGGCCA):          517 1956
bamHI(GGATCC):         1118 3786
banI(GGYRCC):          475 758 934 2108 2174 2817 2931 3089
banII(GRGCYC):         1 323 984 1858 2052 3494 4175 4727 4790
bbvI(GCWGC):           422 547 675 685 785 995 1279 1405 1408 2210 2445 2474 2565 2641 3594 3732 3741
                       3744 4154 4586
bglI(GCCNNNNNGGC):     2575
bglII(AGATCT):         404 1557 2710
bsp1286(GDGCHC):       1 150 323 984 1858 2030 2052 2125 2932 3088 3494 3862 4119 4175 4528 4640 4727
                       4790
bstEII(GGTNACC):       3602
bstNI(CCWGG):          142 361 431 681 750 908 1010 1122 1427 1441 1475 1920 2335 2580 2631 2790 2829
                       3122 3130 3542 3599 3669 3797 3962 4017 4038 4179 4349
ddeI(CTNAG):           25 208 284 329 340 383 765 832 1060 1104 1254 1493 1726 1771 1861 1974 1991
                       2123 2139 2327 2614 2697 2714 2732 3013 3046 3094 3304 3767 3791 3808 3992 3997
                       4026 4318 4361 4446 4460 4480 4519 4638 4664
dpnI(GATC):            139 405 1119 1558 1848 1962 2711 2776 3025 3787
ecoRI(GAATTC):         3508
ecoRV(GATATC):         4342
fnu4HI(GCNGC):         422 547 675 685 785 995 1279 1405 1408 1488 2147 2210 2445 2474 2565 2641 3594
                       3732 3741 3744 4154 4586
fokI(GGATGNNNNNNNNN):  1027 1091 1880 1987 1997 2380 2516 2920 4524 4593
haeI(WGGCCW):          517 1124 1429 1477 1956 2911 2985 3016 3794 4296 4407
haeII(RGCGCY):         782 4588
haeIII(GGCC):          311 518 754 1125 1430 1446 1478 1957 2053 2355 2912 2986 3017 3795 4204 4297
                       4408
hgaI(GACGC):           4265
hgiAI(GWGCWC):         1 1858 2125 3862 4528 4640 4790
```

```
hhaI(GCGC):              783 4589
hincII(GTYRAC):          1096 2538 3351 4413 4749
hindIII(AAGCTT):         1689 4666
hinfI(GANTC):            33 170 212 524 1083 1397 1498 1636 1805 1983 2520 2701 2824 3075 3266 3871 4022
                         4054 4564
hinpI(GCGC):             783 4589
hpaII(CCGG):             756 990 2299 3824
hphI(GGTGA):             36 215 653 856 896 1170 1541 1624 1671 1917 2798 3111 3216 3587 4129 4159 4244
                         4515 4567 4719
kpnI(GGTACC):            475
mboII(GAAGA):            336 370 746 772 1140 1243 2251 3274 3475 3698 3710 4132 4144 4353
mnlI(CCTC):              24 109 195 220 314 328 368 382 427 441 617 762 812 947 977 1005 1037 1062 1179
                         1432 1480 1505 1527 1563 1703 1728 1773 1968 2138 2326 2562 2610 2620 2734 2773
                         2805 2978 3015 3253 3404 3452 3473 3626 3703 3867 3930 3934 3950 3954 3968 3971
                         4030 4117 4242 4333 4406 4448 4714 4717
mspI(CCGG):              756 990 2299 3824
mstII(CCTNAGG):          382 1492 1725 1990 2138 2731 3790 4445 4518
nciI(CCSGG):             989 2299 3824
ncoI(CCATGG):            1113 1875 2161
nheI(GCTAGC):            498 4273
nlaIII(CATG):            40 79 480 521 779 791 915 1114 1146 1272 1876 1913 2130 2162 2258 2269 2513
                         2569 3409 3415 3425 3491 3802 3838 3980 4700 4724
nlaIV(GGNNCC):           15 260 261 364 475 562 758 934 985 1118 1444 1653 1870 2052 2065 2108 2144 2166
                         2174 2261 2322 2338 2367 2406 2589 2817 2931 3089 3125 3672 3786 4174 4202 4607

pstI(CTGCAG):            673 3592
pvuII(CAGCTG):           2694 2812 3010 3745
rsaI(GTAC):              476 507 3223
sacI(GAGCTC):            1 1858 4790
salI(GTCGAC):            2538
sau3AI(GATC):            139 405 1119 1558 1848 1962 2711 2776 3025 3787
sau96I(GGNCC):           261 311 365 753 918 1445 1654 1722 2052 2053 2167 2262 2338 2354 2367 2406 2589
                         4203 4367 4607
scrFI(CCNGG):            142 361 431 681 750 908 989 1010 1122 1427 1441 1475 1920 2299 2335 2580 2631
                         2790 2829 3122 3130 3542 3599 3669 3797 3824 3962 4017 4038 4179 4349
sfaNI(GATGC):            272 1026 1052 2572 2955 3582 4270
sphI(GCATGC):            3408 3424
sspI(AATATT):            1331 2483
styI(CCWWGG):            265 860 1113 1875 2048 2061 2161 2363 2914 4299 4597
taqI(TCGA):              1686 2539 3042 3869 4268 4794
tth111I(GACNNNGTC):      1079
xhoI(CTCGAG):            4793
xhoII(RGATCY):           138 404 1118 1557 2710 3786
xmnI(GAANNNNTTC):        2480
not found:
aatII(GACGTC), ahaIII(TTTAAA), asuII(TTCGAA), avaIII(ATGCAT), bclI(TGATCA), bsmI(GAATGCN), bssHII(GCGCGC),
bstXI(CCANNNNNNTGG), claI(ATCGAT), draI(TTTAAA), fspI(TGCGCA), hpaI(GTTAAC), mluI(ACGCGT), mstI(TGCGCA),
naeI(GCCGGC), narI(GGCGCC), ndeI(CATATG), notI(GCGGCCGC), nruI(TCGCGA), nsiI(ATGCAT), pvuI(CGATCG), sacII(CCGCGG),
scaI(AGTACT), sfiI(GGCCNNNNNGGCC), smaI(CCCGGG), snaBI(TACGTA), speI(ACTAGT), stuI(AGGCCT), thaI(CGCG), xbaI(TCTAGA),
xmaIII(CGGCCG)
```

REFERENCES

Abernethy, T.J., and Avery, O.T. (1941). The occurence during acute phase infections of a protein not normally present in the blood. I. Distribution of the reactive protein in patient's serra and the effect of calcium on the flocculation reaction with C polysaccharide of pneumococcus. J. Exp. Med. 73:173-182.

Adams, J.B., and Wacher, A. (1968). Specific changes in the glycoprotein components of seromucoid in pregnancy. Clin. Chim. Acta 21:155-157.

Agutter, P., McCaldin, B., and McArdle, M.J. (1979). Importance of mammalian nuclear-envelope nucleoside triphosphate in nucleo-cytoplasmic transport of ribonucleoproteins. Biochem. J. 182:811-819.

Agutter, P.S. (1983). An assessment of some methodological criticism of studies of RNA efflux from isolated nuclei. Biochem. J. 214:915-921.

Baumann, H., Firestone, G.L., Burgess, T.L., Gross, T.W., Yamamoto, K.R., and Held, W.A. (1983). Dexamethasone regulation of α_1-acid glycoprotein and other acute phase reactants in rat liver and hepatoma cells. J. Biol. Chem. 258:563-570.

Baumann, H., Held, W.A., and Berger, F.G. (1984). The acute phase response of mouse liver. J. Biol. Chem. 259:566-573.

Baumann, H., Jahreis, G.P., and Gaines, K.C. (1983). Synthesis and regulation of acute plasma proteins in primary culture of mouse hepatocytes. J. Cell Biol. 97:866-876.

Blatter, F.R., Williams, B.G., Blechl, A.E., Denniston-Thompson, K., Moore, D.D., Schumm, J.W., Scheldon, E.L., and Smithies, O. (1977). Charon phages: Safer derivatives of bacteriophage lambda for DNA cloning. Science 196:161-169.

Blin, N., and Stafford, D.W. (1976). A general method for isolating of high molecular weight DNA from eukaryotes. Nucl. Acids Res. 3:2303-2308.

Blobel, G., and Potter, V.R. (1966). Nuclei from rat livers: Isolation method that combines purity with high yield. Science 154:1662-1655.

Bornstein, D.L. (1982). Leukocytic pyrogen: A major mediator of the acute phase reaction. Ann. N.Y. Acad. Sci. 389:323-337.

Brawerman, G. (1981). The role of the poly(A) sequence in mammalian mRNA. CRC Crit. Rev. Biochem. 10:1-38.

Breathnach, R., and Chambon, P. (1981). Organization and expression of eukaryotic split genes coding for proteins. Ann. Rev. Biochem. 50:349-384.

Brown, D.D. (1981). Gene expression in eucaryotes. Science 211:667-674.

Chandra, T., Kurachi, K., Davie, E.W., and Woo, S.L.C. (1981). Induction of α_1 antitrypsin mRNA and cloning of its cDNA. Biochem. Biophys. Res. Commun. 103:751-758.

Chandrasekaran, E.V., Davila, M., Nixon, D., and Mendicino, J. (1984). Structure of the oligosaccharide chains of two forms of α_1-acid glycoprotein purified from liver metastases of lung, colon, and breast tumors. Cancer Res. 44:1557-1567.

Charney, P., Treisman, R., Mellon, P., Chao, M., Axel, R., and Maniatis, T. (1984). Differences in human α_1- and β-globin gene expression in mouse erythroleukemia cell: The role of intragenic sequences. Cell 38:251-263.

Chen, E.Y., and Seeberg, P.H. (1985). Supercoil sequencing: A fast and simple method for sequencing plasmid DNA. DNA 4:165-170.

Chirgwin, J., Pyzybyla, A.E., McDonald, R.J., and Rutter, W.J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 418:5294-5299.

Chiu, K.M., Mortenson, R.F., Osmand, A.P., and Gewurz, H. (1977). Interactions of α_1-acid glycoprotein with the immune system. Immunology 32:997-1005.

Chu, C.Y.-T., Lilian, T.-Y., and Pokala, H.P. (1982). Value of plasma α_1-acid glycoprotein assay the detection of human colorectal cancer: Comparison with carcinoembryonic antigen. J. Natl. Cancer Inst. 68:75-79.

Clawson, G.A., and Smuckler, E.A. (1980). Altered restriction of nuclear RNA during incubation in vitro. Biochem. Biophys. Res. Commun. 95:696-700.

Clawson, G.A., James, J., Woo, C.H., Friend, D.S., Moody, D.E., and Smuckler, E.A. (1980). Pertinence of nuclear envelope nucleoside triphosphatase activity to ribonucleic acid transport. Biochemistry 19:2748-2756.

Cohn, E.J., Strong, L.E., Hughes, W.L., Jr., Mulford, D.J., Ashworth, J.N., Melin, M., and Taylor, H.L. (1946). Preparation and properties of serum and plasma protein. IV. A system for the separation into fractions of the protein and lipoprotein components of biological tissues and fluids. J. Am. Chem. Soc. 68:459-472.

Darnell, J.E., Jr. (1982). Variety in the level of gene control in eukaryotic cells. Nature 297:365-371.

DeRoberts, E.M., Lienhard, S., and Parisot, R.F. (1982). Intracellular transport of microinjected 5S and small nuclear RNA. Nature 295:572-577.

Docherty, P.A., and Aronson, N.N., Jr. (1985). Effect of the threonine analog β-hydroxynorvaline on the glycosylation and secretion of α_1-acid glycoprotein by rat hepatocytes. J. Biol. Chem. 260:10847-10855.

Dynan, W.S., and Tjian, R. (1985). Control of eukaryotic messenger RNA synthesis by sequence-specific DNA binding proteins. Nature 316:774-778.

Feinberg, R.F., Sun, L.-H.K., Ordahl, C.P., and Frankel, F.R. (1983). Identifica-cation of glucocorticoid-induced genes in rat hepatoma cells by isolation of cloned cDNA sequences. Proc. Natl. Acad. Sci. USA 80:5042-5046.

Franzblau, C., Schmid, K., Farris, B., Beldekas, J., Garvin, P., Kagan, H.M., and Baum, B.J. (1976). The interaction of collagen with α_1-acid glycoprotein. Biochim. Biophys. Acta 427:302-314.

Friedman, J.M. (1983). Control of malaria virulence by α_1-acid glycoprotein (orosomucoid), an acute-phase (inflammatory) reactant. Proc. Natl. Acad. Sci. USA 80:5421-5424.

Gorski, J., Taft, D.O., Shyamala, G., Smith, D., and Notides, A. (1968). Hormone receptors: Studies on the interaction of estrogen with the uterus. Rec. Prog. Horm. Res. 24:45-80.

Graham, F., and Van der Eb, A. (1973). A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology 52:456-457.

Gross, V., Andus, T., Trans-Thi, T.-A., Schwarz, R.T., Decker, K., and Heinrich, P.C. (1983). 1-Deoxynojirimycin impair oligosaccharide processing of α_1-proteinase inhibitor and inhibit its secretion in primary cultures of rat hepatocytes. J. Biol. Chem. 258:12203-12209.

Guo, L.H., Yang, R.C., and Wu, R. (1983). An improved strategy for rapid direct sequencing of both strands of long DNA molecules cloned in a plasmid. Nucl. Acids Res. 11:5521-5540.

Hamada, H., Petrino, M.G., and Kakunga, T. (1982). A novel repeated element with Z-DNA-forming potential is widely found in evolutionary divers eukaryotic genomes. Proc. Natl. Acad. Sci USA 79:6465-6469.

Hayashida, H., and Miyata, T. (1983). Unusual evolutionary conservation and frequent DNA segment exchange in class I genes of the major histocompatibility complex. Proc. Natl. Acad. Sci. USA 80:1671-2675.

Hickman, S., Kulczycki, A., Jr., Lynch, R.G., and Kornfeld, S. (1977). Studies of the mechanism of tunicamycin inhibition of IgA and IgE secretion by plasma cells. J. Biol. Chem. 252:4402-4408.

Hohn, B., and Murray, K. (1977). Packaging recombinant DNA molecules into bacteriophage particles in vitro. Proc. Natl. Acad. Sci. USA 74:3259-3263.

Jacobson, A., and Farreau, M. (1983). Possible involvment of poly(A) in protein synthesis. Nucl. Acids Res. 11:6353-6368.

Jamieson, J.C., Friesen, A.D., Ashton, F.E., and Chou, B. (1972a). Studies on acute phase proteins of rat serum. I. Isolation and partial purification of an α_1-acid glycoprotein and α_2-macroglobulin. Can. J. Biochem. 50:856-870.

Jamieson, J.C., Ashton, F.E., Friesen, A.D., and Chou, B. (1972b). Studies on acute phase proteins of rat serum. II. Determination of the contents of α_1-acid glycoprotein, α_2-macroglobulin, and albumin in serum from rat suffering from induced inflammation. Can. J. Biochm. 50:871-880.

Jensen, E.V., Suzuki, T., Kawashima, T., Stumpf, W.E., Jungblut, P.W., and DeSombre, E.R. (1968). A two-step mechanism for the interaction of estradiol with rat uterus. Proc. Natl. Acad. Sci. USA 59:632-638.

Johnson, A.M., Schmid, K., and Alper, C.A. (1969). Inheritance of human α_1-acid glycoprotein (orosomucoid) variant. J. Clin. Invest. 48:2293-2299.

Kampschmidt, R.F., Upchurch, H.F., and Pulliam, L.A. (1982). Characterization of a leukocyte derived endogenous mediator responsible for increased plasma fibrinogen. Ann. N.Y. Acad. Sci. 389:338-353.

Kaufmann, H., and Schmid, K. (1962). The linear amino acid sequence of α_1-acid glycoprotein. Experientia 28:24-25.

Kimura, M. (1968). Evolutionary rate at the molecular level. Nature 217:624-626.

King, R.J.B. (1984). Enlightenment and confusion over steroid hormone receptors. Nature (London) 312:701-702.

Kushner, I. (1981). The acute phase reactants and the erythrocyte sedimentation rate. In: Textbook of Rheumatology. (W.N. Kelley, E.D. Harris, S. Ruddy, and C.B. Sledge, eds.), Saunders, Philadelphia, pp. 669-676.

Kushner, I. (1982). The phenomena of the acute phase response. Ann. N.Y. Acad. Sci. 389:39-48.

Laird, C.D., McConaughy, B.L., and McCarthy, B.J. (1969). Rate of fixation of nucleotide substitutions in evolution. Nature 224:149-154.

Ledford, B.E., and Davis, D.F. (1983). Kinetics of serum protein secretion by cultured hepatoma cells. J. Biol. Chem. 258:3304-3308.

Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L., and Streitz, J.A. (1980). Are snRNP's involved in splicing? Nature 283:220-224.

Lewin, R. (1981). Biggest challenge since the double helix. Science 212:28-32.

Littauer, U.Z., and Soreg, H. (1982). The regulatory function of poly(A) and adjacent 3' sequences in translated RNA. Proc. Natl. Acids Res. Mol. Biol. 27:53-83.

Lodish, H.F., and Kong, N. (1984). Glucose removal from N-linked oligosaccharides is required for efficient maturation of certain secretory glycoproteins from the rough endoplasmic reticulum to the Golgi complex. J. Cell Biol. 98:1720-1729.

Lodish, H.F., Kong, N., Snider, M., and Straus, G.J.A.M. (1983). Hepatoma secretory proteins migrate from rough endoplasmic reticulum to Golgi at characteristic rates. Nature 304:80-83.

Lonberg-Holm, K., and Philipson, L. (1980). Molecular aspects of virus receptors and cell surfaces. In: Cell Membranes and Viral Envelopes, Vol. 2. (H.A. Blough and J.M. Tiffany, eds.). Academic Press, London, pp. 789-848.

MacLeod, C.M., and Avery, O.T. (1941). The occurrence during acute infections of a protein not normally present in the blood. II. Isolation and properties of the reactive protein. J. Exp. Med. 73:183-190.

Mandol, M., and Higa, A. (1970). Calcium-dependent bacteriophage DNA infection. J. Mol. Biol. 53:159-162.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). Molecular Cloning, A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 109-112.

Margoliash, E., and Smith, E.L. (1965). Evolving Genes and Proteins (V. Bryson and H.J. Vogel, eds.), Academic Press, New York, pp. 167-183.

Martinez, H.M., Katzung, B., and Farrah, T. (1984). Sequence analysis program. Biomathematics Computation Laboratory, Department of Biochemistry and Biophysics, University of California, San Francisco.

Maxam, A.M., and Gilbert, W. (1977). A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA 74:560-564.

Mazier, D., Boudoin, R.L., Mellouk, S., Druihle, P., Texier, B., Trosper, J., Miltgen, F., Landau, I., Paul, C., Brandicourt, O., Guguen-Guillouzo, C., and Langlois, P. (1985). Complete development of hepatic stages of plasmodium falciparum in vitro. Science 227:440-442.

Messing, J., and Vieira, J. (1982). A new pair of M-13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene 19:269-276.

Messing, J., Gronenborn, B., Muller-Hill, B., and Hofschneider, P.H. (1977). Filamentous E. coliphage M-13 as a cloning vehicle: Insertion of a Hind III fragment of the lac regulatory region in M-13 replicative form in vitro. Proc. Natl. Acad. Sci. USA 74:3642-3646.

Miesfield, R., Okret, S., Wikstrom, A.-C., Weange, P., Gustafsson, J.-A., and Yamamoto, K.R. (1984). Characterization of a steroid hormone receptor gene and mRNA in wild-type and mutant cells. Nature 312:779-781.

Miyazaki, H., Fukamizu, A., Hirose, S., Hayashi, T., Hori, H., Ohkubo, H., Nakanishi, S., and Murakami, K. (1984). Structure of the human renin gene. Proc. Natl. Acad. Sci. USA 81:5999-6003

Moore, D.D., Marks, A.R., Buckley, A.R., Kapler, G., Payvar, F., and Goodman, H. (1985). The first intron of the human growth gene contains a binding site for glucocorticoid receptor. Proc. Natl. Acad. Sci. USA 81:5999-6003.

Morell, A.G., Gregoriadis, G., Scheinberg, I.H., Hickman, J., and Ashwell, G. (1971). The role of sialic acid in determining the survival of glycoproteins in circulation. J. Biol. Chem. 246:1461-1467.

Morrow, J.F., Stearman, R.S., Peltzman, C.G., and Potter, D.A. (1981). Induction of hepatic synthesis of serum amyloid A protein and actin. Proc. Natl. Acad. Sci. USA 78:4718-4722.

Mount, S.M. (1982). A catalogue of splice junction sequences. Nucl. Acids Res. 10:459-472.

Mount, S.M. (1983). Sequences that signal where to splice. Science 304:309-310.

Nimberg, R.B., Motoyama, T., and Schmid, K. (1971). The amino acid substitutions found in the genetic variants of α_1-acid glycoprotein. J. Biol. Chem. 246:5817-5821.

Nordheim, A., and Rich, A. (1983). The sequence (dC-dA)n • (dG-dT)n forms left-handed Z-DNA in negatively supercoiled plasmids. Proc. Natl. Acad. Sci. USA 80-1821-1825.

Nordheim, A., and Rich, A. (1983). Negatively supercoiled simian virus 40 DNA contains Z-DNA segments within transcriptional enhancer sequences. Nature (London) 303:674-679.

O'Malley, B.W., and Schrader, W.T. (1981). Laboratory Methods Manual for Hormone Action and Molecular Endocrinology, 6th Edition, Huston Biological Association, Inc.

Ohshima, Y.M., Itoh, N., and Miyata, T. (1981). Novel models for RNA splicing that involve a small nuclear RNA. Proc. Natl. Acad. Sci. USA 78:4471-4474.

Payvar, F., DeFranco, D., Firestone, G.L., Edgar, B., Wrange, O., Okret, S., Gustafsson, J.-A., and Yamamoto, K.R. (1983). Sequence-specific binding of glucocorticoid receptor to MTV DNA at sites within and upstream of the transcribed region. Cell 35:381-392.

Pepys, M.B., and Baltz, M.L. (1983). Acute phase proteins with special reference to C-reactive protein and related proteins (pentaxins) and serum amyloid A protein. Adv. Immunol. 34:141-212.

Pepys, M.B., Breathnach, S.M., Black, M.M., Tennent, G., de Beer, F.C., Lanham, J., Zaline, A.M., Tribe, C.R., and Evans, D.J. (1981). Diagnosis and characterization of amyloid deposit by immunohistochemical staining. In: Proceedings of European Amyloid Research Symposium (P.A. Bacon, P. Madison, and J.T. Whicher, eds.). Reedbooks, Chertsey, Surrey, pp. 189-191.

Putnam, F.W. (1975). The Plasma Proteins, 1:2-51. Academic Press, New York.

Reinke, R., and Feigelson, P. (1985). Rat α_1-acid glycoprotein, gene sequence and regulation by glucocorticoids in transfected L-cells. J. Biol. Chem. 260:4397-4403.

Ricca, G.A., and Taylor, J.M. (1981). Nucleotide sequence of rat α_1-acid glycoprotein messenger RNA. J. Biol. Chem. 256:11199-11202.

Ricca, G.A., Hamilton, R.W., McLean, J.W., Conn, A., Kalinyak, J.E., and Taylor, J.M. (1981). Rat α_1-acid glycoprotein messenger RNA. J. Biol. Chem. 256:10362-10368.

Ricca, G.A., McLean, J.W., and Taylor, J.M. (1982). Kinetics of induction of α_1-acid glycoprotein. Ann. N.Y. Acad. Sci. 389:88-105.

Rogers, J., and Wall, R. (1980). A mechanism for RNA splicing. Proc. Natl. Acad. Sci. USA 77:1877-1879.

Salditt-Georgiett, M., and Darnell, J.E., Jr. (1982). Further evidence that the majority of primary nuclear RNA transcript in mammalian cells do not contribute to mRNA. Mol. Cell. Biol. 2:701-707.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.

Sargent, T.D., Wu, J., Sela-Trepat, J.M., Wallace, R.B., Reyes, A.A., and Bonner, J. (1979). The rat serum albumin gene: Analysis of cloned sequences. Proc. Natl. Acad. Sci. USA 76:3256-3260.

Scheidereit, C., and Baeto, M. (1984). Contact between hormone receptor and DNA double helix within a glucocorticoid regulatory element of mouse mammary tumor virus. Proc. Natl. Acad. Sci. USA 81:3029-3033.

Scheidereit, C., Geisse, S., Westphal, H.M., and Beato, M. (1983). The glucocorticoid receptor binds to defined nucleotide sequences near the promotor of mouse mammary tumor virus. Nature (London) 304:749-752.

Schmid, K. (1975). α_1-Acid glycoprotein. In: The Plasma Proteins, Vol. 1 (F. Putnam, ed.), Academic Press, New York, pp. 184-228.

Schmid, K., Binnette, J.P., Kamiyama, S., Pfister, V., and Takahashi, S. (1962). Studies on the structure of α_1-acid glycoprotein. III. Polymorphism of α_1-acid glycoprotein and the partial resolutions and characterization of its variants. Biochemistry 1:959-966.

Schmid, K., Burgi, W., Collins, H., and Nanno, S. (1974). The disulfide bonds of α_1-acid glycoprotein. Biochemistry 13:2694-2697.

Schmid, K., Kaufmann, H, Isemura, S., Bauer, F., Emura, J., Motoyama, T., Ishiguro, M., and Nanno, S. (1973a). Structure of α_1-acid glycoprotein. The complete amino acid sequence, multiple amino acid subsitutions and homology with the immunoglobulins. Biochemistry 12:2711-2724.

Schmid, R.B., Ishiguro, M., Emura, J., Isemura, S., Kaufmann, H., and Motoyama, T. (1971). The amino acid sequences of two large glycopeptides derived from the carbohydrate-carrying region of α_1-acid glycoprotein. Biochem. Biophys. Res. Commun. 42:28-286.

Schreier, P.H., and Cortese, R. (1979). A fast and simple method for sequencing DNA cloned in the single-stranded bacteriophage M-13. J. Mol. Biol. 129:169-172.

Schumm, D.E., and Webb, T.E. (1978). Effect of adenosine 3':5'-monophosphate and guanosine 3':5'-monophosphate on RNA release from isolated nuclei. J. Biol. Chem. 253:8513-8517.

Shearer, R.W., and Smuckler, E.A. (1972). Altered regulation of the transport of RNA from nucleus to cytoplasm in rat hepatoma cells. Cancer Res. 32:339-342.

Simkin, J.L., and Jamieson, J.C. (1968). Studies on the nature of microsome-bound substances involved in the biosynthesis of acidic glycoprotein of guinea-pig serum. Biochem. J. 106:23-28.

Sipe, J.D., and Rosenstreich, D.L. (1981). Serum factors associated with inflammation. In: Cellular Functions in Immunity and Inflammation (J.J. Oppenheim, D.L. Rosenstreich, and M. Potter, eds.). Elsevier, Amsterdam, pp. 411-429.

Smuckler, E.A., and Koplitz, M. (1974). Thioacetamide-induced alteration in nuclear RNA transport. Cancer Res. 34:827-838.

Smuckler, E.A., et al. (1982). Pathology 101 Manual. University of California, San Francisco, pp. 178-179.

Snyder, S., and Coodley, E.L. (1976). Inhibition of platelet aggregation by α_1-acid glycoprotein. Arch. Int. Med. 136:778-781.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by α_1-acid glycoprotein. Arch. Intern. Med. 136:778-781.

Stacey, D.W. (1981). Microinjection of mRNA and other macromolecules into living cells. Methods in Enzymol. 79:76-88.

Sternberg, N., Tiemeier, D., and Enquist, L. (1977). In vitro packaging of a lambda Dam vector containing Eco RI DNA fragment of Escherichia coli and phage P1. Gene 1:255-280.

Subramani, S., and Southern, P.J. (1983). Analysis of gene expression using SV40 vectors (review). Anal. Biochem. 135:1-15.

Tokita, K., Burke, J., Yoshizaki, H., Fischer, S., and Schmid, K. (1966). The constancy of the α_1-acid glycoprotein variant of normal adult under conditions of severe stress. J. Clin. Invest. 45:1624-1630.

Tokes, Z.A., Gendler, S.J., Imin, A., Bullano, T.G., and Ross, K.L. (1983). Deciphering cancer-associated glycoproteins that may have diagnostic relevance. In: Cellular Oncology, Cancer Research Monographs, Vol. 1 (P.J. Moloy and G.L. Nicolson, eds.). Praeger Press, New York, pp. 28-62.

Vannice, J.L., Ringold, G.M., McLean, J.W., and Taylor, J.M. (1983). Induction of the acute-phase reactant, α_1-acid glycoprotein, by glucocorticoids in rat hepatoma cells. DNA 2:205-212.

Vannice, J.L., Taylor, J.M., and Ringold, G.M. (1984). Glucocorticoid-mediated induction of α_1-acid glycoprotein: Evidence for hormone-regulated RNA processing. Proc. Natl. Acad. Sci. USA 81:4241-4245.

Walden, W.E., Godefroy-Colburns, T., and Thach, R.E. (1981). The role of mRNA competition in regulating translation I, II, III. J. Biol Chem. 256:11739-11746; 11747-11754; 11755-11761.

Weimer, H.E., Mehl, J.W., and Winzler, R.J. (1950). Studies on the mucoproteins of human plasma. V. Isolation and characterization of a homogeneous mucoprotein. J. Biol. Chem. 185:561-568.

Wickens, M.P., Woo, S., O'Malley, B.W., and Gurdon, J.B. (1980). Expression of a chicken chromosomal ovalbumin gene injection into frog oocyte nuclei. Nature 285:628-640.

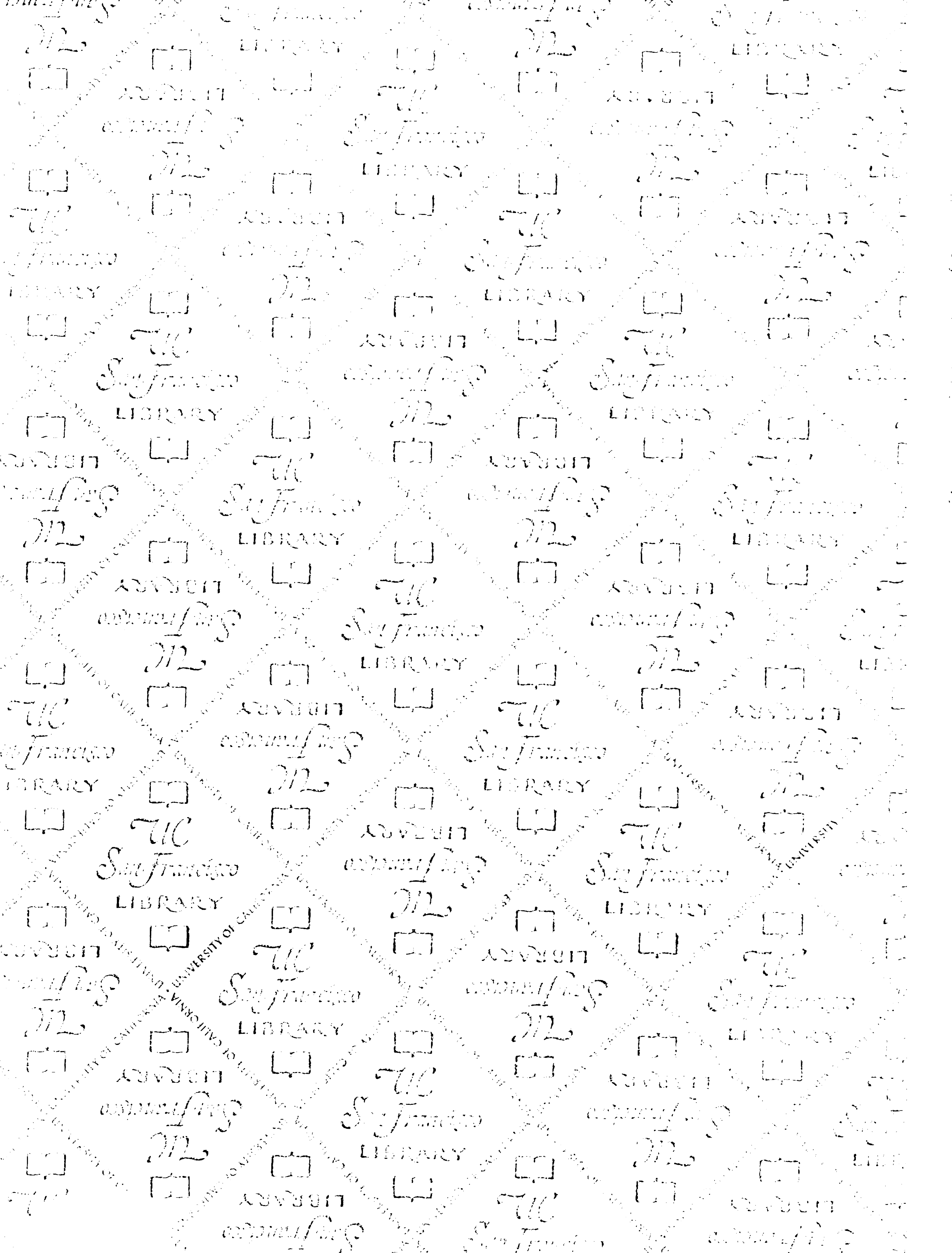
Winzler, R.J. (1955). Determination of serum glycoprotein. Meth. Biochem. Anal. 2:279-311.

Winzler, R.J. (1965). Metabolism of glycoproteins. Clin. Chem. 11:339-345.

Wright, S., Rosenthal, A., Flavel, R., and Grosveld, F. (1984). DNA sequences required for regulated expression of β-globin genes in murine erythroleukemia cells. Cell 38:265-273.

Yamashina, I. (1956). On the carbonhydrate components of the α_1-acid glycoprotein of human plasma. Acta Chem. Scand. 10:1666-1668.

Yoshizaki, H., Hunziker, K., and Schmid, K. (1969). The constancy of the types of α_1-acid glycoprotein variants in patients with uterectomy and irradiation. Clin. Chim. Acta 23:147-151.



FOR REFERENCE

NOT TO BE TAKEN FROM THE ROOM