Bipartite mRNA for chicken α -fibrinogen potentially encodes an amino acid sequence homologous to β - and γ -fibrinogens

(second open reading frame/RNA splicing/gene family/fibrinogen evolution)

LAWRENCE WEISSBACH AND GERD GRIENINGER*

The Lindsley F. Kimball Research Institute of the New York Blood Center, New York, NY 10021

Communicated by Sidney Udenfriend, April 20, 1990 (received for review March 9, 1990)

ABSTRACT Overlapping cDNAs derived from the chicken α -fibrinogen mRNA have been sequenced, beginning from within the coding region for the signal peptide of this subunit and terminating within the poly(A) extension. The predicted size of chicken α -fibrinogen is 54,187 daltons, which is the smallest of any α chain reported; the oligopeptide repeats that characterize the central regions of the other α subunits were conspicuously absent. A further unexpected finding was the presence on the mRNA of a separate, long open reading frame (752 nucleotides), beginning 312 nucleotides downstream from the α -fibrinogen coding sequence and containing intron-like features near its 5' end. The protein sequence predicted from this second open reading frame lacks an initiating methionine but is homologous to the C-terminal regions of all known β - and γ -fibrinogens as well as the C termini of two nonfibrinogen proteins: cytotoxin (tenascin), an extracellular matrix protein, and pT49, a putative protein specific to cytotoxic T cells. The intron-like features of the second open reading frame immediately precede the region of common homology, and the beginnings of the corresponding homologous segments in the β - and γ -fibrinogen sequences are marked by aligned intron positions. Based on these findings, it is proposed that fibrinogen gene evolution included a fusion of two distinct ancestral genes.

Fibrinogen, an acute-phase plasma protein, acts in the final step of blood coagulation by forming a fibrin clot after its N-terminal fibrinopeptides are removed by thrombin cleavage (1). It is a multichain molecule composed of two sets of α , β , and γ subunits in all vertebrate species studied, from human to the most primitive existing vertebrate, lamprey, representing lineages that diverged 400–500 million years ago (2, 3). In spite of their functional distinctions, the three fibrinogen subunits share many structural features, suggesting a common evolutionary history. In particular, within the N-terminal region of each polypeptide, there is a fixed-length α -helical segment, with distinctive Cys-Xaa-Xaa-Xaa-Cys motifs at each end. These join with parallel regions on the other subunit polypeptides to form a disulfide-ring interlocked supercoil (4).

The genetic information for the α , β , and γ subunits resides on three separate genes that are normally transcribed in a stoichiometric and primarily liver-specific manner (5, 6). The human genes are clustered on 50 kilobases (kb) of chromosome 4 (7) and contain two common intron positions in their N-terminal regions (8–10). These facts reinforce the hypothesis that the three subunit genes arose from the same ancestral gene. Based on the closer similarity of the β - and γ -fibrinogens to each other, Doolittle (2) and Henschen *et al.* (3) postulated that a primordial fibrinogen gene duplicated, at about the time of the emergence of plants and animals, to give

rise to an α gene and a pre- $\beta\gamma$ gene, the latter further duplicating to yield separate β and γ genes prior to emergence of the vertebrates. To date, no fibrinogen or fibrinogen-like protein has been identified in invertebrates. However, recent identification of two nonfibrinogen protein sequences sharing homology with β - and γ -fibrinogen—cytotactin, a cellular adhesion molecule also known as tenascin (11), and pT49, a putative cytotoxic T-cell-specific protein (12)—has opened the possibility of a more extended fibrinogen gene family.

To gain further insight into fibrinogen structure and function, we have cloned the cDNA coding for chicken α -fibrinogen,† the subunit whose synthesis has been found in chicken hepatocyte cultures to be the most sensitive to changes in the hormonal milieu (13). We report here not only on aspects of the chicken α -fibrinogen polypeptide sequence but also on the unusual structure of its mRNA that bears an additional open reading frame (ORF). This second ORF appears to encode a truncated homologue of β - and γ -fibrinogens and, as such, represents an intriguing addition to the fibrinogen gene family.

MATERIALS AND METHODS

Antibody screening of the λ gt11 cDNA expression libraries and hybridization with cDNA probes were performed according to standard procedures (14, 15). Plasmid DNA mini-preparations suitable as DNA sequencing templates were prepared by a modification of the boiling method (16). Sequencing of plasmid DNA templates by the Sanger dideoxy chain-termination method (17) using Sequenase (18) was carried out according to guidelines provided with the Sequenase version 2.0 sequencing kit (United States Biochemical), except that T4 gene 32 single-strand DNA binding protein was added during the annealing step to prevent plasmid renaturation during the polymerization reaction (19).

The cDNAs sequenced in this study (see legend to Fig. 1) were obtained from three different chicken liver cDNA libraries. ChFa85 was from a library constructed in pUC8 using mRNA derived from chicken liver polysomes immunoprecipitated with anti-fibrinogen antibody (6, 14). The screening was performed under nonstringent conditions with a rat β -fibrinogen probe (kindly supplied by G. Crabtree, Stanford University) (5). The other clones came from libraries constructed by Clontech in λ gt11. To obtain ChFa42.2, one library was probed with the ChFa85 fragment. Clones ChFa6.1 and ChFa6.2 were isolated independently by antibody screening of the other library, which had been prepared using poly(A)-containing RNA from rooster liver supplied by this laboratory. Residues 16–39 of the deduced α -chain sequence were verified by automated Edman degradation of

Abbreviations: ORF, open reading frame; FASORF, fibrinogen α -subunit second ORF; nt, nucleotide(s).

*To whom reprint requests should be addressed.

†The sequence reported in this paper has been deposited in the GenBank data base (accession no. M34096).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

the purified subunit derived from noncrosslinked fibrin (20), since Murano *et al.* (21) had reported a slightly different protein sequence within this region.

RNA sequencing was performed by the dideoxy chain-termination method using reverse transcriptase (15) with hepatic poly(A)-containing RNA as template and a complementary oligonucleotide within the second ORF of the chicken α -fibrinogen mRNA as primer. Polymerase chain reaction (15) was performed using chicken genomic DNA as template.

Deduced protein sequences were compared with entries in various data bases and were aligned with homologous sequences using the computer program FASTA, which is based on the algorithm developed by Pearson and Lipman (22), and adjusted by visual inspection.

RESULTS

Overlapping partial-length cDNA clones, covering 3000 nucleotides (nt) of the α -fibrinogen mRNA, were isolated from three independently generated chicken liver cDNA libraries by antibody and cDNA screening. They were subcloned into

pGEM-3Z and sequenced by the dideoxy chain-termination technique using primers complementary either to vector sequences or to specific chicken cDNA sequences. Approximately 75% of the total cDNA sequence was determined on both strands.

Nucleotide Analyses. Organization of the sequence was found to be rather unusual for a eukaryotic mRNA: two long ORFs separated by 312 nt (Fig. 1). The first ORF codes for α -fibrinogen, whereas the second potentially encodes a sequence with strong homology to the C-terminal regions of β - and γ -fibrinogen (see below). The second ORF, which we have termed FASORF (for fibrinogen α -subunit second ORF), begins at nt 1830 and covers 752 bases. It is followed by a typical 3' untranslated region containing a poly(A) signal and a poly(A) extension. In Northern analyses with a cDNA probe that overlaps both reading frames, only one major RNA species (*ca.* 3.9 kb) hybridized (unpublished data). FASORF-containing mRNA was confirmed directly by RNA sequencing (see *Materials and Methods*).

Despite the fact that FASORF resides on an apparently mature mRNA, its 5' region contains several nucleotide

FIG. 1. Chicken α -fibrinogen cDNA and deduced protein sequences. The nucleotide sequence was generated from four overlapping clones: ChFa6.1 (1–914); ChFa6.2 (509–1700); ChFa42.2 (1225–2737); and ChFa85 (2520–*ca.* 3000). The coding region for the mature subunit (Gln-1 to Lys-491) extends from nt 43 to 1515. A second ORF of 251 amino acids (italics) occupies most of the 3' end of the sequence (nt 1830–2582). Potential N-linked glycosylation sites are marked above the protein sequences with a diamond (♦). The cysteine at position 49 of the α -coding sequence, which has no counterpart in human α -fibrinogen, is marked with an asterisk (*). The poly(A) signal is denoted by an overline. Intron-like sequences in the 3' second ORF, including a polypyrimidine tract followed by an AG dinucleotide (double underline) and a consensus branch point sequence (underline), are indicated.

motifs that are characteristic of the 3' ends of introns. These include a string of thymidine nucleotides followed by an AG dinucleotide leading up to a presumptive 3' splice site after position 1875 and a potential branch point (where the 5' end of the intron forms a 2'-5' phosphodiester bond to an adenosine residue within the intron during pre-mRNA splicing) that is 37 nt upstream (23). The branch point adenosine, A-1839, is situated within the heptaoligonucleotide TACT-GAT, which matches a proposed consensus branch point sequence.

Deduced Protein Sequence Analyses. The sequence of α -fibrinogen shown in Fig. 1 begins within the signal peptide. Protein sequence data for chicken fibrinopeptide A define the N-terminal amino acid of the α subunit in the secreted fibrinogen as glutamine (24), and we have designated it as residue 1. The α -fibrinogen sequence of 491 amino acids predicts a mature polypeptide of 54,187 daltons prior to posttranslational modification. This is the smallest α chain yet described.

In Fig. 2, the sequence is compared with that of human α -fibrinogen (25–29). As in the human chain, an Arg/Gly thrombin cleavage site (between residues 15 and 16) delineates the end of fibrinopeptide A. There is also a common 111-amino acid-long segment bounded by cysteine motifs as well as an Arg-Gly-Asp sequence (residues 256–258) thought to be involved in fibrinogen binding to the integrin family of cell surface receptors (30). In these chains, the N-terminal region, encompassing the thrombin cleavage site and the conserved cysteine motifs (residues up to 166), displays the most homology (47% identity). Alignment of residues in the C-terminal region is concentrated primarily in the disulfide loop (residues 292–323). According to the algorithm of Chou and Fasman (31), significant portions of the sequence between the cysteine motifs contain α -helix permissive segments that coincide to a large degree with those in the human α chain (29). One end of this region (residues 148–160) shows an unusually high degree of identity, with 13 consecutive amino acids conserved. The physiological importance of this stretch, which corresponds to the location of an intervening sequence in the gene for human α -fibrinogen (8), may be related to its recently identified properties as an epitope that becomes exposed only when fibrinogen is converted to fibrin (32).

FIG. 2. Alignment of chicken and human α -fibrinogen protein sequences. An optimal alignment between chicken and human α -fibrinogen was generated by computer analysis and visual inspection. Single-letter abbreviations for amino acids are used. Aligned identical residues are connected by a vertical dashed line. The cysteine motifs are boxed, and a long, horizontal, hyphenated line is used to indicate that there is no counterpart in the chicken sequence to the region where the 13-amino acid repeats occur in the human polypeptide. The human α -fibrinogen sequence is from Rixon *et al.* (25).

Despite these similarities, however, we found no sequence in the chicken protein that corresponded to the central portion of human α -fibrinogen, a region that encompasses a series of 13-amino acid repeats. Analogous regions are present to varying degrees in other species (8, 33), and the possibility that in some cases they arose independently (33) may mean that the minimalist form of chicken α -fibrinogen most closely resembles that of the ancestral α chain. Two features that are not found in either the human, rat, or lamprey α subunits were also readily apparent. One is a cysteine residue, at position 49, that is within the first motif of the disulfide ring structure that interlocks the three chains. The other is a recognition site for N-glycosylation (residue 62) located between the two cysteine motifs. This site may be more exposed in the oligomeric fibrinogen molecule (1, 4) than potential carbohydrate attachment sites found in α -fibrinogens of other species and could account for the unique presence of glycosylated α chains (up to 40%) in chicken fibrinogen (34).

The amino acid sequence deduced from the second reading frame, FASORF, is equivalent in size to a 28,367-dalton polypeptide, although there is no initiating methionine codon in frame. Computer-based comparison of this sequence with the Dayhoff protein data bank revealed a remarkable homology to the C-terminal regions of all β - and γ -fibrinogens characterized to date. Sequence information we have generated on the chicken β and γ chains indicates that the FASORF is not transcribed from either of these subunit genes (unpublished data). In Fig. 3, the sequence of the deduced FASORF protein is directly compared with the C-terminal regions of the human β and γ chains. The degree of homology exhibited by the FASORF protein sequence with respect to each chain individually is exceptional in this region, the number of identical residues (45% and 44%, respectively) being equivalent to that existing between the two fibrinogen chains themselves (40%). Virtually the same degree of identity is found between the FASORF polypeptide sequence and the β and γ chains of lamprey fibrinogen (36, 37). Of the FASORF protein's two potential N-linked glycosylation sites (Asn-71 and Asn-105), the one at position 71 is closely aligned with a utilized glycosylation site at position 203 in lamprey γ -fibrinogen. Amino acid residues 51–60 in the FASORF polypeptide show a strong resemblance to the recently identified thrombospondin binding domain between residues 243 and 252 of human β -fibrinogen (38). However, the FASORF protein sequence lacks an attachment site for carbohydrate conserved in all β -fibrinogens (at position 364 in humans) and has no parallel to the cysteine found at

FIG. 3. Similarity of the chicken FASORF protein sequence to the C-terminal regions of human β - and γ -fibrinogen. An alignment profile between the chicken FASORF protein and human β - and γ -fibrinogen is shown. Identity between residues in the FASORF protein and either of the two human fibrinogen subunits is depicted by a vertical dashed line. Aligned identical amino acids in all three proteins appear in the line below the compared sequences. The human β - and γ -fibrinogen sequences are from Chung *et al.* (9, 35).

position 286 in human β -fibrinogen (9). Also, no counterpart exists to the 21 C-terminal amino acids of the human γ chain (residues 393–411), which include a crosslinking site as well as a platelet recognition domain (1).

In Fig. 4, the FASORF amino acid sequence is compared with two recently identified nonfibrinogen protein sequences that also share homology with β - and γ -fibrinogen, those of cytotoxin and pT49. (Like the FASORF protein, these protein sequences contain neither the cysteine in β -fibrinogen nor the crosslinking and platelet recognition domains of γ -fibrinogen described above.) Here, too, a striking degree of identity was evident (cytotoxin, 33%; pT49, 39%), overlapping most of the residues conserved between the FASORF polypeptide and the authentic β - and γ -fibrinogen chains.

DISCUSSION

The data presented here demonstrate that the mRNA for chicken α -fibrinogen is bipartite, its 3' sequence containing a second open reading frame, FASORF. Unlike a classic pseudogene, FASORF is transcribed, and, unlike the untranslated regions of RNA, it is free of stop codons. In light of the compelling homologies between the predicted FASORF polypeptide, the β and γ subunits of fibrinogen, pT49, and cytotoxin, the FASORF sequence is probably translated, though not necessarily in hepatocytes.

At present, the reason for FASORF's close physical association with the chicken α -fibrinogen coding region on the same mRNA remains a mystery, although it is tantalizingly suggestive of some role in fibrinogen metabolism. No counterpart has yet been found on other fibrinogen mRNAs (lamprey, rat, human) (8, 25, 33). The only other comparable phenomenon in higher eukaryotes of which we are aware is a recently described, chicken erythroblast polycistronic RNA containing three coding regions of the α -globin gene cluster; its significance is as yet unknown (39).

The strong homologies between the FASORF amino acid sequence and the β - and γ -fibrinogen C termini cover a stretch of about 250 amino acids representing the most highly conserved portion of the entire fibrinogen molecule. Intron-like features in the nucleotide sequence of FASORF suggest an RNA splice site coinciding with the beginning of this " $\beta\gamma$ -C-terminal homologous region." These features include a

polypyrimidine tract and AG dinucleotide, ending with position 1875, 37 nt downstream from a branch point adenosine. The separation by 37 nt is within the range found for almost all branch point/3' splice site regions (18–38 nt) (23). The one AG dinucleotide occurring between the consensus branch point and the polypyrimidine tract is unusual but not unique (40).

The presumptive splice site demarcating FASORF's $\beta\gamma$ -C-terminal homologous segment aligns precisely with intron positions at the very beginning of the corresponding homologous regions of the human β - and γ -fibrinogen chains. The fact that, in cytotoxin and pT49, the strong homologies also begin at this same general location and cover a similar C-terminal region suggests that all these homologues are members of an extended gene family linked by a discrete genetic element that has been preserved as a unit. The fact that the $\beta\gamma$ -homologous segment appears in the C terminus of each polypeptide may be related to the function(s) it performs.

Within the $\beta\gamma$ -C-terminus-like regions of these five homologues, identical residues are conserved in clusters that have remained largely intact (Fig. 4). This pattern of conservation supports the contention that the FASORF sequence is translated since it is usually associated with important structural and functional domains that are subjected to strong evolutionary constraints. For example, the Phe-Xaa-Thr-Xaa-Asp-Xaa-Asp-Xaa-Asp cluster (which includes residues 170–178 in the FASORF protein) appears to be associated with calcium binding (3). Most of the other conserved oligopeptide regions have not yet been assigned functional identities, although they could be important in the physiological associations, with inflamed and damaged tissues and with malignant tumors, that cytotoxin and fibrinogen are known to share (41–44).

It is improbable, however, that FASORF is translated directly from this message since it has no initiating methionine codon in frame and the region between it and the α -fibrinogen coding region is beset with stop codons in all three reading frames, thereby precluding read-through. In light of the intron-like features that characterize the 5' end of FASORF, a possible mechanism for synthesis of the protein might involve formation of a translation-competent mRNA by trans-splicing (45, 46) of FASORF to a sequence encoding an appropriate N-terminal region.

FASORF's intron-like features and protein homologies suggest to us several modifications of the prevailing scheme for fibrinogen gene evolution (Fig. 5). As an alternative to the single ancestral fibrinogen gene hypothesis, we postulate a minimum of two such genes, one representing an N-terminal region common to all three fibrinogen subunits and the other a C-terminal region shared by the β and γ subunits alone. Successive duplications of the $\beta\gamma$ -C-terminal region gene

FIG. 4. Alignment of FASORF protein sequence with cytotoxin and pT49 protein. The homology between the FASORF protein and the C-terminal region of cytotoxin and pT49 protein is presented. Alignment of identical residues between adjacent sequences is depicted by dashed vertical lines. Aligning amino acids shared by all three polypeptides are shown in the line below the compared sequences. Residues that are conserved in all five homologues (FASORF, β - and γ -fibrinogens, cytotoxin, and pT49) are underlined. Thirty-five of these 38 conserved amino acids are also found in lamprey β - and γ -fibrinogens (36, 37). Whereas pT49 shares a slight relationship with the fibrinogen chains outside this zone, cytotoxin does not. The chicken cytotoxin sequence is from Jones *et al.* (11), and the mouse pT49 sequence is from Koyama *et al.* (12).

FIG. 5. Proposed evolutionary scheme for the extended fibrinogen gene family. This scheme, involving at least two progenitor genes, modifies a previous hypothesis (2, 3) to incorporate recent findings. $\alpha\beta\gamma$ -NTR, an ancestral gene homologous to the N-terminal regions of all three fibrinogen subunits (including their second cysteine motifs). $\beta\gamma$ -CTR, a second ancestral gene homologous to the C-terminal regions found in β - and γ -fibrinogens as well as in the FASORF protein, pT49 and cytotoxin.

provided the basis for a branch(es) of the gene family that lead to FASORF, pT49, and cytactin. We estimate that such duplications occurred at least 700 million years ago, when the β and γ subunit genes diverged (2, 3), based upon the fact that, in this region, the homologues all share between 33% and 45% sequence identity with one another. According to our modified scheme, the $\alpha\beta\gamma$ -N-terminal region gene also underwent duplications, one product fusing with a $\beta\gamma$ -C-terminal region gene to form a pre- $\beta\gamma$ -fibrinogen branch and another evolving into the α -fibrinogen gene (possibly involving fusion at some point with an as yet unidentified α -C-terminal gene). These modifications would account for the following observations that are difficult to reconcile with the single gene hypothesis. (i) In the vertebrate fibrinogens sequenced to date, significant homology between all three subunits is found only in their N termini, with the C termini of the α subunits displaying no clear relationship to the corresponding portions of the β and γ subunits (29). (ii) Of the various intron positions located throughout the fibrinogen subunit genes, sites common to all three chains occur only in their N-terminal domains (two); an additional site common to only the β and γ chains occurs at the beginning of their C-terminal regions, and 10 others have no counterpart in any other chain (8–10).

It appears likely that the primordial $\beta\gamma$ -C-terminal region gene was free of introns. In support of this notion, there are no introns in this region either in pT49 (12) or, by polymerase chain reaction-based genomic sequencing, in the gene corresponding to FASORF (unpublished data). Within the corresponding segment of the β - and γ -fibrinogen chains, containing three and four intron positions, respectively, no two intron sites align, suggesting that they were inserted after divergence of the β and γ genes.

Identification and characterization of the proteins encoded by FASORF and pT49 may provide further insights into the functional potential preserved among the members of this extended gene family. Both proteins can be expected to bear some fibrinogen-like properties contributed by their $\beta\gamma$ -C-terminal regions. Further investigation may uncover a host of additional relatives belonging to this family, given the apparent propensity for duplication of the genetic element encoding the $\beta\gamma$ -C-terminal homologous domain. Moreover, since emergence of the vertebrates occurred 400–500 million years ago, considerably later than the postulated duplications of the ancestral fibrinogen genes, it should be possible to identify related proteins, if not bona fide oligomeric fibrinogens, in prevertebrates.

Indeed, during review of this paper, a report has appeared of two $\beta\gamma$ -homologous sequences in an echinoderm (47), a higher invertebrate, and recent computer searches have uncovered an additional related sequence in humans (48). Of particular note is that all these sequences share *ca.* 35% amino acid identity with FASORF, including many of the same conserved amino acid clusters, and all occupy a similar C-terminal position.

We are indebted to Drs. D. Banerjee and C. Redman for sharing their technical expertise and for helpful discussions, to Dr. R. Procyk for assistance with the purification of the subunits of chicken fibrin, and to Dr. R. Fields for performing the automated Edman degradation. We also thank our co-workers: Dr. T.-J. Liang, for his help during the initial cloning; and Dr. P. Plant, for preparing the mRNA and genomic DNA from chicken liver. We are especially grateful to Tellervo Huima for expert graphic services and to Kathe Hertzberg for editorial assistance. This work was supported in part by National Institutes of Health Grant HL-09011 and by a Grant-in-Aid from the American Heart Association.

- Doolittle, R. F. (1984) *Annu. Rev. Biochem.* **53**, 195–229.
- Doolittle, R. F. (1983) *Ann. N.Y. Acad. Sci.* **408**, 13–27.
- Henschen, A., Lottspeich, F., Kehl, M. & Southan, C. (1983) *Ann. N.Y. Acad. Sci.* **408**, 28–43.
- Doolittle, R. F. (1981) *Sci. Am.* **245** (6), 126–135.
- Crabtree, G. R. & Kant, J. A. (1981) *J. Biol. Chem.* **256**, 9718–9723.
- Plant, P. W. & Grieninger, G. (1986) *J. Biol. Chem.* **261**, 2331–2336.
- Kant, J. A., Fornace, A. J., Jr., Saxe, D., Simon, M. I., McBride, O. W. & Crabtree, G. R. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2344–2348.
- Crabtree, G. R., Comeau, C. M., Fowlkes, D. M., Fornace, A. J., Jr., Malley, J. D. & Kant, J. A. (1985) *J. Mol. Biol.* **185**, 1–19.
- Chung, D. W., Que, B. G., Rixon, M. W., Mace, M., Jr., & Davie, E. W. (1983) *Biochemistry* **22**, 3244–3250.
- Rixon, M. W., Chung, D. W. & Davie, E. W. (1985) *Biochemistry* **24**, 2077–2086.
- Jones, F. S., Hoffman, S., Cunningham, B. A. & Edelman, G. M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 1905–1909.
- Koyama, T., Hall, L. R., Haser, W. G., Tonegawa, S. & Saito, H. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1609–1613.
- Grieninger, G., Oddoux, C., Diamond, L., Weissbach, L. & Plant, P. W. (1989) *Ann. N.Y. Acad. Sci.* **557**, 257–271.
- Berger, S. L. & Kimmel, A. R., eds. (1987) *Methods in Enzymology*, Vol. 152.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
- Riggs, M. G. & McLachlan, A. (1986) *Biotechniques* **4**, 310–313.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Tabor, S. & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4767–4771.
- Kaspar, P., Zadrazil, S. & Fabry, M. (1989) *Nucleic Acids Res.* **17**, 3616.
- Procyk, R. & Blombäck, B. (1990) *Biochemistry* **29**, 1501–1507.
- Murano, G., Walz, D., Williams, L., Pindyck, J. & Mosesson, M. W. (1977) *Thromb. Res.* **11**, 1–10.
- Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
- Green, M. R. (1986) *Annu. Rev. Genet.* **20**, 671–708.
- Takagi, T., Finlayson, J. S. & Iwanaga, S. (1978) *Biochim. Biophys. Acta* **534**, 161–164.
- Rixon, M. W., Chan, W.-Y., Davie, E. W. & Chung, D. W. (1983) *Biochemistry* **22**, 3237–3244.
- Blombäck, B., Hessel, B., Iwanaga, S., Reuterby, J. & Blombäck, M. (1972) *J. Biol. Chem.* **247**, 1496–1512.
- Lottspeich, F. & Henschen, A. (1978) *Hoppe-Seyler's Z. Physiol. Chem.* **359**, 1451–1455.
- Lottspeich, F. & Henschen, A. (1978) *Hoppe-Seyler's Z. Physiol. Chem.* **359**, 1611–1616.
- Doolittle, R. F., Watt, K. W. K., Cottrell, B. A., Strong, D. D. & Riley, M. (1979) *Nature (London)* **280**, 464–468.
- Ruoslahti, E. & Pierschbacher, M. D. (1987) *Science* **238**, 491–497.
- Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 222–245.
- Schielen, W. J. G., Voskuilen, M., Tesser, G. I. & Nieuwenhuizen, W. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8951–8954.
- Wang, Y. Z., Patterson, J., Gray, J. E., Yu, C., Cottrell, B. A., Shimizu, A., Graham, D., Riley, M. & Doolittle, R. F. (1989) *Biochemistry* **28**, 9801–9806.
- Grieninger, G., Plant, P. W. & Kossoff, H. S. (1984) *Biochemistry* **23**, 5888–5892.
- Chung, D. W., Chan, W.-Y. & Davie, E. W. (1983) *Biochemistry* **22**, 3250–3256.
- Strong, D. D., Moore, M., Cottrell, B. A., Bohonus, V. L., Pontes, M., Evans, B., Riley, M. & Doolittle, R. F. (1985) *Biochemistry* **24**, 92–101.
- Bohonus, V. L., Doolittle, R. F., Pontes, M. & Strong, D. D. (1986) *Biochemistry* **25**, 6512–6516.
- Bacon-Baguley, T., Ogilvie, M. L., Gartner, T. K. & Walz, D. A. (1990) *J. Biol. Chem.* **265**, 2317–2323.
- Broders, F., Zahraoui, A. & Scherrer, K. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 503–507.
- Nelson, K. K. & Green, M. R. (1989) *Genes Dev.* **3**, 1562–1571.
- Kushner, I. (1982) *Ann. N.Y. Acad. Sci.* **389**, 39–48.
- Dvorak, H. F. (1986) *N. Engl. J. Med.* **315**, 1650–1659.
- Mackie, E. J., Chiquet-Ehrismann, R., Pearson, C. A., Inaguma, Y., Taya, K., Kawarada, Y. & Sakakura, T. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4621–4625.
- Mackie, E. J., Halfter, W. & Liverani, D. (1988) *J. Cell Biol.* **107**, 2757–2767.
- Borst, P. (1986) *Annu. Rev. Biochem.* **55**, 701–732.
- Choquet, Y., Goldschmidt-Clermont, M., Girard-Bascou, J., Kück, U., Bennoun, P. & Rochemaix, J.-D. (1988) *Cell* **52**, 903–913.
- Xu, X. & Doolittle, R. F. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2097–2101.
- Morel, Y., Bristow, J., Gitelman, S. E. & Miller, W. L. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6582–6586.