Etemadi, A. B. (1980) *Biochim. Biophys. Acta 604*, 423–474.
Fisher, K. A. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 173–177.
Fontaine, R. N., & Schroeder, F. (1979) *Biochim. Biophys. Acta 558*, 1–12.
Hale, J. E., & Schroeder, F. (1982) *Eur. J. Biochem. 122*, 649–661.
Heast, C. W. M., Plasa, G., Kamp, D., & Deuticke, B. (1978) *Biochim. Biophys. Acta 509*, 21–23.
Henis, Y. I., Rimon, G., & Felder, S. (1982) *J. Biol. Chem. 257*, 1407–1411.
Hubbell, M. L., & Mc Connell, H. M. (1971) *J. Am. Chem. Soc. 93*, 314–326.
Keith, A., Bulfield, G., & Snipes, W. (1970) *Biophys. J. 10*, 618–629.
Kornberg, R. D., & Mc Connell, H. M. (1971) *Biochemistry 10*, 1111–1120.
Kutchai, H., Huxley, V. H., & Chandler, L. H. (1982) *Biophys. J. 39*, 229–232.
Op den Kamp, J. A. F. (1979) *Annu. Rev. Biochem. 48*, 47–71.
Pagano, R. E., Martin, O. C., Schroit, A. J., & Struck, D. K. (1981) *Biochemistry 20*, 4920–4927.
Paysant, M., Bitran, M., Wald, R., & Polonovski, J. (1970) *Bull. Soc. Chim. Biol. 52*, 1257–1264.
Rimon, G., Hanski, E., & Levitzki, A. (1980) *Biochemistry 19*, 4451–4460.
Rousselet, A., Guthman, C., Matricon, J., Bienvenüe, A., & Devaux, P. F. (1976) *Biochim. Biophys. Acta 426*, 357–371.
Schachter, D., Cogan, U., & Abbot, R. E. (1982) *Biochemistry 21*, 2146–2150.
Seigneuret, M., & Devaux, P. F. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 3751–3755.
Sheetz, M. P., & Singer, S. J. (1977) *J. Cell Biol. 73*, 638–645.
Tanaka, K. I., & Ohnishi, S. I. (1976) *Biochim. Biophys. Acta 426*, 218–231.
Van Deenen, L. L. M. (1981) *FEBS Lett. 123*, 3–15.
Williams, J. H., Kuchmak, M., & Witter, R. (1966) *Lipids 1*, 391–398.
Williamson, P., Bateman, J., Kozarsky, K., & Mattocks, K. (1982) *Cell (Cambridge, Mass.) 30*, 725–733.
Zachowski, A., & Devaux, P. F. (1983) *FEBS Lett. 163*, 245–249.

Amino Acid Sequence of Phosvitin Derived from the Nucleotide Sequence of Part of the Chicken Vitellogenin Gene†

B. Marion Byrne,* Alfred D. van het Schip, Jos A. M. van de Klundert,‡ Annika C. Arnberg, Max Gruber, and Geert AB

ABSTRACT: The amino acid sequence of the egg yolk storage protein phosvitin has been deduced from the nucleotide sequence of part of the chicken vitellogenin gene. Of the phosvitin sequence, 210 amino acids including the N-terminal residue are contained on one large exon, whereas the remaining six amino acids are encoded on the next exon. Phosvitin contains a core region of 99 amino acids, consisting of 80 serines, grouped in runs of maximally 14 residues interspersed by arginines, lysines, and asparagines. The serines of the core region are encoded by AGC and AGT codons exclusively and the arginines by AGA and AGG, which results in a continuous stretch of 99 codons with adenine in the first position. The N-terminal quarter of the phosvitin sequence contains 16 serines grouped in a cluster with alanines and threonines and coded mainly by TCX triplets. The C-terminal part includes 27 serines, preferentially coded by AGC and AGT, 13 histidine residues, and the sequence ...Asn-Gly-Ser... at which the carbohydrate moiety of phosvitin is attached. Heteroduplex formation between cloned DNAs from chicken and *Xenopus* vitellogenin genes shows that the phosvitin sequence contains a stretch of highly conserved sequence.

Phosvitin is an egg yolk protein derived from a large precursor molecule, vitellogenin, which is formed in the liver of oviparous vertebrates under estrogen induction. Containing 10% phosphorus, it is one of the most highly phosphorylated proteins in nature (Levene & Alsberg, 1901; Mecham & Olcott, 1949). The phosphate is monoesterified to mainly serine residues (Lipmann & Levene, 1932; Lipmann, 1983) by an as yet unidentified kinase. Serine accounts for more than 55% of the amino acids in this phosphoprotein (Allerton & Perlmann, 1965). Evidence for the grouping of serines in runs, sometimes of six or more residues long, was found (Williams & Sanger, 1959).

Clark (1970) fractionated hen phosvitin into two discrete phosphoprotein species with apparent molecular weights of 3.4×10^4 and 2.8×10^4. The amino acid sequence of the larger, more abundant, species has been partially determined at the N-terminus (Clark, 1973) and at the site of glycosylation (Shainkin & Perlmann, 1971). A highly phosphorylated fragment of composition $Ser_{43}Asp_2Lys_3Arg_3$ has been isolated by Posternak & Waegell (1964). Determination of the complete amino acid sequence has proven difficult because of high resistance to proteolytic cleavage and the unusual amino acid composition of phosvitin (Posternak & Waegell, 1964; Belitz, 1965; Clark, 1973). Clark & Dijkstra (1980) have recently constructed derivatives of phosvitin to facilitate sequencing.

For our studies on the mechanism of steroid-controlled gene expression, we isolated clones covering the 21-kb[1] vitellogenin gene from a chicken DNA library and established its exon–intron organization (Arnberg et al., 1981). Comparison with

† From the Department of Biochemistry, Groningen University, Nijenborgh 16, 9747 AG Groningen, The Netherlands. *Received July 3, 1984.* This work was carried out under the auspices of the Netherlands Foundation for Chemical Research (S.O.N.) with financial aid from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

‡ Present address: Laboratory for Medical Microbiology, Department of Bacteriology, Wassenaarseweg 64, 2333 AL Leiden, The Netherlands.

[1] Abbreviations: kb, kilobase(s); bp, base pair(s); kDa, kilodalton(s); CNBr, cyanogen bromide.

0006-2960/84/0423-4275$01.50/0

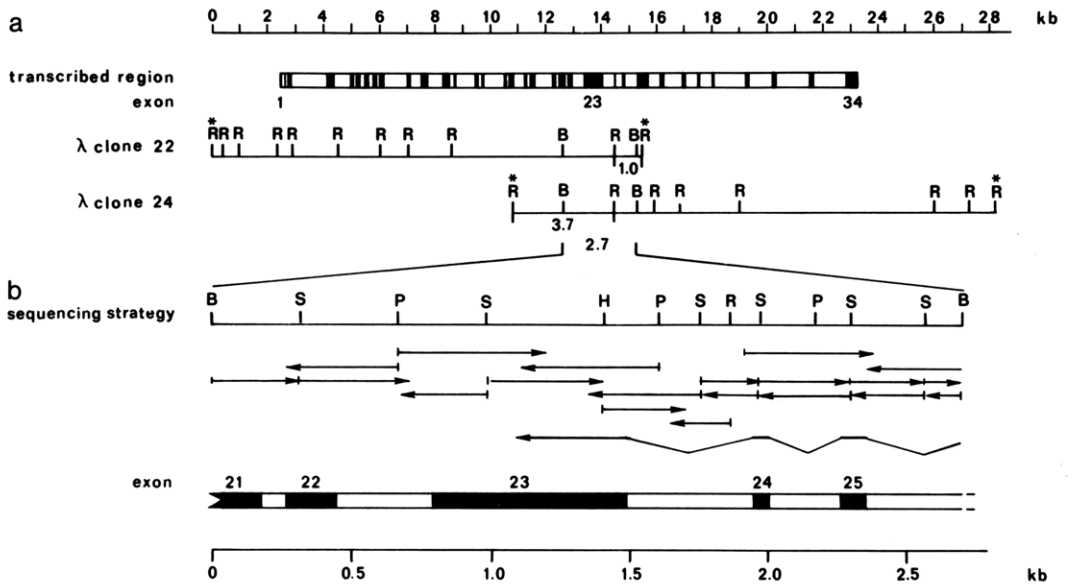

FIGURE 1: Restriction endonuclease map of two genomic clones covering the chicken vitellogenin gene and the sequence strategy for determining part of the nucleotide sequence. (a) The relative positions of the λ Charon A4 clones 22 and 24 with all *Eco*RI sites (R), two of the *Bgl*II sites (B), and the synthetic *Eco*RI sites introduced during the construction of the gene library (R*). *Eco*RI and *Bgl*II fragments used for detailed analysis are indicated by their sizes in kilobases. The transcribed region is drawn above the restriction map with the exons as black bars and the introns as open bars. (b) A 10-fold enlarged restriction map of the central section of the gene with restriction sites for *Bgl*II (B), *Sau*3AI (S), *Pst*I (P), *Hin*fI (H), and *Eco*RI (R). The arrows below the map indicate the direction and extent of the sequence determined for each fragment. The kinked line represents the cDNA clone. Exons 21–25 are indicated.

the vitellogenin A2 gene from *Xenopus laevis* (Wahli et al., 1980) showed a striking similarity in the number and length of the corresponding exons (Arnberg et al., 1981). Nucleotide sequence analysis of the promoter area and the first three exons showed considerable homology of those regions between both genes (Walker et al., 1983).

In heteroduplex analysis of the vitellogenin genes from chicken and *Xenopus* (Arnberg et al., 1981) our attention was drawn to a 300-bp stretch of homology in the central part of these genes, probably located in or close to exon 23, the largest exon. We have now determined the sequence of this section of the chicken vitellogenin gene, which shows that exon 23 codes for phosvitin. Here we present the amino acid sequence of phosvitin derived from its nucleotide sequence in the chicken vitellogenin gene.

Materials and Methods

Isolation of DNA Fragments. Genomic DNA fragments covering the central part of the chicken vitellogenin gene were isolated from λ Charon 4A clones 22 and 24 (Arnberg et al., 1981): two adjacent *Eco*RI fragments of 3.7 and 1.0 kb and a partially overlapping *Bgl*II fragment of 2.6 kb (Figure 1). The *Eco*RI fragments were subcloned in pBR328 (Soberon et al., 1980). cDNA-containing fragments were excised with *Msp*I from a pCR1 clone harboring an internal vitellogenin mRNA sequence (van den Boogaart, 1980) and purified by affinity chromatography on poly(U)–Sepharose using the property of poly(U) to form a triple helix with the poly(dA)·poly(dT) linkers used in the plasmid construction (Flavell & van den Berg, 1975).

DNA Sequence Analysis. *Pst*I and *Sau*3AI fragments of the genomic DNA fragments and the *Msp*I fragments of the cDNA clone were subcloned into M13 phage (Messing & Vieira, 1982) and subjected to sequence analysis by the chain termination method of Sanger et al. (1977). Part of the 3.7-kb *Eco*RI frgament was sequenced by the chemical method of Maxam & Gilbert (1980).

Electron Microscopy. Heteroduplexes between the 3.7-kb fragment of chicken DNA and Xlv128 DNA (Wahli et al., 1980) covering the entire *Xenopus* vitellogenin A2 gene were formed with the experimental conditions described earlier (Arnberg et al., 1981).

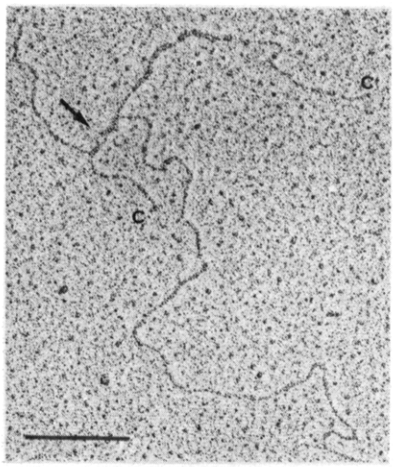

FIGURE 2: Electron micrograph of the heteroduplex structure formed between the 3.7-kb *Eco*RI fragment from the chicken vitellogenin gene and Xlv128 DNA covering the vitellogenin A2 gene from *Xenopus laevis*. Chicken DNA is denoted by C and the hybrid region by an arrow. Only part of the *Xenopus* DNA is visible. The bar is 0.2 μm.

Results

Alignment of the exon–intron map and the restriction endonuclease map (Arnberg et al., 1981) suggested that the region of homology between the chicken vitellogenin gene and the *Xenopus* vitellogenin A2 gene, as well as exon 23, was contained in a 3.7-kb *Eco*RI fragment of λ Charon 4A clone 24. From hybrids with vitellogenin mRNA, the fragment was found to cover a region extending from intron 17 to intron 23 (data not shown). Electron microscopy of heteroduplexes of the 3.7-kb *Eco*RI fragment with the cloned vitellogenin A2 gene from *Xenopus laevis* revealed a short region of homology, in or close to exon 23 (Figure 2), in agreement with earlier observations (Arnberg et al., 1981).

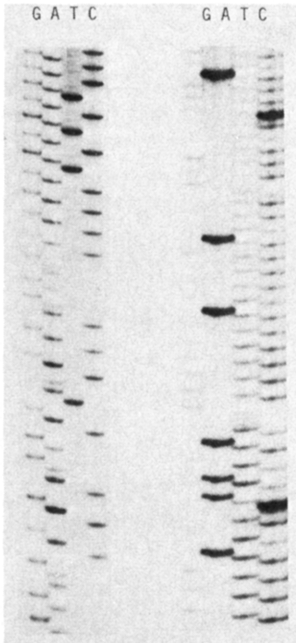

FIGURE 3: Autoradiographs of sequence gels of M13 clones containing exon 23 sequences. Sequence ladders of the messenger RNA like strand and the complementary strand of parts of the *Sau*3AI–*Pst*I fragment covering the 3′ section of exon 23 are shown on the right and on the left, respectively.

Nucleotide Sequence of the Exon 23 and 24 Region. Upon sequencing the presumed exon 23 region in the 3.7-kb *Eco*RI fragment, a highly unusual nucleotide sequence appeared. Representative autoradiographs of sequence gels from both strands are shown in Figure 3. Of the potential reading frames, one contains the serine-encoding triplets AGC and AGT in high abundance, suggesting that we had sequenced a gene segment coding for phosvitin. We proceeded to sequence the adjacent regions. The relevant restriction endonuclease sites and the sequencing strategy are shown in Figure 1.

The completed nucleotide sequence stretching from the *Pst*I site in intron 22 to the *Pst*I site in intron 24 is presented in Figure 4. The exon positions were determined by comparing the sequence of a cDNA clone stretching from exon 23 to exon 26 with the genomic DNA sequence. Although the boundary between intron 22 and exon 23 was not actually confirmed, we consider it highly probable that it is located at the indicated position. This position not only conforms to the consensus sequence for acceptor sites (Mount, 1982) but also yields an exon of the size (690 bp) expected for exon 23 and an open reading frame in phase with that of the preceding exon (data not shown). The amino acid sequence derived from the nucleotide sequence of exons 23 and 24 is also shown in Figure 4.

Discussion

Amino Acid Sequence of Phosvitin. Clark (1973) published the sequence of the first 37 amino acid residues at the N-terminus of the major phosvitin species. This sequence can be well aligned with the amino acid sequence encoded by exon 23, beginning with the alanine designated 1 (Figure 4). The first 15 amino acids are identical, the following 16 are differently grouped with two additional serine residues in our sequence, and the last 6 residues are identical up to the methionine at position 39, which was the site of cleavage with CNBr used by Clark. This establishes the position of the N-terminus of phosvitin in our sequence.

Hen phosvitin contains only one tyrosine residue, which is the C-terminal amino acid (Clark & Joubert, 1971). The first tyrosine in our sequence is encountered at position 216 (Figure 4). We conclude that 210 amino acids of phosvitin are encoded by exon 23 and the last 6 residues at the C-terminus by exon 24. Upstream from the phosvitin region, the vitellogenin gene contains about 3.3 kb of exon sequences, having a coding

```
                                                CTGCAGCTTCCATGGTGAGGCAGGAAAGGCTGCGGTGAGCAGCTTGTCTGACACCCACAGATATTTCACATTGTAAATACAGAAACAGATG
                                        Exon 23                                  -10
                                        Val Ala Asn Lys Thr Arg His Pro Lys Asn Arg Pro Ser Lys Lys Gly Asn Thr Val Leu
TATCCCCTCTCCTGTGATTTCCTTTGAATTGTTTTGCAG GTT GCA AAC AAA ACA CGG CAC CCG AAA AAT CGA CCA TCT AAG AAA GGA AAC ACT GTG CTA

 1                                   10                                      20                                      30
Ala Glu Phe Gly Thr Glu Pro Asp Ala Lys Thr Ser Ser Ser Ser Ser Ser Ala Ser Ser Thr Ala Thr Ser Ser Ser Ser Ser Ser Ala
GCA GAG TTT GGG ACA GAG CCT GAT GCA AAA ACT TCC TCC AGC TCA TCT TCT GCC TCC TCA ACT GCC ACC TCT TCT TCC TCA TCA TCT GCC

                                     40                                      50                                      60
Ser Ser Pro Asn Arg Lys Lys Pro Met Asp Glu Glu Glu Asn Asp Gln Val Lys Gln Ala Arg Asn Lys Asp Ala Ser Ser Ser Ser Arg
TCC TCT CCT AAT CGT AAA AAG CCT ATG GAT GAA GAG GAG AAT GAT CAA GTA AAG CAA GCA AGA AAC AAA GAT GCA AGC AGC AGC AGC AGG

                                     70                                      80                                      90
Ser Ser Lys Ser Ser Asn Ser Ser Lys Arg Ser Ser Ser Lys Ser Ser Asn Ser Ser Lys Arg Ser Ser Ser Ser Ser Ser Ser Ser Ser
AGC AGC AAG AGC AGT AAC AGC AGC AAG AGA AGC AGC AGC AAG AGC AGT AAC AGC AGC AAG AGA AGC AGC AGC AGC AGT AGC AGT AGC AGT

                                    100                                     110                                     120
Ser Ser Ser Arg Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Asn Ser Lys Ser Ser Ser Ser Ser Ser Lys Ser Ser Ser Ser Ser Ser
AGC AGC AGC AGG AGC AGC AGC AGT AGC AGC AGT AGT AGC AGT AAC AGC AAG AGC AGC AGT AGC AGC AGC AAG AGC AGC AGT AGC AGC AGC

                                    130                                     140                                     150
Arg Ser Arg Ser Ser Ser Lys Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Lys Ser Ser Ser Ser Arg Ser Ser Ser
AGG AGC AGA AGC AGC AGC AAG AGC AGC AGT AGC AGC AGC AGC AGT AGC AGC AGC AGC AGC AGC AAA AGT AGC AGT AGT AGG AGC AGT AGC

                                    160                                     170                                     180
Ser Ser Ser Lys Ser Ser Ser His His Ser His Ser His His Ser Gly His Leu Asn Gly Ser Ser Ser Ser Ser Ser Ser Ser Arg Ser
AGC AGC AGC AAG TCA AGC AGT CAC CAT AGC CAT AGC CAT CAT TCA GGG CAT CTA AAT GGC AGC AGC AGC AGC AGC AGC AGC AGC AGG TCA

                                    190                                     200                                     210
Val Ser His His Ser His Glu His His Ser Gly His Leu Glu Asp Asp Ser Ser Ser Ser Ser Ser Ser Ser Val Leu Ser Lys Ile Trp
GTG AGT CAC CAC AGC CAT GAG CAT CAC TCA GGA CAT CTG GAA GAT GAC AGC AGT AGC AGC AGC AGC AGC AGC GTG CTT TCC AAA ATA TGG
GTAACTCATATTTGTCAGTAAATCAAATGATACTGCCTCATGGCAGCTGGTTTTGTATCAAGTTATTAGTTGACACAGCAAATATTAGCAACATAAAATTTTAATATCTATTCTGCAAA
CACAACTGCAGTCATTGGCTCAGCTGTCTCTGTTAGAACATCTCAAATGTTTCTACAAGTTTGCTGAAACTTTTGTTGTGGGCAGGGAAGCATGTTTCACCAATTTATTGCTTTCCGTC
TATACAAATTGTTTGTTTATATCTTTATTTTGTAGTAGGATCATTAATGCTTTATATAGACTATTTCAACCATTTCAACTACTGCATTATGTGCTTACAACTAAGCGTACAGAACTAAG
                                                                                                                Exon 24
                                                                                                                Gly Arg His
CATAAGTGATAATTACTAGAATTCTGACTTAGTTCCCAGTGCTCAGATTTGAATACAATTGCAGCATTTCTTTTCTTCCAATGCTTTTCATCTCCTTCTTTTCAAAG GGG CGT CAT

                        220
Glu Ile Tyr Gln Tyr Arg Phe Arg Ser Ala His Arg Gln Glu
GAG ATT TAT CAG TAT CGC TTT AGA TCA GCA CAC AGA CAA GAG GTGAGTTTTCTATTCTTAGTTAATTGAGTTTGGAATCTGTTACGAGTTAGGTTAAAACATTGA
AGTCCACTTCTTGTCTTTGCTGAGGGAAACTGCAATACTGATAACACCAAATCAGCTGTGGCAGTTCTTCTGAGGCTCATACCAGTAATATGCTCCGTAACAAAGGATATTCTGCAG
```

FIGURE 4: Nucleotide sequence and derived amino acid sequence of a 1.5-kb fragment between the *Pst*I site in intron 22 and the *Pst*I site in intron 24 of the chicken vitellogenin gene. Numbers above the lines indicate the positons of the amino acids, taking the N-terminal amino acid of phosvitin as residue 1.

Table I: Codon Usage for Phosvitin in the Vitellogenin Gene

Phe	TTT	1	Ser	TCT	6	Tyr	TAT	1	Cys	TGT	0
Phe	TTC	0	Ser	TCC	6	Tyr	TAC	0	Cys	TGC	0
Leu	TTA	0	Ser	TCA	8		TAA	0		TGA	0
Leu	TTG	0	Ser	TCG	0		TAG	0	Trp	TGG	1
Leu	CTT	1	Pro	CCT	3	His	CAT	9	Arg	CGT	2
Leu	CTC	0	Pro	CCC	0	His	CAC	4	Arg	CGC	0
Leu	CTA	1	Pro	CCA	0	Gln	CAA	2	Arg	CGA	0
Leu	CTG	1	Pro	CCG	0	Gln	CAG	0	Arg	CGG	0
Ile	ATT	1	Thr	ACT	2	Asn	AAT	3	Ser	AGT	20
Ile	ATC	0	Thr	ACC	1	Asn	AAC	4	Ser	AGC	83
Ile	ATA	1	Thr	ACA	1	Lys	AAA	5	Arg	AGA	4
Met	ATG	1	Thr	ACG	0	Lys	AAG	10	Arg	AGG	5
Val	GTT	0	Ala	GCT	0	Asp	GAT	5	Gly	GGT	0
Val	GTC	0	Ala	GCC	3	Asp	GAC	1	Gly	GGC	1
Val	GTA	1	Ala	GCA	4	Glu	GAA	2	Gly	GGA	1
Val	GTG	2	Ala	GCG	0	Glu	GAG	6	Gly	GGG	3

capacity for a polypeptide of 125 kDa. A peptide of this size, which, as phosvitin, is derived from the precursor vitellogenin, is the larger subunit of β-lipovitellin (Bergink et al., 1974; Wang & Williams, 1980). The smaller 30-kDA polypeptide of β-lipovitellin should be encoded by the gene segment downstream from phosvitin, which contains about 1.8 kb of exon sequences including the 3′ nontranslated region.

The amino acid composition of phosvitin, based on the nucleotide-derived amino acid sequence, is in very good agreement with the amino acid composition determined by Clark and Joubert (Table I). Of the 123 serines, 80 are located in a core section (residues 56–154) that consists of runs of up to 14 serines alternated with the basic amino acids arginine and lysine and, occasionally, asparagines. Obviously, the fragment of phosvitin isolated by Posternak & Waegell (1964) originated from this section of the molecule. The N-terminal portion (residues 1–55) contains 16 serines in a cluster with threonines and alanines. The remaining 27 serines are located in the C-terminal portion (residues 155–216), which is particularly rich in histidines. In general, the amino acids are unevenly distributed in the chain, Ala, Pro, and Thr being restricted to the N-terminal part and Leu, Ile, Val, and His to the C-terminal part.

Phosvitin is a phosphoglycoprotein containing approximately 6.5% carbohydrate. Shainkin & Perlmann (1971) sequenced a glycopeptide from phosvitin containing 18 amino acid residues. In our opinion, this glycopeptide corresponds to the sequence 169–184 (Figure 4) that, however, does not contain the two serines flanking the glycosylated asparagine present in Shainkin and Perlmann's sequence. Therefore, the glycosylated asparagine is present in the sequence ...Asn-Gly-Ser..., which meets the structural requirements for carbohydrate attachment (Marshall, 1972).

Codon Usage. The frequency of each codon in the phosvitin sequence is given in Table II. There is a strong preference for the serine codons AGC and AGT. The serines of the core region are solely encoded by these codons and those of the C-terminal region in 22 out of 27 cases. Interestingly, the serine cluster in the N-terminal part of the phosvitin chain is almost exclusively encoded by TCX codons. Regarding the other amino acids, asparagine has a strong bias toward AGA and AGG.

After induction of phosvitin synthesis by estrogen in roosters, Mäenpää & Bernfield (1975) found that, of the serine isoacceptors, the relative amount of $tRNA_{AGU,AGC}$ was specifically increased in membrane-bound ribosomes of the liver. Their suggestion that this change reflects the codon usage of serines in phosvitin is supported by our data. Preference for the serine codons AGC and AGT has also been reported for the sericin genes from *Bombyx mori* (Okamoto et al., 1982).

Table II: Amino Acid Composition of Hen Phosvitin

amino acid	no. of residues: predicted by nucleotide sequence	no. of residues: determined by aa analysis[a]	amino acid	no. of residues: predicted by nucleotide sequence		no. of residues: determined by aa analysis[a]
Ala	7	8	Asn	7	13	13
Leu	3	3	Asp	6		
Ile	2	2	Gln	2	10	12
Val	3	3	Glu	8		
Pro	3	3	Gly	5		5
Phe	1	1	Lys	15		15
Trp	1	1	Arg	11		10
Met	1	1	His	13		12
Ser	123	125	Cys	0		0
Thr	4	5	total	216		220
Tyr	1	1				

[a] Amino acid (aa) composition of the protein as previously published (Clark & Joubert, 1971).

Evolutionary Aspects. The core section of serine repeats in phosvitin is interspersed with arginines, lysines, and asparagines at irregular positions. Vogel (1983) concluded from NMR studies that, on the one hand, a rapidly exchanging network of salt bridges between the basic amino acids and the phosphate moieties and, on the other hand, hydrogen bonding between the histidines and the phosphate groups confer some degree of folding to an otherwise extended conformation in the protein. The core region is entirely coded by triplets having adenine in the first and a purine in the second position. We speculate that this region might have evolved from a short poly(AGC) stretch by successive tandem duplications, accompanied by mutations to mainly basic amino acids. The conservation of AGC and AGT as the code words for serines in the core region seems plausible in the light of such a model. Duplications of considerable length should be noticed. The sequence Ser_2-Lys-Ser_2-Asn-Ser_2-Lys-Arg-Ser is present at positions 61–71 and 72–82 and is also perfectly duplicated at the nucleotide level. In addition, the sequence Thr-Ser_6-Ala-Ser_2 is repeated in the N-terminal section (positions 11–20 and 23–32), and an imperfect duplication exists in the C-terminal region (positions 154–178 and 179–203).

Phosvitins from various species, although differing in molecular weights, are similar in containing high contents of phosphoserines and low amounts of sulfur-containing and aromatic amino acids (Mano & Lipmann, 1966; Inoue et al., 1971; Weller, 1979). Therefore, some basic elements of the phosvitin structure must have been retained, probably those constituting the core region. Our heteroduplex experiment shows that the vitellogenin A2 gene of *Xenopus laevis* contains a stretch of about 300 nucleotides, homologous to that of the chicken phosvitin sequence. The electron microscopic analysis is not detailed enough to decide exactly which part of exon 23 is involved in the hybrid formation. A definite answer will come pending nucleotide sequence analysis of the relevant part of the *Xenopus* vitellogenin gene.

Added in Proof

While this paper was in press, Dr. J. J. Beintema from our department isolated a peptide with the amino acid sequence His-Glx-Ile-Tyr-Gln from a tryptic digest of phosvitin (Sigma Chemical Corp., St. Louis, MO). This identifies the C-terminal amino acid of phosvitin as glutamine-217 rather than the tyrosine proposed by Clark & Joubert (1971).

Acknowledgments

We thank Dr. W. Wahli for his generous gift of Xlv128 DNA. We thank Peter Bruinenberg for restriction mapping of the cDNA clone, Maaike Hiemstra and John Samallo for

preparing subclones, and Dr. Peter Terpstra for assistance in the computer analysis. We are grateful to Dr. Jaap Beintema for stimulating discussions.

Registry No. Phosvitin (chicken), 91384-58-6.

References

Allerton, S. E., & Perlmann, G. E. (1965) *J. Biol. Chem. 240*, 3892–3898.

Arnberg, A. C., Meijlink, F. C. P. W., Mulder, J., van Bruggen, E. F. J., Gruber, M., & AB, G. (1981) *Nucleic Acids Res. 9*, 3271–3286.

Belitz, H.-D. (1965) *Z. Lebensm.-Unters. -Forsch. 127*, 341–352.

Bergink, E. W., Wallace, R. A., van den Berg, J. A., Bos, E. S., Gruber, M., & AB, G. (1974) *Am. Zool. 14*, 1177–1193.

Clark, R. C. (1970) *Biochem. J. 118*, 537–542.

Clark, R. C. (1973) *Biochim. Biophys. Acta 310*, 174–187.

Clark, R. C., & Joubert, F. J. (1971) *FEBS Lett. 13*, 225–228.

Clark, R. C., & Dijkstra, J. (1980) *Int. J. Biochem. 11*, 577–585.

Flavell, R. A., & van den Berg, F. M. (1975) *FEBS Lett. 58*, 90–93.

Inoue, S., Kaneda-Hayashi, T., Sugiyama, H., & Ando, T. (1971) *J. Biochem. (Tokyo) 69*, 1003–1011.

Levene, P. A., & Alsberg, C. L. (1901) *Z. Physiol. Chem. 31*, 543–555.

Lipmann, F. (1983) *Trends Biochem. Sci. (Pers. Ed.) 8*, 334–335.

Lipmann, F. A., & Levene, P. A. (1932) *J. Biol. Chem. 98*, 109–114.

Mäenpää, P. H., & Bernfield, M. R. (1975) *Biochemistry 14*, 4820–4826.

Mano, Y., & Lipmann, F. (1966) *J. Biol. Chem. 241*, 3822–3833.

Marshall, R. D. (1972) *Annu. Rev. Biochem. 41*, 673–702.

Maxam, A., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.

Mecham, D. K., & Olcott, H. S. (1949) *J. Am. Chem. Soc. 71*, 3670–3679.

Messing, J., & Vieira, J. (1982) *Gene 19*, 269–276.

Mount, S. M. (1982) *Nucleic Acids Res. 10*, 459–472.

Okamoto, H., Ishikawa, E., & Suzuki, Y. (1982) *J. Biol. Chem. 257*, 15192–15199.

Posternak, T., & Waegell, P. (1964) *Helv. Chim. Acta 47*, 1575–1581.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Shainkin, R., & Perlmann, G. E. (1971) *J. Biol. Chem. 246*, 2278–2284.

Soberon, X., Covarrubias, L., & Bolivar, F. (1980) *Gene 9*, 287–305.

van den Boogaart, P. (1980) Thesis, University of Groningen.

Vogel, H. J. (1983) *Biochemistry 22*, 668–674.

Wahli, W., Dawid, I. B., Wyler, T., Weber, R., & Ryffel, G. U. (1980) *Cell (Cambridge, Mass.) 20*, 107–117.

Walker, P., Brown-Luedi, M., Germond, J.-E., Wahli, W., Meijlink, F. C. P. W., van het Schip, A. D., Roelink, H., Gruber, M., & AB, G. (1983) *EMBO J. 2*, 2271–2279.

Wang, S.-Y., & Williams, D. L. (1980) *Biochemistry 19*, 1557–1563.

Weller, M. (1979) in *Protein Phosphorylation* (Lagnado, J. R., Ed.) pp 134–142, Pion Ltd., London.

Williams, J., & Sanger, F. (1959) *Biochim. Biophys. Acta 33*, 294–296.