Ovoinhibitor Introns Specify Functional Domains as in the Related and Linked Ovomucoid Gene*

(Received for publication, December 1, 1986)

Maxwell J. Scott‡, Clark S. Huckaby†, Ikunoshin Kato§§, William J. Kohr∥∥, Michael Laskowski Jr.§§, Ming-Jer Tsai¶¶, and Bert W. O'Malley‡‡

From the ‡Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030 and the §Department of Chemistry, Purdue University, West Lafayette, Indiana 47907

We have isolated cDNA clones and determined the gene structure of chicken ovoinhibitor, a seven domain Kazal serine proteinase inhibitor. Using RNA blot hybridization analysis, the gene was identified initially as a region 9–23 kilobases upstream of the gene for the related inhibitor ovomucoid. Ovoinhibitor RNA appears in oviduct and liver. cDNA clones were identified by screening an oviduct cDNA library with a nick-translated DNA restriction fragment which contained an exon of the gene. The mature protein sequence derived from a cDNA clone is in excellent agreement with that which we obtained from direct sequencing of an ovoinhibitor cDNA. The protein-sequence strategy is reported. The P1 amino acids of the Kazal domains are consistent with the known broad inhibitory specificity of ovoinhibitor. The gene is about 10.3 kilobases in length and consists of 16 exons. Each Kazal domain is encoded by two exons. Like ovomucoid, introns fall between the coding sequences of the ovoinhibitor domains, an arrangement which may have facilitated domain duplication. The intradomain intron occurs in an identical position in all of the ovoinhibitor and ovomucoid Kazal domains, suggesting that this intron was present in the primordial inhibitor gene. We discuss the location of the intradomain intron in relation to the known structure of four Kazal inhibitors and suggest a scheme for the evolution of the ovoinhibitor gene.

Chicken ovomucoid, one of the major egg-white proteins, accounts for about 10% of the protein produced by the tubular gland cells of the oviduct of laying hens (Palmiter, 1972). Ovomucoid is a member of the Kazal family of serine proteinase inhibitors (Laskowski and Kato, 1980). The chicken genome probably contains the genes for at least four Kazal inhibitors (Laskowski et al., 1980). One of these Kazal inhibitors, ovoinhibitor (Matsushima, 1968), is present also in chicken egg white but at about one-tenth the amount of ovomucoid (Liu et al., 1971). Ovoinhibitor is, however, a major proteinase inhibitor in chicken plasma (Laskowski and Kato, 1980).

Ovomucoid consists of three tandem Kazal domains that can be readily separated by hydrolysis of the short connecting peptides (Laskowski et al., 1978; Kato et al., 1987). Each domain can function independently. Proteinase inhibition is due to the formation of a very stable complex between the conformationally constrained reactive site of an ovomucoid domain and the active site of the proteinase (Laskowski and Kato, 1980). All avian ovomucoid domains contain six cysteines in similar locations that form three intradomain disulfide bonds. However, the primary determinant of inhibitor specificity, the P1 amino acid of the reactive site (Schechter and Berger, 1967) and other residues involved in enzyme-inhibitor contact are highly variable among avian ovomucoid third domains (Laskowski et al., 1987).

The third, β-type domain of ovomucoid is distinguished from the first two, α-type domains by a shorter peptide between cysteines I and II (Laskowski et al., 1980). Moreover, the α domains of chicken ovomucoid are more homologous to each other than the β domains. Chicken ovoinhibitor consists of six α domains and one β domain. As in ovomucoid, the β domain of ovoinhibitor is at the carboxyl terminus. Laskowski et al. (1980) suggest that the genes of multidomain Kazal inhibitors such as ovomucoid and ovoinhibitor arose by duplication of a gene that coded for one α and one β domain, the latter occurring at the carboxyl terminus of the protein. Subsequently, the α domains must have duplicated in both genes.

The chicken ovomucoid domains are each encoded by two exons (Stein et al., 1980). The domain boundaries correspond to intron-exon junctions. Since the intradomain intron is located in an identical position in all three domains, Stein et al. (1980) suggest that the primordial inhibitor gene may have consisted of only the first exon of an inhibitory domain, which encodes the reactive site and that the intradomain intron was created by the subsequent addition of the second exon.

The studies presented here were conducted to examine the hypotheses of Laskowski et al. (1980) and Stein et al. (1980) for the evolution of Kazal inhibitors by determining the gene structure of other chicken Kazal inhibitors. In particular we wanted to isolate the gene for ovoinhibitor. As there are many examples of the close linkage of related genes that are expressed in the same tissue (Colbert et al., 1980; Efstratiadis et al., 1980), our approach was to determine which sequences in a 47-kb region around the ovomucoid gene were expressed in the oviduct and homologous to ovomucoid. This analysis identified a region about 9–23 kb upstream from the 5' end.
of the ovomucoid gene that codes for ovoinhibitor. Introns separate the exons coding for the Kazal domains and signal peptide of ovoinhibitor in a manner identical to that of the ovomucoid gene. We find that, the ovoinhibitor gene structure and sequence is consistent with the hypothesis of Stein et al. (1980) and Laskowski et al. (1980). We suggest a possible scheme for the evolution of the ovoinhibitor gene.

EXPERIMENTAL PROCEDURES

General—Procedures for the preparation of phage and plasmid DNA have been previously described (Kulomaa et al., 1986). Individual DNA fragments were purified from agarose gels by the method of Detzet et al. (1981). The sequencing of isolated DNA fragments was performed using either the method of Maxam and Gilbert (1977) or of Sanger et al. (1977). For the latter procedure the DNA fragments were cloned into M13 phage vectors (Messing, 1983). DNA and protein sequences were analyzed using the Microgen program from Beckman. All enzymes were purchased from either New England Biolabs or Boehringer Mannheim and used according to the suppliers’ recommendations.

RNA Blot Hybridization Analysis—Total RNA from various tissues was prepared, resolved on formaldehyde-agarose gels and transferred to nitrocellulose filters as previously described (Sargan et al., 1986). The filters were hybridized to nick-translated DNA fragments according to the procedures of Sargan et al. (1986).

DNA Blotting and Hybridization—Restriction enzyme-digested plasmid DNAs were resolved in agarose gels and transferred to nitrocellulose filters according to the procedures of Southern (1975). The filters were hybridized to nick-translated DNA fragments at moderate stringency using the conditions described by Levine et al. (1985).

Isolation of Ovoinhibitor cDNA Clones—An oligo(dT)-primed hen oviduct cDNA library was screened as previously described (Kulomaa et al., 1986). The probe used was a nick-translated 0.7-kb XhoI fragment from the clone OI4.3. Several cDNA clones were isolated and restriction mapped according to established procedures (Colbert et al., 1980). Overlapping DNA fragments of the longest clone O1c were sequenced.

Protein Sequence of Ovoinhibitor—“Crude ovomucoid” was prepared as described in Kato et al. (1987) and applied to a Bio-Gel P-10 column equilibrated with 5% formic acid. The large molecular weight fraction is “crude ovomucoid,” and the smaller molecular weight fraction is ovomucoid. Crude ovomucoid was further purified on a CM-Sepharose column at pH 5.5 and developed with NaCl gradient. Sequencing was carried out on a Beckman 890C sequenator.

RESULTS

Identification and Localization of the Ovoinhibitor Gene—From a comb library of chicken genomic DNA, we have isolated overlapping clones that contain the ovomucoid gene and about 23 kb of 5' and 17 kb of 3'-flanking DNA. We investigated whether any of the cloned DNA sequences flank- ing the ovomucoid gene were transcribed also in the oviduct. To this end, total RNA from either oviduct or liver was resolved on agarose-formaldehyde gels and transferred to nitrocellulose. The filters were hybridized at high stringency to nick-translated nonrepetitive DNA fragments from subclones of the ovomucoid gene region. DNA fragments from the clones OI3.7, OI2.0, OI1.4, and OI2.3 all hybridized to a 1.7-kb RNA present in both oviduct and liver total RNA (Fig. 1). These clones span a region from about 9–23 kb upstream from the 5' end of the ovomucoid gene region. In contrast, DNA fragments containing the ovomucoid gene hybridize to a 1.1-kb mature RNA that is present only in oviduct RNA (Nordstrom et al., 1979). DNA fragments up to 9 kb upstream of the 5' end or downstream of the ovomucoid gene did not hybridize specifically to oviduct or liver RNAs.

The ovomucoid-linked gene was subsequently shown by sequence analysis of homologous DNA fragments to code for the related Kazal inhibitor, ovoinhibitor. The homologous ovoinhibitor DNA fragments that were to be sequenced were identified by DNA blot hybridization analysis. These hybridization experiments were performed to determine if the ovo- mucoid-linked (ovoinhibitor) gene was related to the ovomucoid gene. For this purpose plasmid DNAs from the 47-kb cloned ovomucoid gene region were digested with various restriction enzymes, resolved on agarose gels, and transferred to nitrocellulose filters. The filters were hybridized under moderate stringency to nick-translated DNA fragments from the ovoinhibitor gene region. A 1.4-kb XhoI-HindIII DNA fragment from the clone OM14.5 hybridized specifically to the 3.4-kb HindIII-EcoRI fragment of the clone OM14.5, which contains the exons that code for the second and third domains of ovomucoid (Lai et al., 1979a) (Fig. 2A). Similar hybridization to the OM14.5 clone was observed with probes from the clones OI2.1 and OI2.3. Additionally, a nick-translated 2.1-kb XhoI-BamHI fragment from the clone OI2.1 hybridized under moderate stringency to three DNA fragments within the ovoinhibitor gene region (Fig. 2, B and C). One of these fragments, a 0.4-kb PstI fragment from the clone OI3.0 hybridized to a 0.6-kb BgIII-PvuII fragment of the clone OI2.1 (Fig. 2D). This hybridization data suggests that the ovoinhibitor gene region is both internally repetitive and homologous to the ovomucoid gene.

* M. J. Scott, M.-J. Tsai, and B. W. O’Malley, manuscript submitted.

* M. J. Scott, unpublished observation.
Chicken Ovoinhibitor Gene Structure

To determine the basis for the internal homology of the ovoinhibitor gene region, the four related fragments were sequenced. The homologous DNA fragments were found to share similar DNA sequences of about 90–140 bp. Over this region the four common sequences were 60–70% homologous. Additionally, the sequences common to each fragment were also all about 60% homologous to the first exons that code for the α Kazal domains of ovomucoid. Moreover, these ovomucoid-related sequences all contained an open reading frame that could potentially code for a peptide homologous to the ovomucoid Kazal domains. We therefore compared these four derived peptide sequences with protein sequences of various Kazal proteinase inhibitors. The Purdue group (I. K., W. K., M. L.) have determined the primary structure of chicken ovoinhibitor by protein sequencing (Fig. 3). It should be noted however, that since the ovoinhibitor sequence was unknown, that since the ovomucoid-linked gene codes for ovoinhibitor, and we had identified by sequence homology the first exons of the first four domains of the protein. Expression of the ovomucoid gene in oviduct and liver was consistent with the presence of the protein in egg white and plasma.

Primary Structure of Ovoinhibitor Protein Derived from a cDNA Clone and as Determined by Protein Sequencing—As the coding sequences of the 0.7-kb XhoI fragment from the clone OI4.3 represent part of the first domain and thus should correspond to sequences near the 5’ end of the ovoinhibitor mRNA, this fragment was used as a probe to screen a chick oviduct cDNA library. The DNA sequence and derived protein sequence of the longest clone that was isolated, OIc is shown in Fig. 3. The amino terminus of the mature protein corresponds to position 24 of the derived preprotein sequence. The strategy used in determining the primary structure of ovoinhibitor by protein sequencing is shown in Fig. 3. The amino acid sequence of the mature protein derived from the cDNA clone is identical to that determined by protein sequencing with one exception. The sole difference is an Asn at position 12 of the fifth domain that was identified as an Asp by sequencing of the protein. As the Asn is adjacent to a Gly we believe the identification of the amino acid as an Asp may have been due to spontaneous deamidation of the protein some time after synthesis (Meinwald et al., 1986). The putative signal peptide sequence conforms with the consensus identified by Watson (1984). The P1 amino acids of the seven Kazal domains are Arg, Arg, Arg, Arg, Phe, Met, and Met. Thus, the first four domains could potentially inhibit trypsin, the fifth chymotrypsin, and the sixth and seventh domains could inhibit both chymotrypsin and elastase (Laskowski and Kato, 1980). Although it is not certain that all seven domains are independently active, chicken ovoinhibitor does contain at least five reactive sites: two for bovine trypsin, two for bovine chymotrypsin, and one for porcine pancreatic elastase4 (Gertler and Feinstein, 1971; Gertler and Ben-Valid, 1980). Moreover, while none of the trypsin binding sites are sensitive to Met oxidation, half of the chymotrypsin sites and all of the elastase sites are sensitive to this modification (Schechter et al., 1977).

Domain Structure of the Ovoinhibitor Gene—To determine the exon-intron arrangement of the ovoinhibitor gene, we sequenced DNA fragments from within the transcribed region that were identified by RNA blot hybridization analysis. Exons were identified by sequence alignment with the OIc cDNA sequence. The DNA sequence of all exons and exon-intron boundaries was determined. The gene exon and OIc cDNA sequences were in complete agreement with the exception of one nucleotide difference in each of the coding and 3’ untranslated sequences. Both nucleotide substitutions were silent changes, that is they did not alter the derived protein sequence shown in Fig. 3. These nucleotide differences probably reflect polymorphisms within the chicken population (Lai et al., 1979b). For those exon-intron boundaries that were

4I. Kato, W. Kohr, and M. Laskowski, Jr., unpublished observation.
Chicken Ovoinhibitor Gene Structure

**FIG. 3.** Complete sequence of an ovoinhibitor cDNA clone and strategy for protein sequencing. The sequence of the mRNA sense strand of the cDNA clone, O1c is shown. The boundaries of the Kazal domains (I−VI) are indicated by vertical bars. The seven P1 amino acids and their respective codons are boxed. The Asn in the fifth domain that was determined to be an Asp by protein sequencing as well as the putative polyadenylation signal AATAAA are underlined. The horizontal lines immediately below the amino acid sequences indicate those portions of the peptides described in the text for which actual protein-sequencing data were obtained. Short vertical lines mark NH2-terminals and in those cases when the entire peptide was sequenced, the COOH-terminals. The letter code indicates the source of the peptides and cleavage method used to produce the peptides as follows: 1) CNBr cleavage of intact ovoinhibitor produced two fragments, the intact first domain (1−68) designated as CN-3 and domains 11−VI (69−499). The second fragment was nicked internally at Met residues 84, 252, 411, and 446, but it was held together by disulfide bridges. Performic acid oxidation yielded five peptides CN-0-1 (69−84), CN-0-2 (85−97), CN-0-3 (98−108), CN-0-4 (109−163), and CN-0-5 (164−289). The complete sequence of the ovoinhibitor cDNA clone, O1c is shown. The boundaries of the Kazal domains (I−VI) are indicated by vertical bars. The seven P1 amino acids and their respective codons are boxed. The Asn in the fifth domain that was determined to be an Asp by protein sequencing as well as the putative polyadenylation signal AATAAA are underlined. The horizontal lines immediately below the amino acid sequences indicate those portions of the peptides described in the text for which actual protein-sequencing data were obtained. Short vertical lines mark NH2-terminals and in those cases when the entire peptide was sequenced, the COOH-terminals. The letter code indicates the source of the peptides and cleavage method used to produce the peptides as follows: 1) CNBr cleavage of intact ovoinhibitor produced two fragments, the intact first domain (1−68) designated as CN-3 and domains 11−VI (69−499). The second fragment was nicked internally at Met residues 84, 252, 411, and 446, but it was held together by disulfide bridges. Performic acid oxidation yielded five peptides CN-0-1 (69−84), CN-0-2 (85−97), CN-0-3 (98−108), CN-0-4 (109−163), and CN-0-5 (164−289).
ambiguous due to sequence redundancy at adjacent junctions, the putative splice site was assigned according to the GT/AG rule (Breathnach and Chambon, 1980).

The ovoinhibitor gene is about 10.3-kb long and consists of 16 exons separated by 15 introns (Fig. 4). The 3' end of the ovoinhibitor gene is about 9.5 kb upstream from the 5' end of the ovomucoid gene. The intron-exon boundary sequences are shown in Table I. All intron sequences begin with a GT, end with an AG, and in general agree with the splice junction consensus sequences determined by Mount (1983). The 5' end of the ovoinhibitor gene was determined by a primer extension assay, using an oligonucleotide complementary to the 5' end of the OIc cDNA clone as a primer, reverse transcriptase, and oviduct poly(A)* RNA. The transcription initiation site occurs 15 bp upstream of the 5' end of the cDNA sequence.

The first two exons of the ovoinhibitor gene code for the signal peptide and the first two amino acids of the mature protein. Each of the seven Kazal domains is encoded by two exons. The sites at which the introns occur in the coding sequences of the domains is shown in Fig. 5. For each domain one intron separates the codon for the first amino acid of the domain from the preceding codon. In addition, one intron occurs at a conserved position within the coding sequences of all the domains. This intradomain intron separates the second and third nucleotides of the codon of the amino acid indicated in Fig. 5. Therefore, the coding sequences of all the a domains are bordered by introns. The coding sequences of the second exon of the seventh domain, which is a b-type, are contiguous with the 3'-noncoding sequences.

Comparative Structure of the Ovoinhibitor and Ovomucoid Genes—Introns separate the coding sequences of the ovomucoid protein domains at sites identical to the corresponding coding sequences in the ovoinhibitor gene (Fig. 6). The structures of the two genes are very similar. The small 20-bp second exon of both the ovoinhibitor and ovomucoid genes codes for the carboxyl-terminal portion of the signal peptide and the first two amino acids of the mature protein. The first exon of the ovoinhibitor second and fifth domains is three bp shorter than the first exon of the other a domains in ovoinhibitor and ovomucoid. On alignment of the a domain sequences, it appears that these deletions are in the region between the codons for the first and second cysteines. Similarly, fewer nucleotides in the region between the codons for Cys I and Cys II of the first exons of the b domains accounts for the smaller size of these exons relative to the first exons of the a domains of the two genes.

We have previously demonstrated that there are two 5' splice sites that are used in the splicing of the second exon of the second chicken ovomucoid domain to the first exon of the third domain (Stein et al., 1980). The most frequently used 5' splice site defines a 67-bp exon. The other 5' splice site, which defines a 61-bp exon, occurs in an identical position at the end of all six ovoinhibitor a domain second exons. The minor splice site thus appears to correspond to the ancestral site whereas the more frequently used splice site appears to have arisen by mutation within the intervening sequence. In support of this conclusion, the dipeptide Val44-Ser350 encoded by the additional 6 bp at the 5' end of the 67-bp ovomucoid sixth exon is present only in ovomucoids of most (not all) birds belonging to the order Galliformes (Laskowski et al., 1987). That is, it appears that the second exon of the ovomucoid second domain was expanded to 67 bp only after the divergence of Galliformes from other birds. The shorter length of the second exon of the ovomucoid first domain (58 bp) relative to the putative 61-bp ancestral exon probably reflects a recent deletion in the exon or a shifting of the 5' splice site.

The sizes and sequences of the corresponding introns in the ovomucoid and ovoinhibitor genes do not appear to be conserved. An exception appears to be the intradomain intron of the b domains which are of similar size. We have, however, sequenced about one-third of the ovomucoid intron, and over this region the ovomucoid and ovoinhibitor introns do not share significant sequence homology. The introns that flank the 20-bp second exon are relatively large in both genes. The interdomain introns are, on the average, about twice the size of the intradomain introns. There is apparently no significant sequence homology between the ovoinhibitor intradomain or interdomain introns. We have, however, only partially sequenced many of these introns. The 5'-flanking sequences of the two genes have little homology, with the exception of the TATA box and a nonanucleotide TGACAGATT, which are

\[ \text{XH, R, K, KpnI, P, PstI, X, XhoI, } \]

FIG. 4. Structural organization of the ovoinhibitor gene. The exons are indicated by boxes: filled boxes, coding sequences; open boxes, noncoding sequences. The sizes of the exons in nucleotides is indicated below the boxes. Exons are shown as lines. The sizes of the introns in nucleotides are shown above the lines. Exon boundaries were determined by sequence analysis. The orientation of the gene is from left to right (5' to 3') and is the same as the ovomucoid gene. For comparison the restriction sites illustrated in Fig. 1B and 2E that occur within or near the gene are shown. B, BamHI; G, BglII; H, HindIII; K, KpnI; P, PstI; Ps, PstI; X, XhoI.
TABLE I

| Exon | Intron | Exon |
|------|--------|------|
| OI 1* | CTTTGCAG | GTGAGTTTC...ACTTTTCCTTTAG | ATATTGCC |
| OI 2 | GAATTGAG | GTGAGTTACAG...GCTGCTTCCCATAG | GGTATTTG |
| OI 3 | TACACAG | GTGAGCTTCTTAG...GCTGCTCTTCCAG | AGAACAG |
| OI 4 | AGTTTAG | GTAAGCCCAAG...GCTGCTTCCCATAG | GTTTAGCT |
| OI 5 | TACACATC | GTAAGCTCTTAG...GCTGCTTCCCATAG | AGAACAG |
| OI 6 | TATCTTGA | GTAAGCTCTTAG...GCTGCTTCCCATAG | AGAACAG |
| OI 7 | CATATG | GTAAGCTCTTAG...GCTGCTTCCCATAG | AGAACAG |
| OI 8 | TACAAAG | GTAAGCTCTTAG...GCTGCTTCCCATAG | AGAACAG |
| OI 9 | ACGTTATG | GTGAGTTACAG...GCTGCTTCCCATAG | GTTTAGCT |
| OI 10 | CACAACAT | GTAAGCTCTTAG...GCTGCTTCCCATAG | AGAACAG |
| OI 11 | AACCATG | GTAAGCTCTTAG...GCTGCTTCCCATAG | AGAACAG |
| OI 12 | TCACGTGA | GTAAGCTCTTAG...GCTGCTTCCCATAG | AGAACAG |
| OI 13 | TACACAC | GTAAGCTCTTAG...GCTGCTTCCCATAG | GAACTT |
| OI 14 | TACAAATG | GTAAGCTCTTAG...GCTGCTTCCCATAG | GAACTT |
| OI 15 | CATATG | GTAAGCTCTTAG...GCTGCTTCCCATAG | GAACTT |

* Introns are numbered consecutively from the 5' end of the ovoinhibitor gene.

Fig. 5. Comparison of intron positions in the ovoinhibitor domains. The structure of the seven Kazal domains of ovoinhibitor is drawn to emphasize the similarities. The disulfide bonds are indicated by solid lines. The positions of the introns which interrupt the coding sequences of the domains are shown with arrows. The P1 amino acids of the domains are indicated by asterisks.

about 30 and 80 bp, respectively, upstream from the predominant transcription initiation sites of both genes. Interestingly, the upstream transcription initiation site of the ovomucoid gene (Lai et al., 1982; Gerlinger et al., 1982) maps to within the conserved nonanucleotide sequence motif.

Evolution of the Multidomain Ovoinhibitor Gene—The a domains of ovoinhibitor (I–VI) are more closely related to each other than to the b-type domain VII (Fig. 7, left panel). The third domain is unusual in that all of the other a domains are more related to the third domain than to any other a domain. This homology suggests that domain III may evolve at a slower rate relative to the other a domains. The a domains...
Chicken Ovoinhibitor Gene Structure

The segments of the ovoinhibitor (left panel) and ovomucoid (right panel) genes that code for the signal peptide plus the first two amino acids of the mature protein, a domains and b domain are indicated in lines 1-4, respectively. Exons are shown as boxes with the length in nucleotides within or above the box. Introns are drawn as lines with the length indicated above or below the line. For the a domain introns of ovoinhibitor, the intron sizes are shown in the same order as the domains are listed on the left.

Amino acid, 55% nucleotide homology) than to the putative peptide sequences of hPSTI (46% amino acid, 43% nucleotide homology for ovoinhibitor and ovomucoid, respectively). Moreover, the Kazal domain sequences of hPSTI (positions 7-58 of mature protein) have similar homology (28-37% amino acid, 40-45% nucleotide) to each of the a- and b-type domains of ovoinhibitor and ovomucoid. These observations suggest that the ovomucoid-ovoinhibitor and hPSTI ancestral genes diverged prior to the duplication that generated the primordial a and b domains of multidomain Kazal inhibitors (Fig. 8).

To estimate the branching order of the ovoinhibitor, ovomucoid, and hPSTI domains, we have used the modified UPG method of Li (1981) which corrects for unequal rates of evolution. The most likely pathway derived by comparing the protein sequences of the 11 domains is shown in Fig. 8. We have also used a protein sequence parsimony method that is similar to Fitch (1971) except that silent changes are not counted. According to this procedure the scheme shown requires the fewest changes. The same phylogenetic tree was also obtained by W. M. Fitch and M. L. by the use of the Fitch (1971) maximal parsimony algorithm on the (then incomplete) protein sequences. As expected from the above sequence comparisons, the a-b two-domain ancestor of ovomucoid and ovoinhibitor and the hPSTI ancestor each descend from a common single-domain primordial inhibitor. The branching order of the first four domains of ovoinhibitor should be regarded as somewhat uncertain because of the similarity of these sequences and also as domain III appears to have a relatively slower rate of evolution. However, it is interesting that the branching order of ovoinhibitor domains appears to be unidirectional in that the newly duplicated domain is always closer to the amino terminus of the protein.

Discussion

Since the formation of the seven-Kazal domain ovoinhibitor gene has involved the repeated duplication of domain-coding sequences, this suggests that there was a selection pressure for domain duplication. One physiological advantage of a large plasma inhibitor, and thus a possible selection pressure for domain duplication, is that it prevents excretion (Laskowski and Kato, 1980). Another possible advantage is that with several independent reactive sites of potentially different specificity, a single polypeptide may be able to inhibit a broad spectrum of serine proteinases. In regard to the latter, the P1 amino acids of the Japanese quail ovoinhibitor domains (Laskowski et al., 1978) are identical to those of the corresponding chicken ovoinhibitor domains except for the fifth domain. The P1 amino acid of this domain is a Tyr in Japanese quail
but a Phe in chicken, a change unlikely to affect inhibitor specificity. The Japanese quail and chicken ovoinhibitor domains are therefore apparently functionally equivalent. Thus, the ability to inhibit a variety of serine proteinases may be physiologically advantageous. However, while most avian ovoinhibitors can inhibit both trypsin and chymotrypsin, the penguin ovoinhibitor can only inhibit trypsin and the ostrich homolog cannot inhibit either proteinase (Liu et al., 1971). Clearly, the reactive sites of many more avian ovoinhibitors need to be sequenced to determine if whether in general, ovoinhibitors can potentially inhibit a broad spectrum of serine proteinases.

We have previously suggested that the primordial Kazal inhibitor gene may have consisted only of the first exon of the inhibitory domain (Stein et al., 1980). This exon encodes the reactive site and five of the six conserved cysteines. Addition of the second exon, which encodes the sixth cysteine, probably created a more efficient inhibitor as breakage of disulfide bonds tends to decrease inhibitor activity. That the intradomain intron occurs in an identical location within the coding sequences of the seven ovoinhibitor domains and the three ovomucoid domains supports the hypothesis that the intron was created by the addition of the second exon of the domain.

Gilbert (1978) has proposed that exons code for polypeptides that have a discrete structure or function. As the tertiary structures of the Kazal inhibitors, bovine pancreatic secretory trypsin inhibitor (Bolognesi et al., 1982) and the third domains of Japanese quail (Papamokos et al., 1982), turkey (Read et al., 1983), and of silver pheasant (Bode et al., 1985) ovomucoids are known, it is possible to examine this hypothesis for the Kazal domains. The tertiary structures of the four Kazal inhibitors are similar. All contain similar elements of second-
correspond to exon-intron junctions whereas the b domain coding sequences are contiguous with the 3'-non-coding sequences. That is, expression of a duplicated b domain would require mutation of the translation stop codon that immediately follows the codon for the important sixth cysteine, as well as creation of a splice site. As domain boundaries correspond to exon-intron junctions, recombination within interdomain introns could have resulted in domain duplication. Although the interdomain introns of the ovoinhibitor gene do not appear to be homologous, duplication of domain exons should also duplicate intron sequences, and it is therefore likely that at one time some or all of the interdomain introns were homologous. Thus, in the gene for the chicken Kazal inhibitor ovoinhibitor, as in the related ovomucoid gene, introns separate functional protein domains, an arrangement which may have facilitated domain duplication.

Acknowledgments—We are grateful to Dr. W.-H. Li for advice on the use of his modified UPG method and to Professor W. F. Fitch for providing a copy of his maximum parsimony program, and Dr. Christina Chang for help with the maximum parsimony analysis. We also thank Dr. D. Sargent for providing chicken tissue total RNAs, W. Beattie and co-workers for sequencing the ovoinhibitor cDNA and genomic clones, C. Wilkinson for preparation of plasmid DNA, and C. McCarthy for careful preparation of this manuscript.

REFERENCES

Bode, W., Epp, E., Huber, R., Laskowski, M., Jr., and Ardelt, W. (1985) Eur. J. Biochem. 147, 387-395
Bolognesi, M., Gatti, G., Menegatti, E., Guanieri, M., Marquant, M., Papamokos, E., and Huber, R. (1986) J. Mol. Biol. 162, 829-866
Breathnach, R., and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349-383
Catterall, J. F., Stein, J. P., Lai, E. C., Woo, S. L. C., Dugaiczyk, A., Mace, M. L., Means, A. R., and O'Malley, B. W. (1979) Nature 281, 323-327
Catterall, J. F., Stein, J. P., Kristo, P., Means, A. R., and O'Malley, B. W. (1980) J. Cell Biol. 87, 480-487
Colbert, D. A., Knoll, B. J., Woo, S. L. C., Mace, M. L., Tsai, M.-J., and O'Malley, B. W. (1980) Biochemistry 19, 5586-5592
Dretzen, G., Bellard, M., Sassone-Corsi, P., and Chambon, P. (1981) Annu. Rev. Biochem. 112, 295-298
Efstratiadis, A., Possikony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., De Riel, J. K., Forget, B. G., Weissman, S. W., Slightom, J. L., Blech, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C., and Proudfoot, N. J. (1980) Cell 21, 653-668
Fitch, W. M. (1971) Syst. Zool. 20, 406-416
Gerlinger, P., Krust, A., LeMuer, M., Perrin, F., Cochet, M., Gannon, F., Dubret, D., and Chambon, P. (1982) J. Mol. Biol. 162, 345-364
Gertler, A., and Feinstein, G. (1971) Eur. J. Biochem. 20, 547-552
Gertler, A., and Ben-Valid, A. (1980) Eur. J. Biochem. 110, 571-577
Gilbert, W. (1978) Nature 271, 501
Kato, I., Schrde, J., Khor, W. J., and Laskowski, M., Jr. (1987) Biochemistry 26, 193-201
Kulomaa, M. S., Weigel, N. L., Kleinsin, D. A., Beattie, W., Connelly, O. M., March, C., Zarucki-Schulz, T., Schrader, W. T., and O'Malley, B. W. (1986) Biochemistry 25, 6244-6251
Lai, E. C., Stein, J. P., Catterall, J. F., Woo, S. L. C., Mace, M. L., Means, A. R., and O'Malley, B. W. (1979) Cell 18, 829-842
Lai, E. C., Woo, S. L. C., Dugaiczyk, A., and O'Malley, B. W. (1979) Cell 16, 201-211
Lai, E. C., Roop, D. R., Tsai, M.-J., Woo, S. L. C., and O'Malley, B. W. (1982) Nucleic Acids Res. 10, 5553-5567
Laskowski, M., Jr., Kato, I., and Khor, W. J. (1978) in Proceedings International Symposium Proteins, Versatility of Proteins (Li, C. H., ed) pp. 307-318, Academic, New York
Laskowski, M., Jr., and Kato, I. (1980) Annu. Rev. Biochem. 49, 593-626
Laskowski, M., Jr., Kato, I., Khor, W. J., March, C. J., and Bogard, W. C. (1980) in Proteins of Biological Fluids (Peeters, H., ed) Vol. 28, pp. 123-128, Pergamon, Oxford
Laskowski, M., Jr., Kato, I., Ardelt, W., Cook, J., Denton, A., Empie, M. W., Khor, W. J., Park, S. J., Parks, K., Schatzley, B. L., Schoenberger, O. L., Tashiro, M., Vichot, G., Whatley, H. E., Wieczorek, A., and Wieczorek, M. (1987) Biochemistry 26, 202-222
Levine, M., Rubin, G. M., and Tjian, R. (1984) Cell 38, 667-673
Li, W.-H. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 1085-1089
Li, W.-H., Means, G. E., and Feeney, R. E. (1971) Biochim. Biophys. Acta 229, 176-185
Longberg, N., and Gilbert, W. (1985) Cell 40, 81-90
March, C. J. (1980) Ph.D. thesis, Purdue University
Marchionni, M., and Gilbert, W. (1986) Cell 46, 133-141
Masatshuka, K. (1958) Science 127, 1178-1179
Mazam, A. M., and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 560-564
Meinwald, V. C., Stanion, E. R., and Scheraga, H. A. (1986) Int. J. Peptide Protein Res. 28, 79-84
Messing, J. (1983) Methods Enzymol. 101, 20-78
Mount, S. M. (1982) Nucleic Acids Res. 10, 459-472
Nordstrom, J. L., Roop, D. R., Tsai, M.-J., and O'Malley, B. W. (1979) Nature 278, 328-331
Norman, D. (1985) The Illustrated Encyclopedia of Dinosaurs, Crescent, New York
Palmier, R. D. (1972) J. Biol. Chem. 247, 6450-6461
Papamokos, E., Weber, E., Bode, W., Huber, R., Empie, M. W., Kato, I., and Laskowski, M. Jr. (1982) J. Mol. Biol. 158, 515-537
Read, R. J., Fujinaga, M., Sielecki, A. R., and James, M. N. G. (1983) Biochemistry 22, 4420-4433
Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463-5467
Sargent, R. S., Tsai, M.-J., and O'Malley, B. W. (1986) Biochemistry 25, 6252-6258
Schechter, Y., and Berger, A. (1967) Biochem. Biophys. Res. Commun. 27, 157-162
Schechter, Y., Burstein, Y., and Gertler, A. (1977) Biochemistry 16, 992-997
Southern, E. M. (1975) J. Mol. Biol. 98, 503-517
Stein, J. P., Catterall, J. F., Kristo, P., Means, A. R., and O'Malley, B. W. (1980) Cell 12, 681-687
Watson, M. E. (1984) Nucleic Acids Res. 12, 5145-5164
Yamamoto, T., Nakamura, Y., Nishide, T., Emi, M., Ogawa, M., Mori, T., and Matsubara, K. (1985) Biochem. Biophys. Res. Commun. 132, 605-612