Phosvitin

Byrne, Brigid Marion

SUMMARY

The main objective of this study was to investigate the molecular basis for the heterogeneity of chicken vitellogenin (Vtg). That the emphasis lies on a particular domain within the Vtg gene, namely the phosvitin-encoding sequence, is largely coincidence. Most of the clones isolated "by my hands" from the VtgII gene (Chapter 3), the chicken liver cDNA library (Chapter 4) and the chicken "mini" gene library (Chapter 4) encoded this highly characteristic domain.

After a short introduction in Chapter 1, an up-to-date review of the state of the art on Vtg in diverse vertebrate and invertebrate species, with particular reference to its evolution, is presented in Chapter 2. The Vtg genes from the tiny roundworm, *Caenorhabditis elegans*, the amphibian *Xenopus laevis*, and the bird, *Gallus domesticus* share a common ancestor. Certain regions in the Vtg polypeptide (200 kDA) are more strongly conserved than others. Exons 1, 12, and 34 are well conserved and are presumably of functional importance. Only the function of exon 1 is known; it encodes the signal peptide sequence which is preferentially conserved in a large number of Vtgs (Chapter 2) and probably mediates the effective secretion of the Vtgs from their site of synthesis into the blood. Two mammalian polypeptides were recently found to share sequences with Vtg. Human apolipoprotein B-100 (apoB) shares common sequences with the N-terminal end of Vtg that encodes the lipovitellin heavy chain, and, to a lesser degree, with a C-terminal sequence. Human von Willebrand factor contains a repeated D domain, that is similar to the C-terminal end of Vtg. A detailed analysis of these relationships and speculations regarding their possible functional implications are outlined in Chapter 2.

Chapter 3 describes earlier work on the cloning of the region of the VtgII gene which hybridizes to the *X. laevis* VtgA2 gene in heteroduplex experiments. The region encodes the major phosvitin of chicken, that contains numerous serine residues which are mostly phosphorylated. Of the 217 amino acids in this phosvitin, 210 are encoded by one large exon, exon 23. It encodes 123 serine residues, 80 by the triplet AGC in the "core" region. The remaining residues in the "core" are arginines, lysines and asparagines, all encoded by triplets which could have arisen by single point mutations of the serine (AGPy) codon. The major chicken phosvitin is the largest of its kind to have been sequenced and has expanded from a smaller ancestor by successive tandem duplications. The, newly discovered, chicken VtgIII gene encodes an unusually small phosvitin-like region, called a phosvette (Chapter 4). Accordingly, the phosvette-encoding exon (tentatively numbered 23) is less than half the size of its VtgII counterpart (321

bp as opposed to 690) and may be highly reminiscent of the ancestral form. Exon 23 encodes only 28 serine residues in VtgIII. Alignment of t VtgII and VtgIII phosvitin sequences showed a lower degree of conservation at the nucleotide and protein level than the peripheral sequences. A similar finding had been made when the chicken VtgII and *X. laevis* Vtg gene sequences were compared, indicating that the phosvitin-encoding domain of the Vtg genes is evolving more rapidly (in size and sequence) than the remainder of the gene. Its rapid evolution may be advantageous for its function which is speculated to be concerned with embryonic bone formation. Most of the serines are phosphorylated and calcium may bind the phosphates. Phosvitin has only been found in the vertebrate Vtgs, to date.

No mammalian equivalent to phosvitin has been recognised. Although casein has a somewhat similar function and often contains short stretches of (phospho)serine residues, no other sequence similarity with phosvitin ha been observed. However, repeats similar to those in phosvitin ARE presen in the genome of eukaryotes. By "computer spotting" (Chapter 6), the AC repeats that encode serine in phosvitin, are found in numerous other transcribed regions, particularily the homeotic domains of *Drosophila melanogaster*, where they usually encode poly-glutamine. In screening a chicken gene library for Vtg gene sequences, using a probe containing th AGC repeats from the phosvitin-encoding region, two sequences were isolated containing such repeats (Chapter 6). Which proteins they may encode remains unknown for the present. In a limited RFLP analysis of genome of different chicken strains, using again an AGC repeat-containing probe, many phosvitin-related sequences were observed, some of which appeared to be strain-specific (Chapter 6).

The finding of a new chicken Vtg gene, VtgIII, by screening an estradiol-induced rooster liver cDNA expression library with antisera and a DNA probe (Chapters 4 and 5), has shown, for the first time, that chicken Vtg is also encoded by a multigene family, similar to the Vtgs i other oviparous vertebrates. Complementary DNA clones, covering about 3. kb of the VtgIII mRNA (total length 5.5 kb, approximately), including th 3' non-transcribed region, have been isolated and partially sequenced (Chapter 5). One region of exceptionally high sequence similarity between VtgII and VtgIII was observed in exon 22. It will be interesting to characterize the 3' non-transcribed region of the Vtg mRNA since it is expected that sequences related to Vtg mRNA stability are harboured there. The library, it is hoped, will serve other purposes in the near future, e.g the isolation of hormone-specific and/or liver specific DNA binding proteins and their mRNAs.