Primary structure of the mouse sperm receptor polypeptide determined by genomic cloning

ROSS A. KINLOCH, RICHARD J. ROLLER*, CAROLYN M. FIMIANI, DAVID A. WASSARMAN†, AND PAUL M. WASSARMAN

Department of Cell and Developmental Biology, Roche Institute of Molecular Biology, Roche Research Center, Nutley, NJ 07110

Communicated by H. Ronald Kaback, May 11, 1988

ABSTRACT The mouse sperm receptor, a glycoprotein called ZP3, is synthesized and secreted by growing oocytes. It is present in more than a billion copies in the unfertilized egg's extracellular coat, or zona pellucida. We have cloned and characterized a region of the mouse (CD-1) genome that spans 10 kilobases of the ZP3 locus. The genomic clones described encompass the entire ZP3 coding region, which contains eight exons. The exons were identified, mapped, and sequenced, yielding the entire primary structure of the ZP3 polypeptide chain (424 amino acids; M_r , 46,300), which includes a 22-amino acid signal sequence. In addition, sequencing of genomic clones has revealed some unusual features of ZP3 mRNA and a region just downstream of the ZP3 gene.

All mammalian eggs are surrounded by a relatively thick extracellular coat, called the zona pellucida, that participates at several steps in the fertilization pathway (1, 2). The mouse egg zona pellucida is $\approx 7 \mu\text{m}$ thick, contains $\approx 3 \text{ ng}$ of protein, and is composed of three glycoproteins, called ZP1 (M_r , $\approx 200,000$), ZP2 (M_r , $\approx 120,000$), and ZP3 (M_r , $\approx 83,000$), that associate with one another through noncovalent interactions to form an insoluble network of crosslinked filaments (1–4). During fertilization of ovulated eggs, ZP3, one of the glycoproteins, serves as both sperm receptor and inducer of the acrosome reaction (3, 5–7). The sperm receptor activity of ZP3 is attributable solely to certain of its O-linked oligosaccharides (3, 8), whereas the acrosome-reaction-inducing activity of ZP3 is attributable to its polypeptide chain as well (3, 9). Thus, ZP3 plays structural and functional roles in the zona pellucida.

ZP1, ZP2, and ZP3 are synthesized and secreted exclusively by growing and fully grown mouse oocytes (3, 10–13). ZP3 is synthesized as a M_r 44,000 polypeptide chain to which either three or four high mannose-type, N-linked oligosaccharides are added cotranslationally, giving rise to precursors with molecular weights of 53,000 and 56,000, respectively (12). The oligosaccharides are then processed to complex-type, presumably in the Golgi, and O-linked oligosaccharides are added, giving rise to the mature, M_r 83,000 form of ZP3 that is secreted by oocytes.

A cDNA encoding a 20-amino acid stretch of ZP3 polypeptide chain was isolated and partially characterized (14). Based on the reported sequence of the ZP3 cDNA fragment, we prepared a synthetic oligonucleotide probe, used it to screen CD-1 mouse genomic libraries, and partially characterized two overlapping clones.‡ Results of experiments presented here provide the entire primary structure of the ZP3 polypeptide chain, as well as information about ZP3 gene organization and some unusual features of ZP3 mRNA.

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.

MATERIALS AND METHODS

Construction and Screening of Mouse Genomic Libraries. Mouse liver DNA was partially digested with *Sau*3A or *Eco*RI, and fragments were inserted between the *Bam*HI sites of λ Charon 35 (15) or the *Eco*RI sites of λ DASH (Stratagene), respectively, as described by Maniatis *et al.* (16). Libraries were amplified on *Escherichia coli* LE392 and screened (2×10^6 plaques) by using an end-labeled, synthetic oligonucleotide complementary to ZP3 (60-mer; ref. 14). Two genomic clones were isolated and characterized, ZP3-G5 from the *Sau*3A library and ZP3-G9 from the *Eco*RI library.

Subcloning and Sequencing of Genomic DNA Fragments. The 10-kilobase (kb) *Eco*RI fragment of ZP3-G9 was isolated and subcloned into the *Eco*RI site of pGEM-3blue (Promega Biotec, Madison, WI), producing plasmid pGEM-G9/R-A. Three subclones, which together span the 10-kb *Eco*RI fragment, were generated as follows: The 4-kb and 1.5-kb *Bgl* II fragments of pGEM-G9/R-A were isolated and cloned into the *Bam*HI site of pGEM-3Z to produce plasmids pGEM-G9/Bg-2B and -G9/Bg-2D, respectively. The 5'-terminal 4.7-kb *Eco*RI fragment of ZP3-G5 was cloned into the *Eco*RI site of pGEM-3blue, to produce pGEM-G5/R1-A. Further subcloning of these three plasmids was carried out by isolating the desired fragments and cloning them into appropriately digested pGEM vectors (Promega Biotec). Sequencing of the pGEM-derived plasmids was carried out by using dideoxy chain-termination (17) as described by Promega Biotec.

Purification of Oocyte RNA. Fully grown oocytes were isolated from 21- to 28-day-old, female, Swiss albino mice (CD-1; Charles River Breeding Laboratories) by poking excised ovaries with fine steel needles under a dissecting microscope. All oocytes were freed of adhering follicle cells by repeated pipetting. Isolation of oocytes was carried out in Earle's modified medium 199 (GIBCO) supplemented with bovine serum albumin (3 mg/ml), sodium pyruvate (30 $\mu\text{g}/\text{ml}$) and containing dibutyl cAMP (100 $\mu\text{g}/\text{ml}$). Isolated oocytes were washed in isotonic phosphate-buffered saline to remove bovine serum albumin, transferred in a minimal volume to 30 μl of extraction buffer (50 mM Tris-HCl, pH 7.5/300 mM NaCl/5 mM EDTA/0.4% NaDodSO₄) containing carrier calf-liver tRNA (15 μg), and frozen on dry ice. After thawing, 20 μl of phenol was added, and the mixture was thoroughly Vortex mixed. Twenty microliters of chloroform/isoamyl alcohol, 24:1 (vol/vol), was then added, and the mixture was thoroughly Vortex mixed again. Aqueous

Abbreviation: nt, Nucleotide(s).

*Present address: Department of Molecular Genetics and Cell Biology, University of Chicago, Chicago, IL 60637.

†Present address: Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511.

‡The sequence reported in this paper is being deposited in the EMBL/GenBank data base (IntelliGenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03851).

and organic phases were separated by centrifugation, and RNA was precipitated from the aqueous phase by addition of ethanol (100 μ l) and stored at -20°C overnight.

Preparation of RNA Probes. High-specific-activity RNA probes were transcribed with T7 or SP6 RNA polymerase from template sequences cloned into pGEM vectors as described by Promega Biotec.

Ribonuclease Protection Assays. RNA preparations to be examined were combined with 2×10^6 dpm of probe RNA, lyophilized, and dissolved in 20 μ l of hybridization buffer containing 80% (vol/vol) formamide, 40 mM Mops (pH 6.6), 0.4 M NaCl, and 1 mM EDTA. RNA was denatured by incubation at 85°C for 5 min and hybridized by incubation at 45°C overnight. Then, unhybridized RNA was digested by adding 200 μ l of a mixture containing 10 mM Tris-HCl (pH 7.5), 300 mM NaCl, 5 mM EDTA, RNase A (40 μ g/ml), and RNase T1 (2 μ g/ml) and incubating for 1 hr at 37°C . Digestion was terminated by adding NaDodSO₄ (0.7%) and proteinase K (150 μ g/ml) and incubating for 15 min at 37°C . Protected RNA was purified by extraction with phenol/chloroform/isoamyl alcohol, 25:24:1 (vol/vol), and then by ethanol precipitation in the presence of carrier tRNA (15 μ g). Purified RNA was analyzed on 5% or 6% polyacrylamide/8 M urea gels. Bands were visualized by autoradiography of dried gels.

RESULTS

Genomic Map of ZP3 Clones. Two overlapping genomic clones were isolated. One, ZP3-G5, was isolated from a *Sau*3A genomic library and contains a 12.5-kb insert. Another, ZP3-G9, was isolated from an *Eco*RI library and contains a 19-kb insert. The relationship of these clones to one another is shown in Fig. 1A. Southern hybridization analysis on blots containing restriction endonuclease digests of ZP3-G5 and -G9 mapped an oligonucleotide probe (60-mer; ref. 14) used to screen the libraries to 5'-terminal *Eco*RI fragments. Accordingly, the 5'-terminal *Eco*RI fragment (10 kb) of ZP3-G9 was isolated, subcloned, and characterized by sequence analysis of both strands. The location of subcloned fragments is shown in Fig. 1B. A partial restriction map for sites employed in subcloning and exon mapping is presented in Fig. 1C.

Exon Number and Order. The pGEM vectors used here permitted synthesis of single-stranded probes from opposite strands of the cloned inserts. Direction of transcription was established by using pGEM-G5/R1-A; only probes transcribed from the T7 promoter hybridized to the 1.5-kb ZP3 mRNA (ref. 14; R.J.R., R.A.K., and P.M.W., unpublished

FIG. 2. Identification of ZP3 exons by RNase protection assays. Antisense RNA, transcribed from pGEM-G9/Bg-2B (A), pGEM-G9/Bg-2D (B), pGEM-G5/R1-A (C), and pGEM-G5/B (D) (see Fig. 1B), was protected with RNA from 250 isolated oocytes (A, lane a; B, lane b; C, lane b; D, lane c), liver RNA (A, lane b; B, lane c), or *E. coli* rRNA (C, lane a; D, lane b). Protected fragments and molecular weight standards (*Msp* I-digested pT7-2) (A, lane a; B, lane a; C, lane c; D, lane a) were resolved on 5% polyacrylamide/urea gels and visualized by autoradiography. Lengths (in nucleotides) of protected fragments (arrowheads) and molecular size standards (hash marks) are indicated. o, Origin.

data). ZP3 exons were identified by RNase protection assays with antisense probes generated by SP6 and T7 polymerase from pGEM-G9/Bg-2B, -G9/Bg-2D, and -G5/R1-A.

Eight protected fragments, 327, 120, 105, 183, 114, 91, 137, and 219 nucleotides (nt) long, were detected in our experiments (Figs. 1C and 2). Given the genomic location of the three subclones (Fig. 1B), the order of each set of protected fragments in the 5'-to-3' direction was established. The

FIG. 1. Organization of the ZP3 genomic locus of CD-1 mice. (A) *Eco*RI map of ZP3 clones, ZP3-G5 (12.5-kb insert) and ZP3-G9 (19-kb insert). Vector sequences are indicated by zigzag lines. The *Eco*RI fragment of clone ZP3-G9 contained in pGEM-G9/R-A is labeled. (B) Location of subclones derived from pGEM-G9/R-A. (C) Partial restriction map of pGEM-G9/R-A. R, *Eco*RI; S, *Sma* I; A, *Apa* I; H, *Hind*III; K, *Kpn* I; Bg, *Bgl* II; P, *Pst* I; B, *Bam*HI. Eight ZP3 exons are denoted by solid black boxes (I-VIII), seven introns are denoted by dashed lines, and exon-intron junctions are denoted by nucleotide number in the 5'-to-3' direction.

FIG. 3. Determination of locations of ZP3 exons by RNase protection assays. Antisense RNA, transcribed from pGEM-G9/S-A (A), pGEM-G9/K-A (B), pGEM-G9/P-A (C), *Pst* I-digested pGEM-G5/R1-A (D), and *Apa* I-digested pGEM-G5/R1-A (E) (see Fig. 1B), was protected with RNA from 250 isolated oocytes (A, C, and E, lane a; B and D, lane b), liver RNA (A, C, and E, lane b; B, lane c), or *E. coli* rRNA (D, lane a). Protected fragments and molecular weight standards (*Msp* I-digested pT7-2) (B, lane a; A, D, and E, lane c) were resolved on 5% polyacrylamide/urea gels and visualized by autoradiography. Lengths (in nucleotides) of protected fragments (arrowheads) and molecular size standards (hash marks) are indicated. (A) Protection of 230-nt truncation of 327-nt exon (Fig. 2A). (B) Protection of 327-nt exon and 97-nt truncation of 120-nt exon (Fig. 2A). (C) Protection of 105-nt exon and 98-nt truncation of 183-nt exon (Fig. 2B). (D) Protection of 219-nt exon and 92-nt truncation of 134-nt exon (Fig. 2C). (E) Protection of 190-nt truncation of 219 exon (Fig. 2C). o, Origin.

locations of five (327, 120, 183, 137, and 219 nt) of the eight protected fragments were then determined either by using subclones that truncated the fragments or by terminating probes within fragments at appropriate internal restriction sites (Fig. 3). For example, by using subclone pGEM-G9/S-A (Fig. 1B), a protected fragment 230 nt long was detected (Fig. 3A) that represents a truncated form of the 327-nt fragment seen in Fig. 2A. Comparisons of truncated and intact fragments permitted positioning of 5' and 3' ends within a few nucleotides. Two fragments, 114 and 91 nt long (Fig. 2C and D), were localized precisely by comparing the sequence of the genomic clone with that reported for a 60-nt ZP3 cDNA fragment (14). One fragment, 105 nt long (Fig. 2B), was localized to the only remaining open reading frame in subclone pGEM-G9/P-A (Figs. 1B and 3C); this location was confirmed by RNA gel blot analysis by using an antisense oligonucleotide probe. Scanning the genome for splice-site donor and acceptor sequences at locations predicted by analyses just described permitted precise mapping of the exons since consensus sequences (18) were found at all 14 intron-exon junctions. Thus the order and sizes of the eight ZP3 exons are 328, 119, 104, 184, 118, 92, 137, and 220 nt in the 5'-to-3' direction (Fig. 1C).

mRNA and Protein Sequence. With the boundaries of eight ZP3 exons defined, sequences may be combined to form a ZP3 mRNA sequence. Certain features of this sequence are worth noting. (i) The ATG initiator codon is located 20 base pairs from the 5' terminus of the first exon in pGEM-G9/R-A

FIG. 4. Primary structure of the ZP3 polypeptide chain predicted from exon sequences. The entire sequence of the 424-amino acid polypeptide is shown. The predicted 22-amino acid signal sequence is italicized and locations of six potential N-linked glycosylation sites (Asn-Xaa-Ser/Thr) are underlined.

(Fig. 1). (ii) The termination codon UAA forms part of the poly(A) addition signal AATAAA (19), resulting in an extremely short 3' untranslated region (≈ 10 nt). (iii) The sequence 5'-GGGATTGGTGGT-3', located 50 nt downstream of the 3' end of the mRNA, is tandemly repeated 21 times.

Translation of the ZP3 mRNA sequence, in the frame defined by the 60-mer complementary to a ZP3 cDNA, reveals an open reading frame of 1272 nt encoding 424 amino acids (M_r 46,300; pI 6.4) (Fig. 4). The N-terminal 22 amino acids are predicted to make up a signal sequence that is missing from the mature glycoprotein, leaving 402 amino acids of polypeptide chain (M_r , 43,900). This polypeptide (minus the signal sequence) has a calculated pI value of 6.5 and an amino acid composition consistent with previous determinations (20). In addition, it has six potential glycosylation sites for N-linked oligosaccharides, consistent with the number of N-linked oligosaccharides found on ZP3 (12). Four glycosylation sites are located within a stretch of ≈ 60 amino acids in the C-terminal region of the polypeptide chain, and two sites are only three amino acids apart (Fig. 4).

Two structure prediction methods were applied (21, 22) to the ZP3 polypeptide chain shown in Fig. 4. Results of both analyses indicate that the polypeptide chain contains little, if any, α -helix-forming potential. Rather, it consists of stretches of 5–29 residues in an extended chain conformation, interrupted by short (4–10 residues) stretches of reverse turns or coils. Analyses of the local hydrophobicity along the polypeptide (23) indicate that it is neither strongly hydrophobic nor hydrophilic, except for the signal sequence (residues 1–22) and two small domains (residues 363–369 and 385–407) that might be sequestered from an aqueous environment (Fig. 5). Finally, examination of the six potential N-linked glycosylation sites suggests that they are found in relatively hydrophilic regions that are predicted to form reverse turns or coil conformations.

ZP3 DNA and polypeptide sequences were used to search

FIG. 5. Hydrophobicity profile for the ZP3 polypeptide chain. The profile was obtained by Kyte-Doolittle (23) computer analysis, using a search length of 9 amino acids. Positive values (above abscissa) indicate hydrophobicity and negative values (below abscissa) indicate hydrophilicity. Vertical dashed lines are present every 25 amino acids along the 424-amino acid polypeptide, with the N terminus at the left. Arrowheads indicate three hydrophobic domains: one domain associated with the N-terminal signal sequence and two domains associated with the C-terminal region of the polypeptide chain. The range of the hydrophobicity scale is from -3.29 to $+3.29$.

GenBank[§] and Protein Identification Resource[¶] data bases, respectively. Significant sequence similarities with data base entries were not found in either case.

DISCUSSION

We have characterized 10 kb of the ZP3 genomic locus in CD-1 mice, as well as the primary structure of the entire ZP3 polypeptide chain. The eight-exon sequences provide 1302 nt of ZP3 mRNA coding sequence. Exons smaller than ≈ 20 nt could have been missed by the methods used due to a low signal (signal intensity is roughly proportional to fragment length) and the tendency of the background on gels to increase in this size range. On the other hand, maintenance of the reading frame throughout the composite sequence and the relative rarity of such short exons argue against the possibility of missed exons.

ZP3 mRNA has unusually short 3' and 5' untranslated regions. The former cannot be located within a single nucleotide, since it forms the 3' end of the ZP3 gene that has a poly(A) addition site and no consensus splice site. As the poly(A) addition site may be located a variable distance from the end of the exon, and there is uncertainty of a few nucleotides in measurements of protected fragment lengths, our assignment of the 3' end also has an uncertainty of a few nucleotides. Similarly, the length of the 5' untranslated region, estimated to be approximately 19 nt long, is significantly shorter than the length of the average (60–70 nt) 5' untranslated region (24). It is tempting to speculate that the unusual size of the ZP3 5' leader may be involved in regulating the dramatic decrease in ZP3 synthesis and mRNA levels during ovulation (refs. 3 and 25; R.J.R., R.A.K., P.M.W., unpublished data). In addition, it may be responsible for the failure of ZP3 mRNA to be translated in *in vitro* systems (P.M.W., unpublished data).

Ten nucleotides surrounding the ATG initiation codon of the ZP3 gene match those of a consensus sequence (GGCG GGACCATGG, underlined nucleotides are cytidines for ZP3) reported for initiation of translation in vertebrates (26). Furthermore, the sequence 5'-GTGTGGTTT-3' found 26 nt downstream from the poly(A) addition site of the ZP3 gene is similar to both the consensus sequence YGTGTTY, where Y is a pyridine (27), and G/T cluster (28). These sequence elements usually are located 24–30 nt downstream from the AATAAA signal and are thought to play a role in the formation of the 3' terminus of many polyadenylated mammalian mRNAs. Thus, the ZP3 sequence may be important in cleavage and processing of nuclear ZP3 RNA.

An unusual sequence consisting of 21 tandem repeats of a 12-mer was found only 50 nt downstream from the 3' end of

the ZP3 gene (Fig. 1C). This sequence may be specific to this region of the mouse genome (R.A.K., D.A.W., and P.M.W., unpublished data). Indeed, a computer search of the GenBank data base[§] revealed only a single related sequence—the reverse complement of the 12-mer found around the CAAT box of the 5' long terminal repeat of intracisternal A-particle genes in the Syrian hamster genome (29). No functional equivalence is implied. However, the rarity and location of this sequence in the mouse genome suggests that it may play a role in regulation of ZP3 gene expression. In this context, enhancers have been located on the 3' side of the chicken histone H5 (30) and β -globin (31, 32) genes, and repeats have been reported to function in regulation of expression in a variety of systems (33–38). Since ZP3 gene expression is stringently regulated with respect to cell type and stage of development (3, 4), it will be of interest to examine the potential role of this repeat as a controlling element.

The predicted primary structure of the ZP3 polypeptide chain (Fig. 4) consists of 424 amino acids, with the N-terminal 22 amino acids predicted to be a signal sequence. The latter has a short polar domain followed by a hydrophobic region consisting of 11 amino acids (39–41). The 4 amino acids around the cleavage site at residue 22 of the signal sequence comply with the $-3, -1$ rule for such sequences (41). Processing of the signal sequence leaves a polypeptide chain of 402 amino acids (M_r , 43,900). This is in good agreement with assessments of the molecular weight of ZP3 (M_r , 44,000) based on NaDodSO₄/PAGE analysis of intracellular precursors metabolically radiolabeled in the presence of tunicamycin (12). Furthermore, the sequence predicted for residues 184–200 is identical to that determined directly for a ZP3 peptide generated by digestion of the glycoprotein with V8 protease (R.J.R. and P.M.W., unpublished data). The N terminus of the ZP3 polypeptide chain is blocked (R.A.R. and P.M.W., unpublished data), suggesting that, after cleavage of the signal sequence, the N-terminal glutamine cyclizes to form a pyroglutamic residue (42).

As compared to the average vertebrate protein (43), ZP3 contains an unusually high number of histidine, valine, serine, threonine, and proline residues and a low number of methionine, tyrosine, lysine, and isoleucine residues. Serine plus threonine (72 residues) and proline (30 residues) represent $\approx 17\%$ and $\approx 7\%$, respectively, of the total amino acid composition. This high serine plus threonine content of ZP3 is not unexpected given the relatively high degree of O-linked glycosylation (4, 8). It is of interest to note that the middle third of the polypeptide chain (residues 119–259) includes 17 of the 30 proline residues, with two proline-rich domains (residues 119–159, 7 prolines; residues 208–259, 9 prolines) separated by a proline-poor domain (residues 159–207, 1 proline). The C-terminal 165 amino acids (residues 260–424) includes only 6 proline residues.

Examination of ZP3 primary structure reveals six potential glycosylation sites for N-linked oligosaccharides, four of which are clustered within 60 amino acids of one another in

[§]EMBL/GenBank Genetic Sequence Database (1987) GenBank (Bolt, Beranek, and Newman Laboratories, Cambridge, MA), Tape Release 54.0.

[¶]Protein Identification Resource (1987) Protein Sequence Database (Natl. Biomed. Res. Found., Washington, DC), Release 15.0.

the C-terminal half of the polypeptide chain (Fig. 4). The sequence, Asn-Xaa-Ser/Thr, where Xaa may be any amino acid (except possibly aspartic acid), is necessary for N-linked glycosylation and is always glycosylated *in vitro* when presented as a peptide or part of a denatured protein (44–46). However, the sequence is not always used *in vivo*, perhaps, because it is not always accessible to the oligosaccharide transferase as nascent protein folds (46). Mature ZP3 possesses either three or four complex-type N-linked oligosaccharides (12). Therefore, the six acceptor sequences are sufficient to account for the N-linked oligosaccharides found on ZP3. The extreme closeness of acceptor asparagine residues at positions 327 and 330 may mitigate against the use of both of them due to steric considerations.

Two computerized methods (21, 22) were used to analyze the primary structure of ZP3. Both analyses suggest that the sequence consists of regions of 5–29 residues in an extended chain conformation, interrupted by short, 4–10 residue, stretches of reverse turns or coils. Determination of local hydrophobicity along the sequence (23) indicates that the sequence is neither strongly hydrophobic nor hydrophilic, although two short hydrophobic domains are present at the C terminus, in addition to the signal sequence hydrophobic domain at the N terminus of the polypeptide chain (Fig. 5). An interesting aspect of these analyses is the finding that all six potential, N-linked glycosylation sites are located in relatively hydrophilic stretches of sequence that are predicted to form reverse turns or coils (46).

Knowledge of the primary structure of the ZP3 polypeptide should aid in designing experiments that address the functional and structural roles of ZP3 during development. For example, in the former case, it should now be possible to locate precisely those O-linked oligosaccharides involved in sperm binding to ZP3 (3, 8, 9). Whether the oligosaccharides are clustered in one region of the polypeptide chain sequence or are brought together by folding of the polypeptide chain will be of considerable interest. In the case of the structural role of ZP3, primary structure information should permit analysis of those regions of the polypeptide chain that interact with ZP2 to form dimers that serve as the repeating unit of zona pellucida filaments (2, 4, 47).

Note Added in Proof. While this report was in press, an identical primary structure for ZP3 polypeptide, based on cDNA cloning, was described by Ringuette *et al.* (48).

- Wassarman, P. M. (1987) *Science* **235**, 553–560.
- Wassarman, P. M. (1987) *Annu. Rev. Cell Biol.* **3**, 109–142.
- Wassarman, P. M., Bleil, J. D., Florman, H. M., Greve, J. M., Roller, R. J., Salzmann, G. S. & Samuels, F. G. (1985) *Cold Spring Harbor Symp. Quant. Biol.* **50**, 11–19.
- Wassarman, P. M. (1988) *Annu. Rev. Biochem.* **57**, 415–442.
- Bleil, J. D. & Wassarman, P. M. (1980) *Cell* **20**, 873–882.
- Bleil, J. D. & Wassarman, P. M. (1983) *Dev. Biol.* **91**, 121–130.
- Bleil, J. D. & Wassarman, P. M. (1986) *J. Cell Biol.* **102**, 1363–1371.
- Florman, H. M. & Wassarman, P. M. (1985) *Cell* **41**, 313–324.
- Florman, H. M., Bechtol, K. B. & Wassarman, P. M. (1984) *Dev. Biol.* **106**, 243–255.
- Bleil, J. D. & Wassarman, P. M. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1029–1033.
- Greve, J. M., Salzmann, G. S., Roller, R. J. & Wassarman, P. M. (1982) *Cell* **31**, 749–758.
- Salzmann, G. S., Greve, J. M., Roller, R. J. & Wassarman, P. M. (1983) *EMBO J.* **2**, 1451–1456.
- Shimizu, S., Tsuji, M. & Dean, J. (1983) *J. Biol. Chem.* **258**, 5858–5863.
- Ringuette, M. J., Sobieski, D. A., Chamow, S. M. & Dean, J. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 4341–4345.
- Loenen, W. A. M. & Blattner, F. R. (1983) *Gene* **26**, 171–179.
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Mount, S. M. (1982) *Nucleic Acids Res.* **10**, 459–472.
- Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214.
- Bleil, J. D. & Wassarman, P. M. (1980) *Dev. Biol.* **76**, 185–203.
- Chou, P. Y. & Fasman, G. D. (1978) *Adv. Enzymol.* **47**, 47–148.
- Garnier, J., Osguthorpe, D. & Robson, B. (1978) *J. Mol. Biol.* **120**, 97–120.
- Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
- Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872.
- Philpott, C. C., Ringuette, M. J. & Dean, J. (1987) *Dev. Biol.* **121**, 568–575.
- Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125–8131.
- McLauchlan, J., Gaffney, D., Whitten, G. L. & Clements, J. B. (1985) *Nucleic Acids Res.* **13**, 1347–1368.
- Birnsteil, M. L., Busslinger, M. & Strub, K. (1985) *Cell* **41**, 349–359.
- Ono, M. & Ohishi, H. (1983) *Nucleic Acids Res.* **11**, 7169–7179.
- Trainor, C. O., Stamler, S. J. & Engel, J. D. (1987) *Nature (London)* **328**, 827–830.
- Choi, O. R. & Engel, J. D. (1986) *Nature (London)* **323**, 731–734.
- Hesse, J. E., Nikol, J. M., Lieber, M. R. & Felsenfeld, G. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 4312–4316.
- Fujita, T., Shibuya, H., Hotta, H., Yamanishi, K. & Taniguchi, T. (1987) *Cell* **49**, 357–367.
- Kaneda, S., Takeishi, K., Ayusawa, D., Shimizu, K., Serio, T. & Altman, S. (1987) *Nucleic Acids Res.* **15**, 1259–1270.
- Herr, W. & Clarke, J. (1986) *Cell* **45**, 461–470.
- Hearing, P. & Shenk, T. (1986) *Cell* **45**, 229–236.
- Zenke, M., Grundstrom, T., Matthes, H., Wintzerith, M., Schatz, C., Wildeman, A. & Chambon, P. (1986) *EMBO J.* **5**, 387–397.
- Barrett, P., Clark, L. & Hay, R. T. (1987) *Nucleic Acids Res.* **15**, 2719–2736.
- von Heijne, G. (1983) *Eur. J. Biochem.* **133**, 17–21.
- von Heijne, G. (1984) *J. Mol. Biol.* **173**, 243–251.
- von Heijne, G. (1985) *J. Mol. Biol.* **184**, 99–105.
- Andrews, P. C., Nichols, R. & Dixon, J. E. (1987) *J. Biol. Chem.* **262**, 12692–12699.
- Doolittle, R. F. (1986) *Of Urfs and Orfs* (University Science Books, Mill Valley, CA).
- Pless, D. D. & Lennarz, W. J. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 134–138.
- Hart, G. W., Brew, K., Grant, G. A., Bradshaw, R. A. & Lennarz, W. J. (1979) *J. Biol. Chem.* **254**, 9747–9758.
- Struck, D. K. & Lennarz, W. J. (1980) in *The Biochemistry of Glycoproteins and Proteoglycans*, ed. Lennarz, W. J. (Plenum, New York), pp. 35–83.
- Greve, J. M. & Wassarman, P. M. (1985) *J. Mol. Biol.* **181**, 253–264.
- Ringuette, M. J., Chamberlin, M. E., Bauer, A. W., Sobieski, D. A. & Dean, J. (1988) *Dev. Biol.* **127**, 287–295.