Characterization of human haptoglobin cDNAs coding for $\alpha^{2FS}\beta$ and $\alpha^{1S}\beta$ variants

Ariane van der Straten, Albert Herzog, Teresa Cabezón* and Alex Bollen+

Laboratory of Genetics, University of Brussels, rue des Chevaux 67, B-1640 Rhode-St-Genèse and **Smith-Kline/RIT, rue de l'Institut 89, B-1300 Rixensart, Belgium*

Received 9 January 1984

A human liver library, derived from a heterozygous (Hp2-1) donor, has been used to isolate cDNA clones coding for the haptoglobin (Hp) $\alpha^{1S}\beta$ and $\alpha^{2FS}\beta$ variants. DNA sequencing has shown that the two variants are identical except for the α^{F} duplicated segment in Hp $\alpha^{2FS}\beta$. Four nucleotide changes have been found between the phenotypically different F and S regions of the Hp α^{2} gene, resulting in an Asp,Lys/Asn,Glu substitution.

Haptoglobin *Cloning* *DNA sequence* *Gene duplication* *Allele* *Phenotype*

1. INTRODUCTION

Haptoglobin (Hp) is a serum glycoprotein synthesized in the liver. The protein has a tetrameric structure $(\alpha\beta)_2$ with disulfide bonds linking the polypeptide chains [1]. Hp binds free hemoglobin (Hb) in plasma of humans and other mammals, in the ratio of one Hb$\alpha\beta$ subunit per Hp$\alpha\beta$ subunit, and forms an irreversible complex which is taken up by liver cells and digested [2,3]. There is good evidence to show that the Hb binding site on haptoglobin is located on the β chain [1,4]. The β polypeptide is 245 residues long; it has a well conserved sequence as very few β variants have been described. This is consistent with its functional role, whereas α chain which shows greater variability, does not directly participate in Hb binding [1]. Inherited variations in the smaller subunit, the α chain, are responsible for Hp polymorphism in human populations, the α chain having two major allelic forms, α^1 with 83 residues and α^2 with 142 residues. These alleles determine the 3 major phenotypes Hp1-1, Hp2-2 and Hp2-1 [1].

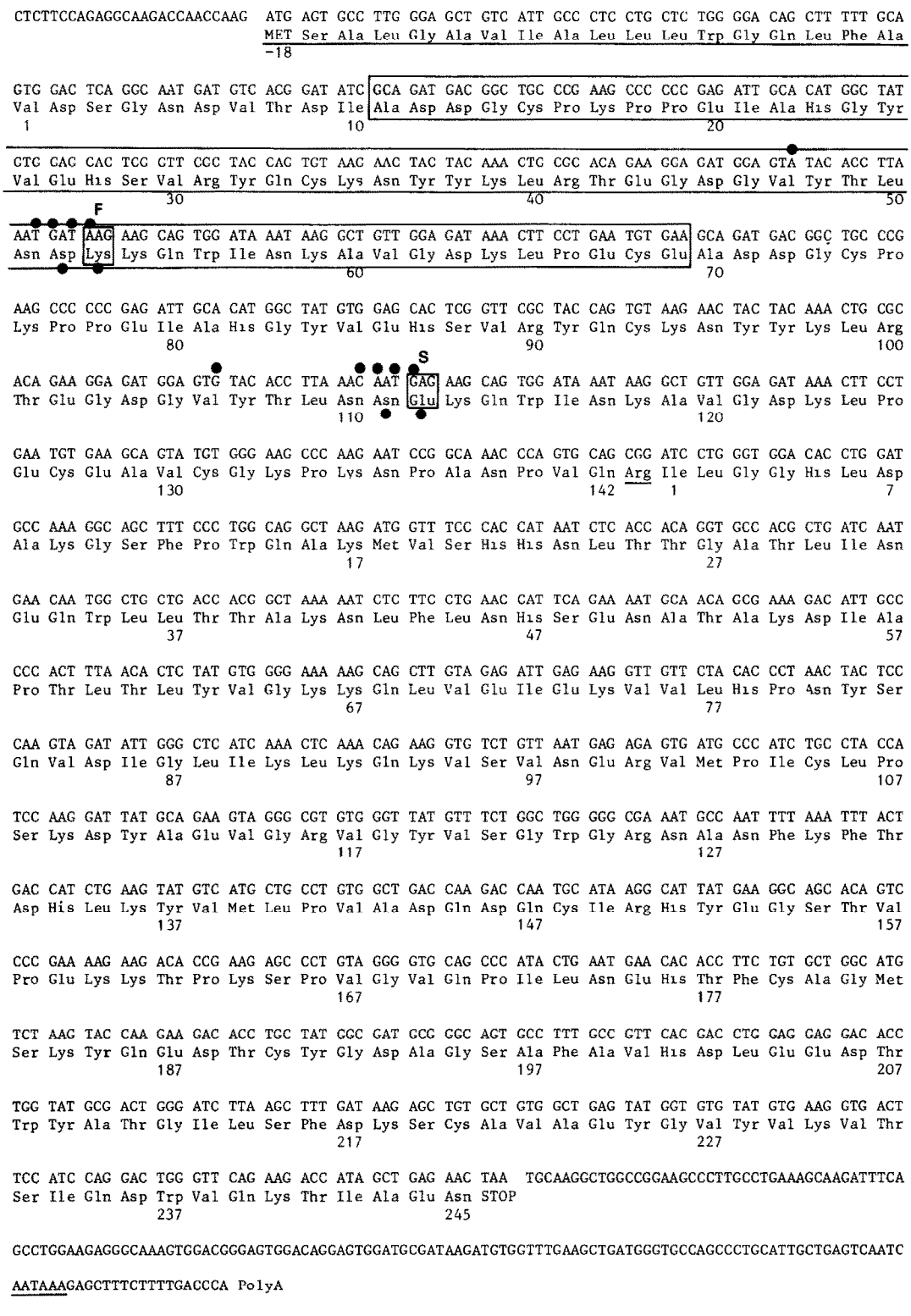
+ To whom correspondence should be addressed

There are also two electrophoretic types of Hpα^1 chains, α1F (fast) and α1S (slow), differing by a Lys/Glu amino acid substitution at position 53 [1]. Two alleles control their structure, Hpα^{1S} and Hpα^{1F}. The third allele Hpα^2 is the product of a partial gene duplication possibly resulting from an unequal crossover event in a heterozygous genotype Hpα^{1F}/Hpα^{1S} [1].

As we previously showed, Hp mRNA codes for both α and β polypeptides in tandem [5]. The two chains are linked on the $\alpha\beta$ precursor by a single Arg residue, which is released during the proteolytic maturation generating the α and β subunits [6–8]. This cleavage mechanism gives further support to the hypothesis of a common ancestor for Hp and the serine protease family [9].

We report here the characterization of recombinant clones containing cDNA genes coding for Hpα^{2FS} (Lys 53/Glu 112) and Hpα^{1S} (Glu 53) variants. Their DNA sequences are identical except for the Ala 11 ⟶ Glu 69 duplicated portion of α^2. In addition to the Lys/Glu amino acid substitution at position 53, another amino acid change and two nucleotide differences are found between the F and S regions of the Hpα^2 gene. β sequences are identical for Hpα^{2FS} and Hpα^{1S} variants at the DNA level.

Published by Elsevier Science Publishers B.V.
00145793/84/$3 00 © 1984 Federation of European Biochemical Societies


CTCTTCCAGAGGCAAGACCAACCAAG ATG AGT GCC TTG GGA GCT GTC ATT GCC CTC CTG CTC TGG GGA CAG CTT TTT GCA
MET Ser Ala Leu Gly Ala Val Ile Ala Leu Leu Leu Trp Gly Gln Leu Phe Ala
-18
GTG GAC TCA GGC AAT GAT GTC ACG GAT ATC GCA GAT GAC GGC TGC CCG AAG CCC CCC GAG ATT GCA CAT GGC TAT
Val Asp Ser Gly Asn Asp Val Thr Asp Ile Ala Asp Asp Gly Cys Pro Lys Pro Pro Glu Ile Ala His Gly Tyr
1 10 20
CTG GAG CAC TCG GTT CGC TAC CAG TGT AAG AAC TAC TAC AAA CTG CGC ACA GAA GCA GAT GGA GTA TAC ACC TTA
Val Glu His Ser Val Arg Tyr Gln Cys Lys Asn Tyr Tyr Lys Leu Arg Thr Glu Gly Asp Gly Val Tyr Thr Leu
30 40 50
F
AAT GAT AAG AAG CAG TGG ATA AAT AAG GCT GTT GGA GAT AAA CTT CCT GAA TGT GAA GCA GAT GAC GGC TGC CCG
Asn Asp Lys Lys Gln Trp Ile Asn Lys Ala Val Gly Asp Lys Leu Pro Glu Cys Glu Ala Asp Asp Gly Cys Pro
60 70
AAG CCC CCC GAG ATT GCA CAT GGC TAT GTG GAG CAC TCG GTT CGC TAC CAG TGT AAG AAC TAC TAC AAA CTG CGC
Lys Pro Pro Glu Ile Ala His Gly Tyr Val Glu His Ser Val Arg Tyr Gln Cys Lys Asn Tyr Tyr Lys Leu Arg
80 90 100
S
ACA GAA GGA GAT GGA GTG TAC ACC TTA AAC AAT GAG AAG CAG TGG ATA AAT AAG GCT GTT GGA GAT AAA CTT CCT
Thr Glu Gly Asp Gly Val Tyr Thr Leu Asn Asn Glu Lys Gln Trp Ile Asn Lys Ala Val Gly Asp Lys Leu Pro
110 120
GAA TGT GAA GCA GTA TGT GGG AAG CCC AAG AAT CCG GCA AAC CCA GTG CAG CGG ATC CTG GGT GGA CAC CTG GAT
Glu Cys Glu Ala Val Cys Gly Lys Pro Lys Asn Pro Ala Asn Pro Val Gln Arg Ile Leu Gly Gly His Leu Asp
130 142 1 7
GCC AAA GGC AGC TTT CCC TGG CAG GCT AAG ATG GTT TCC CAC CAT AAT CTC ACC ACA GGT GCC ACG CTG ATC AAT
Ala Lys Gly Ser Phe Pro Trp Gln Ala Lys Met Val Ser His His Asn Leu Thr Thr Gly Ala Thr Leu Ile Asn
17 27
GAA CAA TGG CTG CTG ACC ACC GCT AAA AAT CTC TTC CTG AAC CAT TCA GAA AAT GCA ACA GCG AAA GAC ATT GCC
Glu Gln Trp Leu Leu Thr Thr Ala Lys Asn Leu Phe Leu Asn His Ser Glu Asn Ala Thr Ala Lys Asp Ile Ala
37 47 57
CCC ACT TTA ACA CTC TAT GTG GGG AAA AAG CAG CTT GTA GAG ATT GAG AAG GTT GTT CTA CAC CCT AAC TAC TCC
Pro Thr Leu Thr Leu Tyr Val Gly Lys Lys Gln Leu Val Glu Ile Glu Lys Val Val Leu His Pro Asn Tyr Ser
67 77
CAA GTA GAT ATT GGG CTC ATC AAA CTC AAA CAG AAG GTG TCT GTT AAT GAG AGA GTG ATG CCC ATC TGC CTA CCA
Gln Val Asp Ile Gly Leu Ile Lys Leu Lys Gln Lys Val Ser Val Asn Glu Arg Val Met Pro Ile Cys Leu Pro
87 97 107
TCC AAG GAT TAT GCA GAA GTA GGG CGT GTG GGT TAT GTT TCT GGC TGG GGG CGA AAT GCC AAT TTT AAA TTT ACT
Ser Lys Asp Tyr Ala Glu Val Gly Arg Val Gly Tyr Val Ser Gly Trp Gly Arg Asn Ala Asn Phe Lys Phe Thr
117 127
GAC CAT CTG AAG TAT GTC ATG CTG CCT GTG GCT GAC CAA GAC CAA TGC ATA AGG CAT TAT GAA GGC AGC ACA GTC
Asp His Leu Lys Tyr Val Met Leu Pro Val Ala Asp Gln Asp Gln Cys Ile Arg His Tyr Glu Gly Ser Thr Val
137 147 157
CCC GAA AAG AAG ACA CCG AAG AGC CCT GTA GGG GTG CAG CCC ATA CTG AAT GAA CAC ACC TTC TGT GCT GGC ATG
Pro Glu Lys Lys Thr Pro Lys Ser Pro Val Gly Val Gln Pro Ile Leu Asn Glu His Thr Phe Cys Ala Gly Met
167 177
TCT AAG TAC CAA GAA GAC ACC TGC TAT GGC GAT GCC GGC AGT GCC TTT GCC GTT CAC GAC CTG GAG GAG GAC ACC
Ser Lys Tyr Gln Glu Asp Thr Cys Tyr Gly Asp Ala Gly Ser Ala Phe Ala Val His Asp Leu Glu Glu Asp Thr
187 197 207
TGG TAT GCG ACT GGG ATC TTA AGC TTT GAT AAG AGC TGT GCT GTG GCT GAG TAT GGT GTG TAT GTG AAG GTG ACT
Trp Tyr Ala Thr Gly Ile Leu Ser Phe Asp Lys Ser Cys Ala Val Ala Glu Tyr Gly Val Tyr Val Lys Val Thr
217 227
TCC ATC CAG GAC TGG GTT CAG AAG ACC ATA GCT GAG AAC TAA TGCAAGGCTGGCCGGAAGCCCTTGCCTGAAAGCAAGATTTCA
Ser Ile Gln Asp Trp Val Gln Lys Thr Ile Ala Glu Asn STOP
237 245
GCCTGGAAGAGGGCAAAGTGGACGGGAGTGGACAGGAGTGGATGCGATAAGATGTGGTTTGAAGCTGATGGGTGCCAGCCCTGCATTGCTGAGTCAATC
AATAAAGAGCTTTCTTTTGACCCA PolyA

2. MATERIALS AND METHODS

All methods for RNA preparation, cDNA synthesis and cloning, and hybridization selection of Hp mRNA have been described in [5].

2.1. *In situ hybridization of bacterial colonies*

Escherichia coli transformants were screened for the presence of Hp-encoding sequences by hybridization of colonies transferred onto nitrocellulose filters (Sartorius type SM50) with ^{32}P-labelled nick-translated DNA as in [10].

2.2. *DNA sequencing*

DNA sequence determination was performed as in [11]. Labelling of DNA fragments was done either with T4 polynucleotide kinase and [γ-^{32}P]dATP or with DNA polymerase I and [α-^{32}P]dNTP [12].

3. RESULTS AND DISCUSSION

Starting from human liver mRNA, a human liver cDNA clone bank was constructed using pBR322 as cloning vector and bacterial clones containing Hp cDNA sequences were identified as in [5]. One of them, pULB1148, has an insert of ~1.4 kb and covers the complete Hp$\alpha^{2FS}\beta$ gene. Its DNA sequence shows that α^2 and β sequences are contiguous [5]. The open reading frame starts with an initiation triplet (ATG) 18 codons upstream from the triplet encoding the first amino acid of the mature α^2 protein (Val). The stretch of 18 amino acids preceding the Val 1 is hydrophobic, and probably constitutes a signal peptide involved in haptoglobin secretion [13]. The first stop signal (TAA) is found at the 3′-end of the β coding sequence. α and β sequences are in the same reading frame, separated by a CGG triplet coding for an Arg residue. The polyadenylation signal AATAAA is found in the cDNA sequence at position −23 from the poly(A) track.

Further sequencing of the Hp gene allowed us to complete and correct the published sequence of pULB1148, where a few non-identified bases (methylated C) remained [5].

In order to verify our suggestion [5] that the ATG codon of pULB1148 is not artefactual, as it is found contiguous to the oligo(dG) tail of the clone, the bank has been screened for longer 5′-terminal Hp clones, using a probe corresponding to the leader sequence plus the first 19 amino acids of α chain. In one of these clones, the longest 5′-terminal cDNA insert begins 26 nucleotides upstream from the −18 ATG which remains the first one encountered on the sequence (fig.1) and thus appears to be the effective initiation codon. Hp α^1 sequences should be present in the liver library if the donor was a heterozygous Hp2-1. Restriction analysis is a simple way to verify this possibility because partial gene duplication generates two *Ava*I sites in Hpα^2 whereas only one should be found for Hpα^1 sequences (fig.2).

Unique *Ava*I Hp clones were selected and clone pULB5741, which carries a ~1.2 kb insert covering the complete Hp$\alpha^{1S}\beta$ sequence, has been identified. Hp$\alpha^{2FS}\beta$ and Hp$\alpha^{1S}\beta$ cDNAs have identical sequences except for the α^F duplicated segment (Ala 11 ⟶ Glu 69) of α^2 (fig.1). Differences between the α^F and α^S homologous regions were found at the nucleotide level as well as at the amino acid level (table 1). In addition to the Lys (AAG)/Glu (GAG) substitution occurring in positions 53 and 112, DNA sequencing has shown that residues in the preceding position (52 and 111) also vary according to F or S phenotype: Asp (GAT) and Asn (AAT) were found respectively before the Lys and Glu amino acids. Both amino acids changes result from a G⟶A transition at the nucleotide level. Moreover, two other nucleotide differences have also been found between codons for homologous residues Val (47 and 106) and Asn (51 and 110). The 4 base pair transition events located within a 7-codon long region seem

←

Fig.1. Complete nucleotide sequence of Hp$\alpha^{2FS}\beta$ and Hp$\alpha^{1S}\beta$ cDNAs. The amino acid sequence is deduced from the DNA sequence. The peptide signal and Arg residue released during proteolytic maturation of the precursor polypeptide are underlined. The polyadenylation site is also underlined. The duplicated segment in the Hpα^2 sequence is boxed and must be omitted when reading the Hpα^1 sequence. Nucleotide and amino acid differences between the F and S regions of Hpα^2 gene are indicated by (●). Characteristic F and S amino acids at positions 53 and 112 are boxed and indicated by F and S capitals.

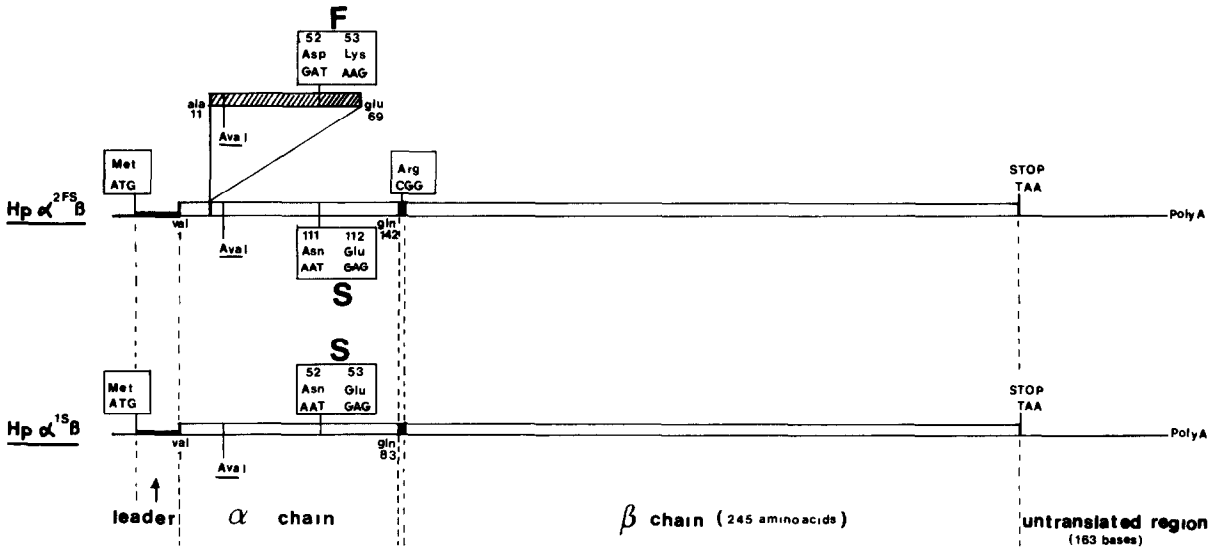


Fig.2. Schematic representation of the cDNAs coding for Hp$\alpha^{2FS}\beta$ and Hp$\alpha^{1S}\beta$ variants. Codons and amino acids of particular interest have been pointed out. Duplicated α^F portion of Hpα^2 has been represented in parallel to its homologous α^S region.

therefore to characterize two different α haptoglobin genotypes which lead, at the amino acid level, to a paired Asp,Lys/Asn,Glu substitution constituting the electrophoretically determined F and S phenotypes.

Two other groups [14,15] have also sequenced cDNAs coding for Hp$\alpha^{2FS}\beta$ variants. Both published sequences are almost identical to ours: the only amino acid difference in α^2 appears at residue 52 where authors in [15] determined an Asn residue instead of the Asp found by authors in [14] and ourselves. In the β sequence, authors in [14] found an Asn in position 206 instead of an Asp.

Table 1

Relevant base and amino acid substitutions between F and S regions of Hp$\alpha^2\beta$ and Hp$\alpha^1\beta$ clones

F region			S region			
Amino acid	Position in α^2	Codon	Amino acid	Position in α^1	Position in α^2	Codon
Val	47	GTA	Val	47	106	GTG
Asn	51	AAT	Asn	51	110	AAC
Asp	52	GAT	Asn	52	111	AAT
Lys	53	AAG	Glu	53	112	GAG

ACKNOWLEDGEMENTS

We thank Dr M. De Wilde (Smith-Kline/RIT) for helpful discussions and generous hospitality in his laboratory. A.vd S. is an IRSIA grantee and A.B. is Maître de Recherches at the Fonds National de la Recherche Scientifique. This work was financially supported by Smith-Kline/RIT.

REFERENCES

[1] Bowman, B.H. and Kurosky, A. (1982) Adv. Human Genet. 12, 189–454.
[2] Putnam, F.W. (1975) in: The Plasma Proteins (Putnam, F.W. ed) vol.2, pp.1–50, Academic Press, New York.
[3] Higa, Y., Oshiro, S., Kino, K., Tsunoo, H. and Nakajuma, H. (1981) J. Biol. Chem. 256, 12322–12328.
[4] Greer, J., Liao, W.-D.L. and Brown, W.E. (1981) J. Biol. Chem. 256, 8771–8774.
[5] Van der Straten, A., Herzog, A., Jacobs, P., Cabezón, T. and Bollen, A. (1983) EMBO J. 2, 1003–1007.
[6] Haugen, T.M., Hamley, J.M. and Heath, E.C. (1981) J. Biol. Chem. 256, 1055–1057.
[7] Misumi, Y., Tanaka, Y. and Ikehara, Y. (1983) Biochem. Biophys. Res. Commun. 114, 729–736.
[8] Chow, V., Murray, R.K., Dixon, J.D. and Kurosky, A. (1983) FEBS Lett. 153, 275–279

[9] Kurosky, A., Barnett, D.R., Lee, T.H., Touchstone, B., Hay, R.E., Arnott, M.S., Bowman, B.H. and Fitch, W.M. (1980) Proc. Natl. Acad. Sci. USA 77, 3388–3392.
[10] Grunstein, M. and Hogness, D. (1975) Proc. Natl. Acad. Sci. USA 72, 3961–3965.
[11] Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560–564.
[12] Grossman, L. and Moldave, K. (1980) Methods Enzymol. 65, 497–560.
[13] Blobel, G. (1980) Proc. Natl. Acad. Sci. USA 77, 1496–1500.
[14] Rangei, G., Bensi, G., Colantuoni, V., Romano, V., Santoro, C., Costanzo and Cortese, R. (1983) Nucleic Acids Res. 11, 5811–5819.
[15] Yang, F., Brune, J.L., Baldwin, W.D., Barnett, D.R. and Bowman, B.H. (1983) Proc. Natl. Acad. Sci. USA 80, 5875–5879.