Volume 9 Number 16 1981 Nucleic Acids Research

Mouse heavy chain variable regions: nucleotide sequence of a germ-line V_H gene segment

Roger Ollo, Charles Auffray, Jean-Louis Sikorav and François Rougeon

Unite de Genetique et Biochimie du Developpement, E.R.A. C.N.R.S. 851, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France

Received 17 June 1981

ABSTRACT

We have constructed a library of Balb/c mouse embryo DNA in the vector Charon 4A. The library was searched for sequences homologous to the V_H region of a cloned cDNA of the UPC10 heavy chain mRNA. In this paper, we describe the structure and the partial nucleotide sequence of one of such clones (V_H441). The nucleotide sequence of this germ-line gene indicates that it encodes amino-acids 1-98 of the X44 and J601 galactan-binding V_H regions, but that it differs from the UPC10 V_H segment by four single base changes. The V_H gene appears to contain a 101 bases long intervening sequence within a precursor sequence identical to the precursor sequence of UPC10. The 3' non coding sequence of the V gene contains the two conserved sequences found in embryonic V DNA segments, CACAGTG and ACATGAACC, separated by 23 nucleotides and a sequence CACTGTG separated by 33 nucleotides from the first heptamer.

INTRODUCTION

The variable regions of immunoglobulins contain three short polypeptide loops, the complementarity determining regions or CDRs, forming the antigen-binding site. Comparative studies of V regions have shown that the CDRs exhibit a higher amino-acid sequence variability than does the remainder of the V region. They are termed the hypervariable regions (HVs). It is generally assumed that sequence variations of the HV regions are correlated with the functional diversity of immunoglobulins (1-3). In the case of heavy chains, it has been shown that the third HV region or diversity region (D) is encoded separately by a discrete DNA segment. An active gene segment encoding a complete V_H region results from a two-step somatic recombination which positions a V_H gene segment next to a D gene segment and the D segment next to one of the four joining (JH) segments (4, 5, 6). As in the case for light chains, the junctional va-

riations may add to the diversity (7).

Hybridization experiments using V_H cDNA probes suggest that there are only 10 or 20 V_H germ-line genes corresponding to all members of the Balb/c V_HIII subgroup (8). The comparison of the sequences of V_H germ-line genes and rearranged V_H genes suggests that somatic mutations may take place during B cell differentiation (5, 6, 9).

We have recently described a cDNA plasmid containing a complete transcript of a γ2a heavy chain mRNA isolated from plasmacytoma UPC10 (10). This heavy chain bears the U10-173 determinant located on the V_H regions and belonging to the V_HIII subgroup (11, 12).

Using the V_H cDNA probe of the plasmid pG2a-10-21, we have isolated several germ-line genes for the V_H region of immunoglobulins heavy chain. We describe here the characterization of a V_H gene prepared from a λ Charon 4A library of the Balb/c embryo DNA. The nucleotide sequence of the germ-line gene establishes that it belongs to the U10-173 family and that it encodes the V_H segments of two anti-galactan myeloma proteins. The 3' non coding sequence reveals an unexpected organization of recognition sequences for V-D joining (4, 5, 6).

MATERIALS AND METHODS

1 - Chemicals and enzymes

T4 polynucleotide kinase, EcoR1, BamH1, restriction endonucleases were purified according to published procedures (13, 14, 15).

T4 DNA ligase, E. coli DNA polymerase I (Klenow fragment) and all other restriction endonucleases were purchased from New England Biolabs.

(γ - ^{32}P)ATP and (α - ^{32}P)dATP were obtained from Radiochemical Centre Amersham (England).

2 - Bacteria, phages and plasmids

The bacterial strains used for cloning experiments are E. coli 1106 (r_k^- m_k^-, supE supF) ; E. coli C600 (r_k^- m_k^-, recBC). The lysogens used for preparation of packaging mixtures BHB 2688 N205 rec A$^-$ (λ imm. 434 b2 red3 Eam 4 Sam 7)/λ and BHB 2690 N205 rec A$^-$ (λ imm. 434 cI ts red3 Dam 15 Sam 7)/λ were obtained from

B. Hohn. Charon 4A phage was obtained from F.R. Blattner and plasmid pBR325 (obtained from Boyer and Kochsts) was prepared according to Katz et al. (16). The plasmid pG2a-10-21 containing a full length transcript of a γ2a chain mRNA has been described previously (10).

3 - Construction of mouse embryo gene library

High molecular weight DNA from 15 days-old Balb/c mouse embryos was prepared according to Maniatis et al. and was partially digested with restriction endonuclease EcoR1 (17). The products were size fractionated on a 10-40% sucrose gradient. DNA fragments in the 12-20 kb range were isolated, pooled and ethanol precipitated. Charon 4A arms were prepared according to Maniatis et al. (17). The procedure used for in vitro packaging of recombinant DNA into phage particles was as described by Hohn and Murray and Collins and Hohn (18, 19). The packaging efficiency of ligated DNA was 6.10^5 phages/µg of DNA.

Fifteen separate packaging reactions were performed to obtain 8×10^6 in vitro packaged phages. The percentage of non recombinant phages in the preparation was determined by testing the phage for the lac 5 function as described by Blattner (20). The background of non recombinant phage DNA packaged was below 1%.

The library was screened for variable region genes by in situ plaque hybridization technique of Benton and Davis (21) using as a probe a nick-translated cloned cDNA of the UPC10 heavy chain mRNA (10).

Phages from plaques that were positive were replated three times until more than 95% of the phages gave positive hybridizations. To obtain a large amount of DNA, the EcoR1 digested fragments of the insert in V_H441 clone were subcloned into the EcoR1 site of pBR325 (see restriction map). The two subclones pV_H441-3 and pV_H441-4, which contain the 5' and 3' of the coding regions of V_H441 gene segment, were identified by rapid alkaline lysis method (22) and by Southern gel blotting according to Wahl et al. (23) and were used for sequencing.

4 - Restriction endonuclease analysis of cloned DNA

The restriction map was constructed by digestion with combination of restriction endonucleases and by two dimensional gel

electrophoresis with Sea plaque agarose (Marine Colloïds). All the probes were labeled by nick-translation as described (24). Two dimensional Southern blotting experiments were performed as described by Sato et al. (25).

5 - DNA sequence analysis

DNA fragments were labeled either at 5' end using γ - ^{32}P ATP and polynucleotide kinase in the exchange reaction ($\rightarrow$) or at the 3' end by filling in protruding restriction sites with E. coli DNA polymerase I using α - ^{32}P deoxynucleotides triphosphate. Partial chemical degradation was performed according to Maxam and Gilbert (26). Four base reactions were used (G, G + A, C + T, C). The products were analysed on 20% and 8% 0,35 mm thick urea polyacrylamide gels according to Sanger and Coulson (27).

RESULTS AND DISCUSSION

1 - Isolation of clones containing V_H genes

About 2.10^5 phages were screened by plaque hybridization with the pG2a-10-21 probe. We detected several positive clones and selected those hybridizing with the V_H cDNA probe. Five out of six clones yielded different EcoR1 digestion patterns. One recombinant phage V_H441 containing a 13.8 kb DNA insert has five EcoR1 fragments : 0.5 kb, 0.7 kb, 1.3 kb, 4.5 kb and 6.9 kb. Southern blots of the V_H441 clone digested with EcoR1 and hybridized to the UPC10 V_H cDNA probe gave two hybridizing fragments : the 0.5 kb and the 4.5 kb fragments.

In order to determine the number of V_H genes contained in this insert, a two dimensional blot according to Hutchinson et al. was performed (25). Using BamH1 fragments of the plasmid pG2a-10-21 as probes which gives 3' and 5' V_H cDNA probes, only two spots were detected ; the 5' V_H cDNA probe hybridized with the 0.5 kb fragment and the 3' V_H cDNA probe hybridized with the 4.5 kb fragment. The 13.8 kb insert bears only one V_H gene.

The two EcoR1 fragments with homology to our V_H probe were subcloned in the pBR325. Two clones bearing the 0.5 kb (pV_H441-3) and the 4.5 kb (pV_H441-4) fragments were detected by hybridization to the V_H DNA probe. A restriction map of pV_H441 clone was constructed by digestion with several enzymes alone or

in combinations and by two dimensional gel electrophoresis on low-melting agarose gels (Fig. 1).

2 - DNA sequence of the germ-line V_H441 clone

Using the method of Maxam and Gilbert (26), 727 nucleotides surrounding the V_H gene were sequenced. The sequencing strategy employed for the subclones pV_H441-3 and pV_H441-4 is outlined in Figure 2. As shown in Figure 3, this gene segment encodes a hydrophobic signal peptide identical to the peptide signal of UPC10 (10) and the first 98 amino-acids of a mature heavy chain of the U10-173 family (see below). As is the case for V_L and V_H genes, an intervening sequence (101 bp) is located between amino-acid position -5 and -4 (5). The nucleotide sequence includes 98 bp of the 5' flanking region and 180 bp at the 3' side of the codon 98.

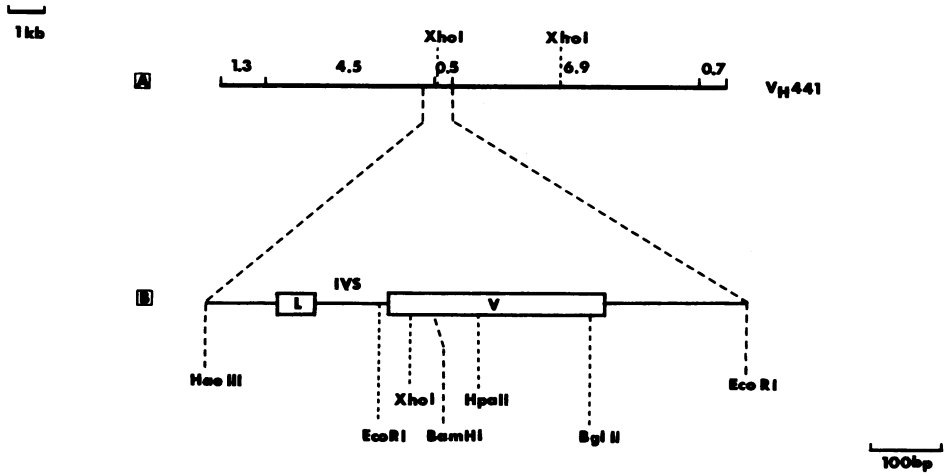

Fig. 1 : Diagram of restriction endonuclease cleavage sites in the V_H441 clone :

A - The top line represents the cloned 13,8 Kb EcoRl fragments in Charon 4A V_H441. Xho site was also indicated.

B - A magnified detail of 0,73 Kb HaeIII EcoRl fragment containing 5' untranslated region, leader (L), intervening sequence (IVS), variable structural germ-line gene (V) and 3' flanking region was figurated below. Restriction sites are indicated at bottom.

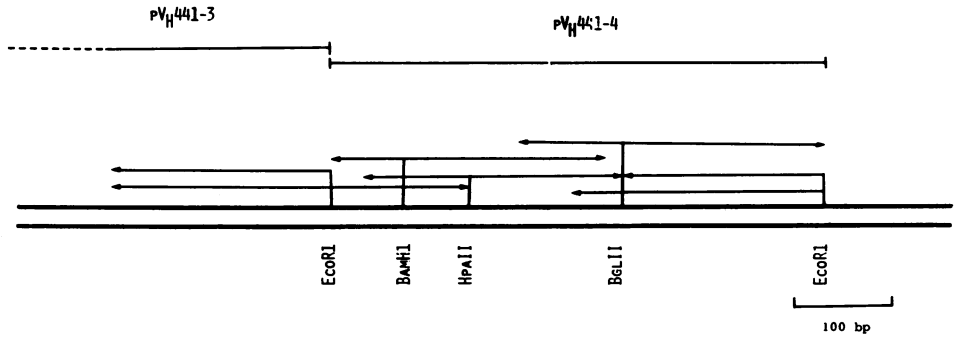

Fig. 2 : Sequencing strategy of the pV_H441-3 and pV_H441-4 subclones :

Cutting sites of restriction endonucleases used are indicated at bottom. Arrows show direction and extent of nucleotide reading.

```
                                                                                                                -18
                                                                                                                M   D
TCATGAAAACGAGTTCTGAACTAACCTTGAATCTGAAGCAAAGGGGATCAGCCCGAGATTCTCATTCAGTGATCAACACTGAACACACATCCCTTACC ATG GAT
1                                                  50                                                      100
     -15                  -10                  -5
  F  G  L  I  F  F  I  V  A  L  L  K
TTT GGG CTG ATT TTT TTT ATT GTT GCT CTT TTA AAA GGTAATTCATGGAAAAGAGATACTGAGTGTGTTACTGGTCATGAGCAAGATAGATGGT
                                                   150
                                        -4         -1  1              5                  10
                                        G  V  Q  C  E  V  K  L  L  E  S  G  G  G  L  V
GAGCCTGTATGGCAGTTTGCTGACAGAATTCTCTGTGTTTTCAG GG GTC CAG TGT GAG GTG AAG CTT CTC GAG TCT GGA GGT GGC CTG GTG
200                                             250
         15                  20                  25                  30                  35
  Q  P  G  G  S  L  K  L  S  C  A  A  S  G  F  D  F  S  R  Y  W  M  S  W  V  R  Q
CAG CCT GGA GGA TCC CTG AAA CTC TCC TGT GCA GCC TCA GGA TTC GAT TTT AGT AGA TAC TGG ATG AGT TGG GTC CGC CAG
         300                                                    GGA*  350
                                                                 G
 4C                  45                  50                  55                  60                  65
 A  P  C  K  G  L  E  W  I  G  E  I  N  P  D  S  S  T  I  N  Y  T  P  S  L  K  D
GCT CCA GGG AAA GGG CTA GAA TGG ATT GGA GAA ATT AAT CCA GAT AGC AGT ACG ATA AAC TAT ACG CCA TCT CTA AAG GAT
                        400                                                          TTT*          450
                                                                                      F
         70                  75                  80                  85                  90
  K  F  I  I  S  R  D  N  A  K  N  T  L  Y  L  Q  M  S  K  V  R  S  E  D  T  A  L
AAA TTC ATC ATC TCC AGA GAC AAC GCC AAA AAT ACG CTG TAC CTG CAA ATG AGC AAA GTG AGA TCT GAG GAC ACA GCC CTT
                                                   TTC*       500
                                                    F
     95          98
  Y  Y  C  A  R
TAT TAC TGT GCA AGA CCCACAGTGAGGAAATCTCAGTTTGTACCCAGACATGAACCTCACTGTGAGGTCGCTCACAACCACCAGGGGGCGATGAAAACCAGA
    TTC*              550                                                  600
     F

CACAGAACACAGGATCGCCCCAGGAAATGAGAAACAAGGAAGCATTCCTCTCATGTATCTGGGCAGCCTTTTTTTCTTAGTTTCTAGGAATTC
                  650                                                  700
```

Fig. 3 : Nucleotide sequence of the V_H441 gene :

The nucleotide sequence of the strand corresponding to the mRNA is displayed 5' to 3'. The amino-acid predicted by the nucleotide sequence is shown in italic letters (29). Only those codons that differ from the V_H441 segment are indicated for the V_H segment of UPC10.

3 - Comparison between V_H germ-line V_H441 sequence and V_H sequences sharing the U10-173 subgroup marker

The mouse cloned cDNA probe containing variable region used to screen the library was derived from myeloma UPC10, which belongs to the family which bears the U10-173 determinant described by Bosma et al. (11,12). We compared the amino-acid sequence deduced from the nucleotide sequence determined in this study with available protein sequences of those myelomas ; some of which have different ligand-binding specificities : 2-6 levan (UPC10) (Auffray et al., submitted), 1-6 D galactan (X44, X24, J539 and T601) (28).

As shown in Figure 3, the protein coding sequence of V_H441 and pG2a-10-21 differs by four single base changes responsible for four amino-acid substitutions. It is interesting to observe that there are no silent mutations and that variations are found both in hypervariable regions and in the framework region. The pattern of amino-acid substitution suggests that the UPC10 V_H region could derive from the V_H441 germ-line gene segment by a mechanism of somatic diversification.

We have also compared the amino-acid sequence derived from the V_H441 gene segment to V_H sequences of IgA antibodies that bind galactan (28). As shown in Figure 4, two of these V_H regions (X44 and T601) could derive from the germ-line V_H441 gene segment without somatic mutations. The two remaining J539 and X24 V_H regions could have arisen from the V_H441 germ-line gene segment by three or two single base mutations.

The analysis by Gearhart et al. (9) of antibodies to phosphorylcholine has shown that :

1° V_H regions of IgG antibodies exhibit more diversity than V_H prototype sequence ;

2° the V_H segments demonstrate diversity in both framework and hypervariable regions ;

3° most of amino-acid substitutions can be explained by single base changes ;

4° IgA antibodies fall into two categories : those having V_H regions identical to the prototype V_H sequence and those having V_H regions differing from the prototype sequence.

Although we have not demonstrated that V_H regions of

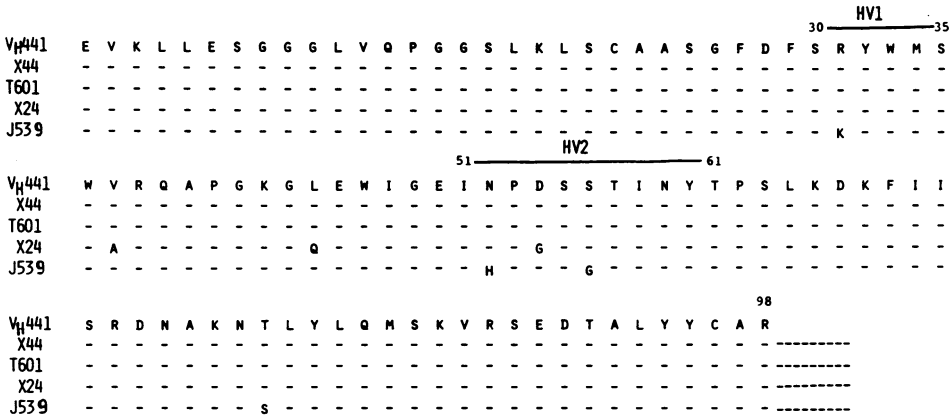

Fig. 4 : Comparison of the amino-acid sequence derived from the V_H441 nucleotide sequence with amino-acid sequences of four anti-galactan proteins (X44, T601, X24, J539) (28) :

The sequence of V_H441 shown on the top line was determined in this study. Below the amino-acid sequences of four anti-galactan proteins determined by Rao et al. (28) were aligned. Homologous positions were figurated with dashes and substitutions were mentioned in letter (29).

(HV : hypervariable region)

UPC10 IgG and antigalactan IgA molecules have arisen from the V_H441 gene segment, the patterns of variability in the two V_H families appear similar. It is obvious that only the analysis of the five V_H germ-line gene segments hybridizing to the cDNA V_H probe in stringent hybridization conditions will allow definitive conclusions.

4 - Sequence analysis of the 3' non protein coding region

Two blocks of short relatively well conserved sequences are found near the 3' end of germ-line V gene : the basic sequence, one heptamer CACAGTG and one nonamer ACAAAAACC (4, 5). The spacer between the two sequences is 12 $\pm$ 1 or 23 $\pm$ 1 base pairs long. These sequences are also found as inverted repeats at the 5' non coding region of the J segments. The D segments of heavy chain genes appear also to be flanked by two sets of sequences related to the consensus heptamer and nonamer sequences (6). It has been proposed that the two sequences are the recog-

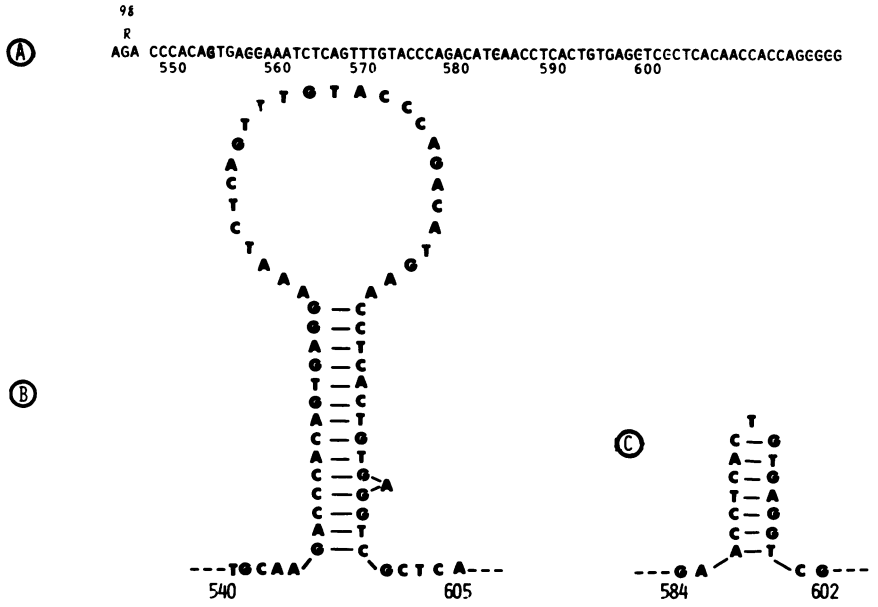

Fig. 5 : Alternative hairpin loop structures for the V_H441 3' non coding region sequence :

A - 3' non coding sequence ;

B - hairpin loop structure between the two self complementary heptamer sequences ;

C - palindromic sequence around the central T of the second heptamer.

nition sites for a recombinase (6).

As shown in Figure 5 A, the V_H441 gene contains the heptamer sequence CACAGTG near the 3' end of the protein coding segment and a nonamer sequence ACATGAACC separated from the heptamer by a 23 base pairs long spacer. Surprisingly, the V_H441 gene contains also an inverted repeat of the sequence of the heptamer CACTGTG located precisely 33 nucleotides downstream from the 3' end of the first heptamer. As shown in Figure 5B, the 10 base pairs long sequences surrounding the two heptamers can form a perfectly matched stem. Furthermore, a segment of 15 nucleotides including the second heptamer can be formed into a hairpin (Figure 5C). To suggest a possible role for the second

heptamer in a V-D joining recombination step would require the determination of the 5' nucleotide sequences of germ-line D segments.

ACKNOWLEDGEMENTS

We thank R. NAGEOTTE and B. CHAMBRAUD for excellent technical assistance and W. ROSKAM for helpful advice. We are grateful to J. SINGER and U. HIBNER for critical reading of the manuscript.

This work was supported by grants from the C.N.R.S. (A.T.P. 3663 and 4247) and the Fondation pour la Recherche Médicale Française.

REFERENCES

1 - Wu, T.T. and Kabat, E.A. (1970) J. Exp. Med. 132, 211-240.
2 - Kabat, E.A. (1976) Structural concept in Immunology and Immunochemistry, 2nd Ed. (New York, Rinchart, Winston).
3 - Amzel, L.M., Poljak, R.J., Saul, F., Varga, J.M. and Richards, F.F. (1974) Proc. Natl. Acad. Sci. USA 71, 1427-1430.
4 - Early, P., Huang, H., Davis, M., Calame, K. and Hood, L. (1980) Cell 19, 981-992.
5 - Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. and Tonegawa, S. (1980) Nature 286, 676-683.
6 - Sakano, H., Kurosawa, Y., Weigert, M. and Tonegawa, S. (1981) Nature 290, 562-565.
7 - Weigert, M., Perry, R., Kelley, D., Hunkafiller, T., Schilling, J. and Hood, L. (1980) Nature 283, 497-499.
8 - Rabbits, T.H., Matthyssens, G. and Hamlyn, P.H. (1980) Nature 284, 238-244.
9 - Gearhart, P.J., Johnson ,N.D., Douglas, R. and Hood, L. (1981) Nature 291, 29-34.
10 - Auffray, C., Nageotte, R., Chambraud, B. and Rougeon, F. (1980) Nucl. Acids Res. 8, 55-65.
11 - Bosma, M.J., DeWitt, C., Potter, M., Owen, J. and Taylor, B. (1977) Immune System : Genetic and Regulation. Sercarz, E.E., Herzenberg, L.A. and Fox, C.F. eds. (Acad. Press) 6, 99-105.
12 - Bosma, M.J., DeWitt, C., Hausman, S.F., Marks, R. and Potter, M. (1977) J. Exp. Med. 146, 1041-1053.
13 - Panet, A., VandeSande, J.H., Loewen, P.C., Khorana, H.G., Raae, A.J., Lillehaug, J.R. and Kleppe, K. (1973) Biochemistry 12, 5045-5049.
14 - Greene, P.J., Heyneker, H.L., Bolivar, F., Rodriguez, R.L., Betlach, M.C., Covarrubias, A.A., Backman, K., Russel, D.J., Tait, R. and Boyer, H.W. (1978) Nucl. Acids Res. 5, 2373-2380.
15 - Old, R., Murray, K. and Roizes, G. (1975) J. Mol. Biol. 92, 331-339.

16 - Katz, L., Kingsbury, D.T. and Helinsky, D.R. (1973) J. Bact. 114, 557-591.
17 - Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Efstratiadis, A. (1978) Cell 15, 687-701.
18 - Hohn, B. and Murray, K. (1977) Proc. Natl. Acad. Sci. USA 74, 3259-3264.
19 - Collins, J. and Hohn, B. (1978) Proc. Natl. Acad. Sci. USA 75, 4242-4246.
20 - Blattner, F.R., Blech, A.E., Denniston-Thompson, K., Faber, H.E., Richards, J.E., Slighton, J.L., Tucker, P.W. and Smithies (1978) Science 202, 1279-
21 - Benton, W.D. and Davis, R.W. (1977) Science 196, 180-182.
22 - Birnboim, H.C. and Doly, J. (1979) Nucl. Acids Res. 7, 1513-1523.
23 - Wahl, G.M., Stern, M. and Stark, G.R. (1979) Proc. Natl. Acad. Sci. USA 76, 3683-3688.
24 - Rigby, P.W., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
25 - Sato, S., Hutchinson, C.A. III and Harris, J.I. (1977) Proc. Natl. Acad. Sci. USA 74, 542-546.
26 - Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-566.
27 - Sanger, F. and Coulson, A.R. (1978) F.E.B.S. Letters 87, 107-116.
28 - Rao, D.N., Rudikoff, S., Krutzsch, H. and Potter, M. (1979) Proc. Natl. Acad. Sci. USA 76, 2890-2894.
29 - Dayhoff, M.O. (1976) Atlas of Protein Sequences and Structure (National Biomedical Research Foundation, Silver Spring, M.D.) 5, 189-190.