

Amino Acid Sequences of Ovomuroid Third Domains from 27 Additional Species of Birds

Izydor Apostol,^{1,2} Anthony Giletto,¹ Tomoko Komiyama,^{1,3} WenLei Zhang,¹ and Michael Laskowski, Jr.^{1,4}

Received February 16, 1993

Ovomucoids consist of a single polypeptide chain which is composed of three tandem Kazal domains. Each Kazal domain is an actual or putative protein inhibitor of serine proteinases. Ovomuroid third domains were already isolated and sequenced from 126 species of birds (Laskowski *et al.*, 1987, 1990). This paper adds 27 new species. A number of generalizations are made on the basis of sequences from 153 species. The residues that are in contact with the enzyme in enzyme-inhibitor complexes are strikingly hypervariable. While the primary specificity residue, P_1 , is the most variable; substitutions occur predominantly among aliphatic, hydrophobic residues. Consensus sequences for an avian ovomucoid third domain, for a b-type Kazal domain (i.e., a COOH terminal domain of multidomain inhibitors) and for a general Kazal domain are given. Finally, the individual new sequences are briefly discussed.

KEY WORDS: Ovomuroid third domain sequences; protein proteinase inhibitors; evolution of birds.

1. INTRODUCTION

Ovomucoids are abundant glycoproteins in avian eggwhites. Their polypeptide chains consist of three tandem homologous domains (Fig.1) (Kato *et al.*, 1976, 1987). Each of the three domains is a member of the Kazal inhibitor family of serine proteinase inhibitors (for definition of inhibitor families, see Laskowski and Kato, 1980). While easy to isolate, entire ovomucoids are cumbersome to study. In favorable cases, they can be cleaved into their three component domains without any intradomain nicking (Kato *et al.*, 1987, also see Fig. 1). This can be achieved either by limited proteolysis or by CNBr cleavage or by a combination of both. The production of third domains by limited proteolysis proved most efficacious. We have isolated third domains

from 126 species of birds and determined their amino acid sequences (Kato *et al.*, 1987; Laskowski *et al.*, 1987, 1990). This paper adds the sequences of ovomucoid third domains from 27 more species, thus raising the number of species with known third domain sequences to 153.

Many investigators show only limited interest in large sets of sequences of homologous proteins from closely related species. This is because they expect, and generally find, that the structurally and functionally important residues in such sets are unvaried or strongly conserved. The variation is expected to occur in unimportant positions and is a reflection of genetic drift (Kimura, 1983).

Protein proteinase inhibitors of serine proteinases in general (Laskowski and Kato, 1980; Creighton and Darby 1990), and ovomucoids (Rhodes *et al.*, 1960), and ovomucoid third domains in particular (Laskowski *et al.*, 1987, 1988, 1990) are conspicuous exceptions to these expectations. In these inhibitors, the positions which are involved in enzyme-inhibitor contact (see Table IV in the Results and Discussion section) fix mutations far more rapidly than the remaining positions. This rapid

¹ Department of Chemistry, Purdue University, West Lafayette, Indiana 47907.

² Present address: Somatogen, Inc., 5797 Central Ave., Boulder, Colorado 80301.

³ Present address: Duke University Medical Center, Department of Pathology, Durham, North Carolina 27710.

⁴ To whom all correspondence should be addressed.

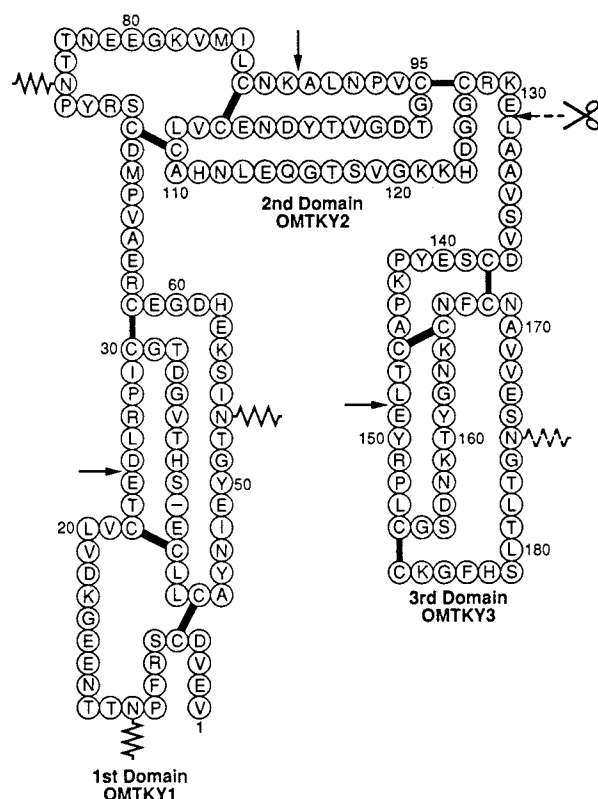


Fig. 1. Amino acid sequence of turkey (*Meleagris gallopavo*) ovomucoid. The $\ominus$ indicates a deletion in the turkey ovomucoid first domain; $\equiv$ mark the disulfide bridges; $\rightarrow$ the reactive site peptide bonds of the three domains, $\sim$ the glycosylation positions. In the third domain $\sim$ emphasizes that the glycosylation is partial. The $\sim$ indicates the position of limited proteolysis site by Staphylococcal proteinase V8 (endoproteinase Glu-C) often used in the generation of ovomucoid third domains (see Table 1).

change in the enzyme-inhibitor contact regions is accompanied by huge changes in the inhibitory specificity and in the strength of enzyme-inhibitor interaction. This unusual evolution is one important reason for study of comparative sequences of proteinase inhibitors.

Adherence to neutral theory of evolution makes homologous proteins from closely related species virtually useless for structure to function studies. On the other hand, ovomucoid third domain sequences proved quite useful. We have now measured the enzyme-inhibitor association equilibrium constants K_a for most of the 92 unique sequence ovomucoid third domains we have isolated (Empie and Laskowski, 1982; Laskowski *et al.*, 1981, 1983, 1989; Komiyama *et al.*, 1991; Bigler *et al.*, 1993; Park, 1985; Wynn, 1990) with numerous serine proteinases. For many of these systems, we also measured

the rate of constants for enzyme-inhibitor association (Forwith, 1992) and two important properties of the inhibitor alone: the melting temperature, T_m , (Otlewski, Qasim and Laskowski, manuscripts in preparation) and K_{hyd} , the equilibrium constant for hydrolysis of the reactive site peptide bond (Ardelt and Laskowski, 1991).

As this work progressed, it became clear that while a valuable database was generated, further interpretation could be aided far more efficiently by designed variants obtained by semisynthesis (Sealock and Laskowski, 1969; Wiczorek and Laskowski, 1982; Wiczorek *et al.*, 1987; Bigler, 1991; Bigler *et al.*, 1993), site-specific mutagenesis (Lu *et al.*, 1992), or by total synthesis (Warne, 1990; Green and Laskowski, in preparation). Therefore, this paper is probably the last third domain sequencing effort from our laboratory for some time. As the sequence to reactivity algorithms becomes better developed, we plan to return to natural sequences with the hope of understanding the reasons for the various specific changes that occurred in the evolution of closely related ovomucoids.

In reading this paper, the reader will be greatly aided by having Laskowski *et al.* (1987, 1990) at their side, as this paper is a simple extension of the older ones.

2. MATERIAL AND METHODS

2.1. Nomenclature and Acquisition of Eggwhites

Table 1 (analogous to Table 1 of Laskowski *et al.*, 1987, 1990) contains the three-letter abbreviations we have coined for the species of birds, their English names and their Latin names. The names and their order (roughly corresponding to the classical phylogenetic classification) are from Gruson (1976), which is generally in good agreement with the latest Sibley and Monroe (1990) and Sibley and Ahlquist (1990) bird classification. The next entry lists the suppliers to whom we are indebted for the eggs or eggwhite. In the case of WPA, where we obtained material from two suppliers, both are reported even though the material from only one supplier has been used since frequently it is difficult to obtain additional material from the original supplier.

The next three columns list the yield obtained from the isolation of entire ovomucoid, the enzyme used for partial digestion to cut the third domain from entire ovomucoid, and the third domain yield. These figures, especially the last, should not be taken

very seriously as they are mostly the results of unreplicated experiments. There is also considerable variation in the quality and quantity of the starting sample which effects yields.

The last column indicates by (+) cases where we isolated a glycosylated (on Asn⁴⁵) third domain and by (−) cases where nonglycosylated third domain could be obtained. In one case, OMBTP, both glycosylated and nonglycosylated forms of third domain were obtained. In all cases, we preferred to work with nonglycosylated domains as they are easier to characterize.

2.2. Preparation of Avian Ovomucoids

This procedure was slightly modified from the one given in Kato *et al.* (1987) and described later by Laskowski *et al.* (1990). Egg-whites were placed in a Waring blender. Two volumes of freshly prepared, cold precipitating solution (1 vol of acetone and 2 vol of 0.13 M trichloroacetic acid) were added, and the mixture was homogenized for 3–5 min. The resulting suspension was gently mixed overnight at 4°C in a rotary shaker and then centrifuged at 3000 g for 25 min. Ovomucoid, which is in the supernatant, was precipitated by adding 2–2.5 vol of cold acetone. The suspension was kept at 4°C overnight. The precipitate was recovered by centrifugation and decantation and then dissolved in deionized water and freeze-dried.

2.3. Preparation of Ovomucoid Third Domains

This was carried out as described in detail in the preceding paper by Laskowski *et al.* (1987). Third domains were isolated from the digest by size-exclusion chromatography (Bio-gel P-10) and were further purified by ion-exchange chromatography. Nonglycosylated domains were purified on Q-Sepharose in 50 mM Tris buffer pH 9.0 and eluted by 0–0.2 M NaCl gradient. Glycosylated domains were purified on S-Sepharose in 50 mM acetate buffer pH 3.6 and eluted by 0–0.5 M NaCl gradient.

2.4. Amino Acid Analysis

Twenty to fifty µg of newly prepared third domains were hydrolyzed in 6 N HCl at 110°C for 20 hr. Most recently, the temperature was raised to 165°C and the time shortened to 2 hr in the Savant Automatic Hydrolyzer AP 100. Dithiodipropionic acid was added to all hydrolyses in order to improve

cystine quantitation (Barkhold and Jensen, 1989). The hydrolysates were analyzed on a Beckman 7300 amino acid analyzer. Color factor for Cys-X (mixed disulfide of cysteine and thiopropionic acid) was determined as 0.950 of the color factor of cystine ($CF_{Cys-X} = 0.950 \times CF_{Cys/2}$). The results were fitted to the number of residues per molecule by an integer fit program utilizing Hoy *et al.* (1974) and Scott *et al.* (1975) fitting functions.

2.5. Sequencing

Most sequencing was carried out on the modified Beckman 890C as described previously (Laskowski *et al.*, 1990). However, as gas-phase sequencers require far less material and chemicals and are much less cumbersome to use, we also used a Porton 2020 gas-phase sequencer equipped with a homemade automatic sample injector and Waters isocratic HPLC system operating with a DuPont Zorbax PTH column. The system was controlled by Spectra Physics ChromJet Integrator which was also used to log and quantitate the data acquired by Bioanalytical Systems BAS UV-108 programmable detector which was operated at 269 nm. The Porton was easier, faster, and cheaper to use, but a complete third domain sequence (typically 56 residues) could not be reliably completed. Reliable sequences tended to terminate at 46–50 residues. However, most ovomucoid third domains (see Tables II and IV below) have either Asp⁴³ or Glu⁴³. After the long run which determined which residue was present at position 43, the reduced, carboxymethylated ovomucoid third domain was extensively digested either with *Staphylococcal* proteinase V8 (Endoproteinase Glu-C) or with Endoproteinase Asp-N as appropriate. The peptides were separated on a Poros R/H reversed-phase column and the COOH terminal peptide 44–56 or 43–56, respectively, was isolated and sequenced.

3. RESULTS AND DISCUSSION

Table II is a list of the 28 sequences of ovomucoid third domains determined in this work. The reason for 28 sequences from 27 species is that in Chilean tinamou, NTI, the third domain is dimorphic with both Ala³² and Pro³². In the complete data set (a composite of Kato *et al.*, 1987; Laskowski *et al.*, 1987, 1990, and this paper) there are sequences from 153 species, six of which are dimorphic. Thus, 159 ovomucoid third domain sequences have now been reported.

Table 1. Names of Birds Whose Ovomuroid Third Domains Were Sequenced: Sources of Egg Supply and Isolation Data

Abbrev.	Common name	Latin name	Supplier code ^a	Entire ovomuroid yield (mg/ml of egg white)	Proteinase used	Third domain yield (mg/ml of egg white)	Glycosylation
DWC	Dwarf cassowary	<i>Casuarius bennetti</i>	s1 (G)	9.4	Spase V8	0.8	(+)
GTI	Great tinamou	<i>Tinamus major</i>	s1 (G)	13.9	Thermolysin	0.1	(+)
RTI	Red-winged tinamou	<i>Rhynchotus rufescens</i>	s1 (G)	16.9	Thermolysin	0.3	(+)
NTI	Chilian tinamou	<i>Nothoprocta perdicaria</i>	s1 (G)	13.8	Thermolysin	0.4	(+)
SOS	Southern screamer	<i>Chauna torquata</i>	s2 (G)	7.3	Thermolysin	0.5	(+)
EMG	Emperor goose	<i>Anser canagicus</i>	s3 (G)	1.0	Thermolysin	<0.1	(-)
EGG	Egyptian goose	<i>Alopochen aegyptiaca</i>	s4 (G)	7.5	Spase V8	0.4	(-)
RED	Redhead	<i>Aythya americana</i>	s5 (G)	9.7	Thermolysin	0.3	(-)
RPO	Red-crested pochard	<i>Netta rufina</i>	s5 (G)	16.6	Thermolysin	0.5	(-)
HDV	Hooded vulture	<i>Neophron monachus</i>	s6 (G)	9.3	Spase V8	0.1	(-)
WBT	Wattled brush turkey	<i>Aepyptodius arfakianus</i>	s1 (G)	8.8	Spase V8	0.1	(+)
TTR	White-tailed ptarmigan	<i>Lagopus leucurus</i>	s7 (G)	9.9	Spase V8	0.3	(-)
WPA	Crested wood-partridge	<i>Rollulus roulroul</i>	s8 (G), s9 (G)	3.4	Thermolysin	<0.1	(-)
BTP	Blyth's tragopan	<i>Tragopan blythii</i>	s10 (G)	8.1	Spase V8	0.2	(+)
CTP	Cabot's tragopan	<i>Tragopan caboti</i>	s11 (G)	6.8	Spase V8	0.4	(-)
WEP	White-eared pheasant	<i>Crossoptilon crossoptilon</i>	s10 (G)	11.4	Spase V8	0.3	(-)
SFP	Siamese fireback Pheasant	<i>Lophura diardi</i>	s8 (P)	9.3	Spase V8	0.4	(-)
PPP	Palawan peacock-Pheasant	<i>Polyplectron emphanum</i>	s10 (G)	3.5	Pepsin	<0.1	(+)
HTZ	Hoatzin	<i>Opisthocomus hoazin</i>	s1 (G)	4.3	Pepsin	<0.1	(+)
JPC	Japanese crane	<i>Grus japonensis</i>	s11 (G)	8.5	Spase V8	0.2	(+)
COC	Crowned crane	<i>Balearcica pavonina</i>	s12 (G)	6.7	Spase V8	0.2	(+)
CBC	Channel-billed cuckoo	<i>Scythrops novaehollandiae</i>	s1 (G)	10.0	Pepsin	0.3	(+)
GGC	Guira cuckoo	<i>Guira guira</i>	s1 (G)	6.4	Thermolysin	0.1	(+)
BLB	Blackbird	<i>Turdus merula</i>	s1 (G)	8.3	Spase V8	0.4	(+)
HFN	House finch	<i>Carpodacus mexicanus</i>	s1 (G)	4.2	Trypsin	0.1	(+)
CPC	Pied crow	<i>Corvus albus</i>	s1 (G)	5.5	Thermolysin	0.1	(+)
CEJ	Jackdaw	<i>Corvus monedula</i>	s1 (G)	7.7	Pepsin	<0.1	(+)

^aList of suppliers: (G) indicates a gift; (P) indicates a purchase.

s1—Sibley Collection located at Purdue University, Department of Chemistry, West Lafayette, IN 47907, donated by Dr. Charles G. Sibley.

s2—National Zoological Park, Smithsonian Institution, 3100 Connecticut Ave. N.W., Washington, D.C. 20008; Paul Tommasoni, Curator.

s3—Lincoln Park Zoological Gardens, 2200 North Cannon Drive, Chicago, IL 60614; Kevin J. Bell, Curator of Birds.

s4—Woodland Park Zoological Gardens, 5500 Phinney Avenue North, Seattle, WA 98103; Gregory S. Toffie, Curator of Birds.

s5—St. Louis Zoological Park, Forest Park, St. Louis, MO 63110; Bruce W. Bohmke, Curator of Birds.

s6—Bronx Zoo, New York Zoological Park, 185th and Southern Blvd., Bronx, N.Y.; Dan Wharton, Curator.

s7—Dr. Clait Braun, Colorado Wildlife Research Center, 317 West Prospect, Ft. Collins, CO 80526.

s8—Richard & Joan Durrant, 1140 West State Road 4, Cathlamet, WA 98612.

s9—Denver Zoological Foundation, City Park, East 23rd Avenue and Steele, Denver, CO 80205; Susie Haefner, Curator of Birds.

s10—Harry Hardy, South View Aviaries, 4566 Portland St, Burnaby, British Columbia, Canada V5J2N9.

s11—Internal Crane Foundation, Rt. 1, Box 230 C, Shady Lane Road, Baraboo, WI 53913.

s12—San Antonio Zoo, 3903 North St. Mary's Street, San Antonio, TX 78212; Mary Healy, Curator of Birds.

Table II. Amino Acid Sequences of 28 Ovomuroid Third Domains (27 Species Plus One Polymorphism)^{a,b}

Table II. Amino Acid Sequences of 28 Ovicornuinae (27 Species)																																																										
	1	2	3	4	5	5 _A	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6																					
OMDWC3	F	A	T	-	-	-	V	D	C	S	E	Y	P	K	P	V	C	S	P	E	Y	M	P	L	C	G	S	D	S	K	T	Y	N	N	K	C	D	F	C	S	A	V	V	E	S	N	G	T	L	T	L	G	H	F	G	K	C	
OMGTI3	.	.	.	V	T	-	V	D	C	S	G	Y	P	K	P	A	C	T	A	E	Y	F	F	P	L	C	G	S	D	N	Q	T	Y	S	N	K	C	A	F	C	N	A	A	V	E	K	N	V	-	-	T	L	G	H	F	G	E	C
OMRTI3	.	.	.	V	T	-	V	D	C	S	G	Y	P	K	P	D	C	T	L	E	S	F	F	P	L	C	G	S	D	N	Q	T	Y	S	N	K	C	A	F	C	N	A	A	V	E	R	N	V	-	-	T	L	R	H	L	G	E	C
OMNTI3A	.	.	.	V	T	-	V	D	C	S	G	Y	P	K	P	A	C	T	L	E	Y	F	F	P	L	C	G	S	D	N	Q	T	Y	A	N	K	C	T	F	C	N	A	V	E	K	N	V	-	-	T	L	N	H	L	G	E	C	
OMNTI3P	.	.	.	V	T	-	V	D	C	S	G	Y	P	K	P	A	C	T	L	E	Y	F	F	P	L	C	G	S	D	N	Q	T	Y	P	N	K	C	T	F	C	N	A	V	E	K	N	V	-	-	T	L	N	H	L	G	E	C	
OMSOS3	V	A	T	-	-	-	V	D	C	S	D	Y	P	K	P	A	C	T	L	E	Y	M	P	L	C	G	S	D	N	Q	T	Y	S	N	K	C	S	F	C	N	A	V	V	D	S	N	G	T	L	T	L	S	H	F	G	K	C	
OMEMG3	V	A	T	-	-	-	V	D	C	S	D	Y	P	K	P	A	C	T	V	E	Y	M	P	L	C	G	S	D	N	K	T	Y	G	N	K	C	N	F	C	N	A	V	V	D	S	N	G	T	L	T	L	S	H	F	G	K	C	
OMEGG3	V	A	T	-	-	-	V	D	C	S	G	Y	P	K	P	A	C	T	M	E	Y	M	P	L	C	G	S	D	N	K	T	Y	G	N	K	C	N	F	C	N	A	V	V	D	S	N	G	T	L	T	L	S	H	F	G	E	C	
OMRED3	V	A	T	-	-	-	V	N	C	S	G	Y	P	K	P	A	C	T	M	E	Y	M	P	L	C	G	S	D	N	B	K	T	Y	G	N	K	C	N	F	C	N	A	V	V	D	S	N	G	T	L	T	L	S	H	F	G	E	C
OMRPO3	V	A	T	-	-	-	V	N	C	S	G	Y	P	K	P	A	C	T	M	E	Y	M	P	L	C	G	S	D	N	K	T	Y	G	N	K	C	N	F	C	N	A	V	V	D	S	N	G	T	L	T	L	S	H	F	G	E	C	
OMHVD3	I	A	I	-	-	-	V	D	C	S	D	Y	P	K	P	V	C	S	L	E	Y	M	P	L	C	G	S	D	S	K	T	Y	S	N	K	C	D	F	C	N	A	V	V	D	S	N	G	T	L	T	L	S	H	F	G	K	C	
OMWBT3	V	V	I	-	-	-	V	D	C	S	D	Y	P	T	H	G	C	T	L	E	K	P	I	C	G	S	D	N	Q	T	Y	S	N	K	C	C	G	F	C	N	A	V	A	Q	S	N	G	T	L	T	L	S	H	F	G	K	C	
OMTTR3	L	A	A	V	S	-	V	D	C	S	E	Y	P	K	P	A	C	T	M	E	Y	R	P	L	C	G	S	D	N	K	T	Y	S	N	K	C	N	F	C	N	A	V	V	E	S	N	G	T	L	T	L	S	H	F	G	K	C	
OMWPA3	L	A	A	V	N	-	V	D	C	S	G	Y	P	K	P	A	C	T	M	E	Y	R	P	L	C	G	S	D	N	K	T	Y	S	N	K	C	N	F	C	N	A	V	V	E	S	N	G	T	L	T	L	S	Y	P	G	V	C	
OMBTP3	L	A	A	V	S	-	V	D	C	S	E	Y	P	K	P	A	C	T	M	E	Y	R	P	L	C	G	S	D	N	K	T	Y	G	N	K	C	N	F	C	N	A	V	V	E	S	N	G	T	L	T	L	S	H	F	G	K	C	
OMCTP3	L	A	A	V	S	-	V	D	C	S	E	Y	P	K	P	A	C	T	M	E	Y	R	P	L	C	G	S	D	N	K	T	Y	G	N	K	C	N	F	C	N	A	V	V	E	S	N	G	T	L	T	L	S	H	F	G	K	C	
OMWER3	L	A	A	V	S	-	V	D	C	S	E	Y	P	K	P	A	C	T	L	E	Y	R	P	L	C	G	S	D	N	K	T	Y	G	N	K	C	N	F	C	N	A	V	V	E	S	N	G	T	L	T	L	S	H	F	G	K	C	
OMSFP3	L	A	A	V	S	-	V	D	C	S	E	Y	P	K	P	A	C	T	M	E	Y	R	P	L	C	G	S	D	N	K	T	Y	S	N	K	C	N	F	C	N	A	V	V	E	S	N	G	T	L	T	L	S	H	F	G	K	C	
OMPPP3	.	.	.	.	.	.	V	D	C	S	E	Y	P	Q	P	A	C	T	T	E	R	R	P	V	C	G	S	D	N	K	T	Y	S	N	K	C	N	F	C	N	A	V	V	E	S	N	G	T	L	T	V	S	H	F	G	K	C	
OMHTZ3	T	A	T	-	-	-	V	D	C	S	D	Y	P	K	P	A	C	L	L	E	Y	M	P	L	C	G	S	D	N	K	T	Y	D	N	K	C	S	F	C	N	A	V	V	D	S	N	G	T	L	S	L	S	H	F	G	K	C	
OMJPC3	T	A	T	-	-	-	V	D	C	S	D	Y	P	K	P	A	C	T	L	E	Y	M	P	F	C	G	S	D	S	K	T	Y	S	N	K	C	N	F	C	N	A	V	V	D	S	N	G	T	L	T	L	S	H	F	G	K	C	
OMCOC3	T	A	M	-	-	-	V	D	C	S	D	Y	P	K	P	A	C	T	L	E	Y	M	P	F	C	G	S	D	S	K	T	Y	S	N	K	C	N	F	C	N	A	V	V	D	S	N	G	T	L	T	L	S	H	F	G	K	C	
OMCBC3	.	.	.	T	-	-	V	D	C	S	D	Y	P	K	P	V	C	T	L	E	E	M	P	L	C	G	S	D	N	K	T	Y	G	N	K	C	N	F	C	N	A	V	V	D	S	N	G	T	L	T	L	S	H	F	G	K	C	
OMGGC3	I	T	T	-	-	-	V	D	C	S	D	Y	P	K	P	V	C	S	L	E	Y	M	P	L	C	G	S	D	N	K	T	Y	S	N	K	C	N	F	C	N	A	V	A	D	S	N	G	T	L	T	L	S	H	F	G	K	C	
OMBLB3	I	V	T	-	-	-	V	D	C	S	D	Y	P	K	P	V	C	T	L	D	Y	M	P	L	C	G	S	D	N	T	T	Y	S	N	K	C	N	F	C	N	A	V	A	D	S	N	G	T	L	T	L	S	H	F	G	K	C	
OMHFN3	I	V	A	-	-	-	V	D	C	S	D	Y	P	K	P	V	C	S	L	D	Y	M	P	L	C	G	S	D	N	T	T	Y	S	N	K	C	I	F	C	N	A	V	V	D	S	N	G	T	I	T	L	S	H	F	G	K	C	
OMCP3	I	V	T	-	-	-	V	D	C	S	D	Y	P	R	P	V	C	T	L	D	Y	M	P	L	C	G	S	D	N	K	T	Y	S	N	K	C	N	F	C	N	A	V	V	D	S	N	G	T	I	T	L	S	H	F	G	R	C	
OMCEI3	I	V	T	-	-	-	V	D	C	S	D	Y	P	K	P	V	C	T	L	D	Y	M	P	L	C	G	S	D	N	K	T	Y	S	N	K	C	N	F	C	N	A	V	V	D	S	N	G	T	I	T	L	S	H	F	G	R	C	

^a For abbreviation of bird name, see Table I.

^b The third domains proper start at position 6 and continue to position 56; thus, except for tinamou sequences, they contain 51 residues. Positions 1-5A are formally in the second domain. They are part of interdomain connecting peptide (see Fig. 1). For a discussion, see Laskowski *et al.* (1987).

Table III. Species Having Identical Ovomucoid Third Domain Amino Acid Sequences^a

REA = LER	Ratitae
<u>DWC</u> = ASC	Ratitae
APE = KNC = ADR	Sphenisciformes/Pelecaniformes
NOS = <u>SOS</u>	Anseriformes
WTD = HTD = BTD = FTD	Anseriformes
SWN = MSN = RUD = LKD	Anseriformes
GOO = BHG	Anseriformes
<u>EMG</u> = CNG	Anseriformes
UPG = BWG = ORG = <u>EGG</u> = WSD = CRD = CAW = DUK	
= MND = WOD = KBG = KDM	Anseriformes
<u>RED</u> = <u>RPO</u>	Anseriformes
TTE ^b = <u>HDV</u> = CPV = AHH = WTE	Accipitridae
GUA = RGN = PGN ^b	Cracidae
<u>TTR</u> = GPC = SGG = <u>BTP</u> = <u>CTP</u> = STP = TTP = HMP = BEP = <u>WEP</u> = BRP + EDP	
= SVP = SWP = KOP = CHP = ELP = HUP = CPP	Phasianoidae
CAQ = GMQ	Phasianoidae
CHU = RPA	Phasianoidae
BTf = ERF ^c	Phasianoidae
<u>SFP</u> = CFP	Phasianoidae
CHI = CJF = GJF(A) ^d	Phasianoidae
RNP = VCP	Phasianoidae
LAP = GLP	Phasianoidae
IPF = GPF ^c	Phasianoidae
GUI = VGU	Phasianoidae
TKY = OTK	Phasianoidae
WTC = <u>JPC</u> = WNC = DEC = <u>COC</u> ^b	Gruiformes
GRM = COO	Gruiformes
SPL = GGL = BGL	Charadriiformes

^a Underlined species are those provided by this paper in addition to those provided by Laskowski *et al.* (1987, 1990).

^b Connecting peptide fragments are different.

^c Connecting peptides differ in length; different enzymes were used in cutting.

^d GJF ovomucoid third domain is polymorphic. One of the two forms is identical to those of CHI and CJF.

Of the 28 sequences in Table II, 12 duplicate previously published sequences. These 12 are underlined in the cumulative table of duplicate sequences (Table III). Additionally underlined in this table is a new pair of duplicate sequences, OMRED3 = OMRPO3 found in this paper. Thus, we have 15 new unique sequences and 13 duplicate ones. In the whole set, there are now 92 unique and 67 duplicate sequences (listed in Table III).

3.1. Ornithology Texts

In Tables I, II, and III, the species are listed in an "ornithological" order based on the book of Gruson (1976). We adopted this book in Laskowski *et al.* (1987) and continued its use in Laskowski *et al.* (1990) both for naming the various species and for placing them in order. At that time it was clearly the best choice. Recently, a more complete and more authoritative list of avian species was published by Sibley and Monroe (1990). Although it now would be best to order all of our sequences on the basis of that text, the current paper is

incremental not total, and we have opted for consistency with earlier work to retain Gruson (1976). Note that this does not simply affect the order, but also the definition of species. Within the genus *Phasianus*, Gruson (1976) recognizes two species *Phasianus colchicus*, ringneck pheasant, RNP, and *Phasianus versicolor*, green pheasant, VCP. We sequenced (Laskowski *et al.*, 1987) the third domains from both and found them identical (RNP = VCP in Table III); Sibley and Monroe (1990), on the other hand, recognize only one species of *Phasianus* (i.e., *Phasianus colchicus*). Following them would eliminate the RNP = VCP equality from Table III and the OMVCP3 sequence from Laskowski *et al.* (1987), thus reducing the total number of species, whose third domains were sequenced from 153 to 152.

Within the genus *Crossoptilon*, Gruson (1976) recognized only three species: *Crossoptilon crossoptilon*, white-eared pheasant, WEP; *Crossoptilon mantchuricum*, brown-eared pheasant, BRP; and *Crossoptilon auritum*, blue-eared pheasant, BEP. This paper completes the sequencing of all three and thus completes the genus. Sibley and Monroe

(1990) divide *Crossoptilon crossoptilon* into two possible species *Crossoptilon [crossoptilon] crossoptilon*, white-eared pheasant, and *Crossoptilon [crossoptilon] hormani*, Tibetan-eared pheasant. Of this pair, we have sequenced only the first so that the genus is not completed. On the whole, there are relatively few differences between Gruson (1976) and Sibley and Monroe (1990) that affect this work, but readers with ornithological interests should consult both texts.

3.2. Consensus Sequences and Variability of Sequence Positions

Table IV lists vertically the turkey ovomucoid third domain sequence (OMTKY3). Horizontally it lists the various amino acid residues found at corresponding positions in the sequences from 153 species of birds. The subscripts denote the number of species in which given residues are found. The six dimorphisms are dealt with by assigning a weight of 0.5 to each dimorphic form. Therefore, the sum of the subscripts in each horizontal line is 153.

In a previous version of this table (Laskowski *et al.*, 1990), the residue with the largest subscript in each horizontal line was the residue present in turkey ovomucoid third domain. Now this is true only for 49 of the 51 positions. At position 21 Met is more common than Arg which is present in turkey. At position 43 Asp is slightly more common than Glu which is present in turkey. None of the thus far determined natural sequences correspond to the "most probable" sequence. Since having a "wild type" is very convenient in inhibitor design and structure to function studies, we continue to use turkey ovomucoid third domain as our wild type (Komiya *et al.*, 1991; Lu *et al.*, 1992; Bigler *et al.*, 1993).

Also listed in Table IV are the residues of OMTKY3 making contact with human leukocyte elastase (Bode *et al.*, 1986; and Bode, personal communication), SGPB (Read *et al.*, 1983), and bovine chymotrypsin A α (Fujinaga *et al.*, 1987) in their enzyme-inhibitor complexes. On the basis of this data and of detailed molecular modeling, we chose the set of residues 13, 14, 15, 16, 17, 18, 19, 20, 21, 32, 33, 36 as consensus contact residue set. These are boxed in Table IV. We expect most of these residues to be in contact in complexes of any avian ovomucoid third domain with any serine proteinase. We expect that contact involving other residues will be moderately rare, for any avian ovomucoid third domain interacting with any serine proteinase.

Inspection of Table IV (including the subscripts) shows that 40 of the 51 residues are strongly conserved. This is defined here as greater than 85% or 130/153 conservation. These 40 residues, with X's denoting the remaining positions, are said to be the consensus sequence of avian ovomucoid third domains and listed as OM (Avi)3 in Table V.

The remaining 11 residues: 10, 15, 17, 18, 20, 21, 28, 32, 36, 43, and 55 are moderately to strongly variable. The underlined 7 of these residues are among the 12 consensus contact residues discussed above and shown in Table IV. Furthermore, these seven residues are exceptionally highly variable (more variable than the remaining four). The very large coincidence between the most variable positions and the contact positions is the basis of hypervariability of contact residues in protein proteinase inhibitors in general (Laskowski and Kato, 1980) and in avian ovomucoid third domains in particular (Laskowski *et al.*, 1987). This hypervariability coupled with the finding that ovomucoid third domains from closely related species have strikingly different inhibitory specificities (Empie and Laskowski, 1982; Park, 1985; Wynn, 1990) has been tentatively interpreted as evidence for positive Darwinian selection in proteinase inhibitor evolution (Laskowski *et al.*, 1988).

The P₁ residue (residue 18) is the most important in determining enzyme-inhibitor specificity of interaction with most enzymes. This residue is also clearly the most variable in the set. This variation is not unlimited. Thus far, as only 11 out of the possible 20 residues were found here. No new ones are added in this paper. None of the P₁ residues are aromatic and only two are charged (both of these are Lys). The uncharged polar residues are much less common than the hydrophobic aliphatic Leu and Met.

Now let us turn to the conserved residues. These are the residues of the OM(Avi)3 line in Table V. These residues seem to be conserved because they are structurally important. Some possible explanations of this structural importance were already given in Laskowski *et al.* (1987). However, another question can also be addressed that was not addressed before — which of these residues are important to the structure of all Kazal family inhibitor domains and which are important only to the structure of avian third domains? Furthermore, is there some intermediate group encompassing the third domains, but smaller than the entire set of Kazal domains, for which a consensus sequence can usefully be written?

Table IV. Composite Amino Acid Sequence of Ovomuroid Third Domains from 153 Species of Birds^a

												HLE	CHY	SGPB	
6	Val ₁₅₁	Ile ₂													
7	Asp ₁₅₀	Asn ₃													
8	Cys ₁₅₃														
9	Ser ₁₅₃														
10	Glu _{67.5}	Asp _{58.5}	Gly ₂₇												
11	Tyr ₁₄₅	His ₈													
12	Pro ₁₅₃														
13	Lys ₁₄₂	Gln ₆	Arg ₂	Met ₁	Thr ₂								■	▲	●
14	Pro ₁₅₀	His ₁	Ser ₁										■	▲	●
15	Ala ₁₀₇	Val ₃₀	Asp ₈	Glu ₂	Gly ₃	Thr ₂	Ser ₁						■	▲	●
16	Cys ₁₅₃												■	▲	●
17	Thr ₁₁₁	Ser ₂₈	Leu ₅	Pro ₄	Met ₃	Ala ₁	Arg ₁						■	▲	●
18	Leu _{65.5}	Met ₆₀	Ala _{6.5}	Gln ₅	Val _{4.5}	Thr ₃	Pro ₃	Lys ₂	Ser _{1.5}	Gly ₁	Ile ₁		■	▲	●
19	Glu ₁₃₉	Asp ₁₃	Leu ₁										■	▲	●
20	Tyr ₁₂₄	Asp ₈	Gln ₅	His ₄	Glu ₄	Arg ₂	Ser ₂	Leu ₂	Phe ₁	Asn ₁			■	▲	●
21	Arg ₆₃	Met ₇₂	Lys ₇	Phe ₅	Val ₄	Leu ₁	Thr ₁						■	▲	●
22	Pro ₁₅₃														
23	Leu ₁₃₀	Val ₉	Phe ₈	Ile ₆											
24	Cys ₁₅₃														
25	Gly ₁₅₃														
26	Ser ₁₅₃														
27	Asp ₁₅₁	Asn ₂													
28	Asn ₁₁₉	Ser ₃₄													
29	Lys ₁₃₂	Gln ₁₃	Ile ₄	Thr ₃	Glu ₁								■		
30	Thr ₁₅₀	Ser ₂	Ile ₁										■		
31	Tyr ₁₅₃														
32	Gly _{79.5}	Ser _{46.5}	Asp ₁₃	Ala _{5.5}	Asn ₅	Val ₁	His ₁	Arg ₁	Pro _{0.5}				■		●
33	Asn _{151.5}	Asp ₁	Ser _{0.5}										■	▲	
34	Lys ₁₅₁	Arg ₁	Glu ₁												
35	Cys ₁₅₃														
36	Asn ₁₁₈	Asp ₁₈	Ser ₆	Ala ₄	Gly ₂	Tyr ₁	Thr ₁	Ile ₁					■	▲	●
37	Phe ₁₅₃														
38	Cys ₁₅₃														
39	Asn ₁₅₁	Ser ₂												▲	
40	Ala ₁₅₃														
41	Val ₁₄₉	Ala ₃	Phe ₁												
42	Val ₁₄₁	Ala ₁₀	Met ₁	Leu ₁											
43	Glu ₇₁	Asp ₇₄	Lys ₅	Gln ₂	His ₁										
44	Ser ₁₄₈	Lys ₄	Arg ₁												
45	Asn ₁₅₂	Ser ₁													
46	Gly ₁₄₈	Val ₅													
47	Thr ₁₄₈	del ₅													
48	Leu ₁₄₄	Ile ₄	del ₅												
49	Thr ₁₄₈	Asn ₂	Ser ₂	Ile ₁											
50	Leu ₁₅₀	Val ₂	Phe ₁												
51	Ser ₁₃₃	Asn ₁₁	Arg ₅	Gly ₄											
52	His ₁₄₆	Arg ₅	Tyr ₁	Asn ₁											
53	Phe ₁₄₈	Leu ₃	Ile ₁	Pro ₁											
54	Gly ₁₅₀	Glu ₃													
55	Lys ₁₂₈	Glu ₂₀	Gln ₂	Arg ₂	Val ₁	Thr ₁								▲	
56	Cys ₁₅₃														

^a Residues are numbered according to a notation introduced by Weber *et al.* (1981). The third domain proper starts at residue 6 (see Table II and Laskowski *et al.*, 1987). Given vertically is the sequence of turkey ovomucoid third domain. Given horizontally are all the kinds of residues found at a given position in 153 species (Kato *et al.*, 1987; Laskowski *et al.*, 1987, 1990; and the present paper). The subscripts indicate the number of species in which the particular residue was found. In cases of dimorphism, an arbitrary weight of 0.5 is given to each dimorphic form. On the left the residues which are in contact between turkey ovomucoid third domain in its complexes with human leukocyte elastase (HLE) (Bode *et al.*, 1986; and Bode, personal communication), bovine chymotrypsin A α , CHY (Fujinaga *et al.*, 1987) and *Streptomyces griseus* proteinase B, SGPB (Read *et al.*, 1983) are indicated by ■, ▲, ●, respectively. On the basis of this data and of extensive modeling of the OMTKY3 complexes in which the various residues of the inhibitor were substituted, a consensus set of contact residues between OM(Avi)3 and serine proteinases was constructed. The contact symbols for these positions are boxed.

Sequences^a

^a OMAL14, fourth domain of alligator (*Alligator mississippiensis*) ovomucoid (March 1980); OICH17, seventh domain of chicken (*Gallus gallus*) ovomucoid (Scott *et al.*, 1987); SMCAN2, second domain of dog (*Canis familiaris*) submandibular inhibitor (Reisinger *et al.*, 1987); OMI(Avi)3, consensus sequence of avian ovomucoid third domains based on the data (Table IV) and obtained as described in the text; b-domain, consensus sequence of b-type domains obtained from the four sequences above; Kz domain, consensus sequence of all Kazal domains based on 400 such domains in our database. (See text.)

^a OMAL14, fourth domain of alligator (*Alligator mississippiensis*) ovomucoid (March 1980); OICH17, seventh domain of chicken (*Gallus gallus*) ovomucoid (Scott *et al.*, 1987); SMCAN2, second domain of dog (*Canis familiaris*) submandibular inhibitor (Reisinger *et al.*, 1987); OMI(Avi)3, consensus sequence of avian ovomucoid third domains based on the data (Table IV) and obtained as described in the text; b-domain, consensus sequence of b-type domains obtained from the four sequences above; Kz domain, consensus sequence of all Kazal domains based on 400 such domains in our database. (See text.)

speciation. All other ratite birds third domains (ostrich, greater and lesser rhea and emu) whose sequences are known have P₁Leu. P₁Leu is quite common (Table IV), and inhibitors with P₁Leu are very efficient inhibitors of many enzymes.

The next three entries in Table I (four entries in Table II because of a dimorphism) deal with tinamous. The sequences of two other tinamous were already known (Laskowski *et al.*, 1987). Tinamous are the only birds whose third domains have an internal deletion in the third domains. This deletion, common to all five tinamou species studied, involves two residues following the Asn⁴⁵ residue, which serves as a glycosylation site. In almost all nontinamou birds⁵ the sequence there is ... Asn⁴⁵Gly⁴⁶Thr⁴⁷... In all tinamous, the sequence is ... Asn⁴⁵Val⁴⁸Thr⁴⁹... We have previously suggested that not only the deletion of two residues but the mutation of Gly⁴⁶ to Val⁴⁸ can be a result of the same event — deletion of six nucleotides (Laskowski *et al.*, 1987). The persistence of Val⁴⁸ in all tinamous lends some additional credence to this suggestion.

In addition, all five tinamou species have Gly¹⁰ (however, so do many other species of birds, Table IV), Phe²¹ (which is thus far limited to tinamous), and Gin²⁹ (a few other species have this also). Tinamou third domains are the only third domains sequenced thus far that lack Ser⁴⁴. Instead, four species have Lys⁴⁴ and one (RTI) has Arg⁴⁴, which may be an evolutionary intermediate in Ser⁴⁴ → Arg⁴⁴ → Lys⁴⁴ conversion. Four of the five tinamous lack the highly conserved (but not unvaried in other birds) Phe⁵³. However, OMGTI3 contains Phe⁵³. It is possible that the Phe⁵³ → Leu⁵³ and/or Phe⁵³ → Ile⁵³ fixations in tinamous followed the common deletion event that set the tinamou third domains apart. On the basis of just two first (unpublished) and third domain sequences of tinamous, Laskowski and Fitch (1989) concluded that tinamou ovomucoid sequences are the most distant from the sequences of all other birds. This conclusion was strongly reinforced by the four new sequences reported here.

Since tinamou sequences are so different from other bird ovomucoid sequences, it was interesting to ask whether this set of sequences also shows the contact residue hypervariability characteristic of avian ovomucoids in general. We note changes at four contact residues: Ala¹⁵ → Asp¹⁵ in OMRTI3, Leu¹⁸ → Ala¹⁸ in OMGTI3, Thr²⁰ → Phe²⁰ in

OMCTI3, Ser³² → Thr³² in OMNTI3. The dimorphism at position 32 in OMNTI3 is one of the six dimorphisms found in the third domains from 153 species. Of the six, five (including OMNTI3) occur in contact positions (see Table IV), in striking support of the hypervariability conclusion. It seems clear that the contact positions in tinamou third domains are as hypervariable as in third domains of other birds.

The screamers are a small group of three species in two genera, generally classified as one of the two superfamilies of *Anseriformes*. The only known ovomucoid sequences of screamers were those of the first (unpublished) and third (Laskowski *et al.*, 1990) domains of northern screamer, NOS, *Chauna chavaria*. These sequences caused difficulties in phylogenetic tree construction (Laskowski and Fitch, 1989). While quite different from them, OMNOS1 and OMNOS3 showed some similarities to the two then available tinamou sequences, OMBTI and OMCTI, and in most construction NOS was the nearest neighbor of the two tinamous. Part of the reason for this may have been accidental; and such accidental congruences are much more likely when distant sets each composed of only one (e.g., northern screamer) or two (the two tinamous) taxa are examined. The gift of the Sibley collection allowed us to add several more tinamou sequences as well as the southern screamer, SOS, *Chauna torquata*. However, disappointingly the sequence of OMSOS3 turned out to be identical to the already known OMNOS3 (Table III).

We have previously sequenced two third domains from the goose genus *Anser*—OMGOO3 from *Anser anser* — graylag goose — and OMBHG3 from *Anser indicus* — bar-headed goose. Both gave identical sequences (GOO = BHG in Table III). On the other hand, the only sequence we have from the genus *Branta*, that of the Canada goose — *Branta canadensis*, OMCNG3 — differs from the two above by having an Asp³² → Gly³² replacement. It is surprising that emperor goose, *Anser canigulus*, third domain OMEMG3 is identical to OMCNG3 and not to the two *Anser* sequences.

The Egyptian goose, EGG, *Alpochen aegyptiaca*, is a goose in name only. It belongs to the tribe *Tadornini*, of shelducks and sheldgeese (Johnsgard, 1978). Other members of this tribe have third domains sequences characteristic of the large group of ducks (see UPG = ... in Table III), and so rather expectedly does OMEGG3.

⁵ Ostrich is the only exception, its sequence is ... Ser⁴⁵Gly⁴⁶Thr⁴⁷....

One might have expected identity of all duck-related third domains. This is not the case. The next two sequences — OMRED3, redhead, *Anthaya americana* and OMRPO3, red-crested pochard *Netta rufina* — represent the two extant genera of the tribe *Aynithini* or pochards. Their sequences differ from most duck sequences (see OMEGG3) by having an Asp⁷ → Asn⁷ replacement. As is seen in Table IV and V, Asp⁷ is very strongly conserved. The reason for this conservation is not hard to find as all available three-dimensional structures of avian ovomucoid third domains show a hydrogen bond between unvaried Ser⁹ and the strongly conserved Asp⁷. As expected, we have shown that OMRED3 has a lower melting temperature, T_m , than OMDUK3. Again, as expected, this difference disappears at low pH (Lu and Giletto, unpublished).

There are some interesting variations among the third domain sequences of birds of prey belonging to the group *Accipitridae*. This was not the case for hooded vulture, HDV, whose sequence joins the identity cluster TTE = HDV = ... (Table III).

As the megapodes are phylogenetically interesting, we were quite anxious to enlarge our sequence set from two to three members. Wattled brush turkey, WBT, third domain sequence differs from the previously determined (Laskowski et al., 1987) Malee fowl by three replacements, while it differs from the common scrub hen by 12.

Sequences of three grouse third domain species are already known. Two of them are not only identical (sage grouse, SGG, and greater prairie chicken, GPC), but belong to the large identity cluster (starting with TTR in Table III) of third domains of many, but not all pheasants and especially (see below) of all tragopans. The third ruffed grouse, RFG, differs from these are two positions. Therefore, it was interesting to sequence another grouse third domain from white-tailed ptarmigan, TTR. As already pointed out it turned out to be a member of the identity cluster.

The crested wood partridge sequence, OMWPA3, was a large technical effort, as it was difficult to obtain the third domain. Once obtained, the sequence proved very surprising. While on the basis of this sequence, there are no grounds for questioning the standard assignment of WPA as a phasianoid bird, several changes observed in this sequence are unusual. There are nine differences in sequences between OMWPA3 and a "typical" phasianoid belonging to the phasianoid identity cluster starting with OMTTR3 (Table III). These are: Ser⁵ → Asn⁵, Glu¹⁰ → Gly¹⁰, Tyr²⁰ → Ser²⁰, Leu²³ → Ile²³,

Lys²⁹ → Thr²⁹, Gly³² → Ser³², His⁵² → Tyr⁵², Phe⁵³ → Pro⁵³, Lys⁵⁵ → Val⁵⁵. Thus far, all phasianoids had Ser⁵ and only cracids had Asn⁵. Gly¹⁰ is extremely rare among phasianoids, only OMBPA3 has it. Ser²⁰ never occurs in other phasianoids. In the total sequence set, it occurs only twice (Table IV). Its other occurrence is in red winged tinamou ovomucoid third domain (OMRTI3) discussed above. Ile²³ is rare in the total sequence set (Table IV) and Thr²⁹ is even less common. In addition to OMWPA3, it was only found in two of the four *Passeriformes* sequences discussed below. Gly³² → Ser³² is a very common change in the total sequence set, although rare among phasianoids. Tyr⁵² is unique in the total sequence set, and His⁵² — which it replaces — is rather strongly conserved. Pro⁵³ is also unique. It replaces Phe⁵³, which is completely conserved in all birds other than tinamous (see above). Surprisingly, the Phe⁵³ → Pro⁵³ requires at least two nucleotide changes. This is very uncommon for changes observed at virtually unvaried positions in ovomucoids. It may be worth remarking that the Ramachandran angles of Phe⁵³ in OMSVP3 (Bode et al., 1985) would allow for a Phe⁵³ → Pro⁵³ replacement without main chain distortion. Val⁵⁵ is unique; most residues at this position are either Lys⁵⁵ or Glu⁵⁵. The overall impression is OMWPA3 was freed from many constraints that applied in the evolution of other third domains in general and of phasianoid third domains in particular.

The BTP and CTP sequences are two more sequences in the genus *Tragopan*. Two more were determined in the past, STP and TTP. All four fall into the identity cluster starting with TTR in Table III. In part on the basis of the identity of most grouse third domain sequences with the then two known tragopan third domain sequences. Laskowski and Fitch (1989) suggested that grouse, tragopans, and Himalayan monal pheasant (HMP) are especially closely related. The present results (i.e., BTP = CTP = STP = TTP = HMP = TTR = GPC = SGG) are consistent with this conclusion. It should be noted that the conclusion is principally based on first rather than third domain sequences.

The sequence of white-eared pheasant ovomucoid third domain, OMWEP3 provides the sequence of the third and last species (Gruson, 1976) in the genus *Crossoptilon*. They are all identical. All belong to the cluster starting with TTR (i.e., BEP = WEP = BRP).

In contrast to the genus *Crossoptilon*, where all the extant species have the same third domain

sequence, there is considerable variation within the pheasant genus *Lophura*, where many species are present. Sequences of third domains of seven *Lophura* species are now known. Of these, three belong to the large pheasant identity cluster starting with TTR (i.e., Edwards' pheasant, EDP; silver pheasant, SVP; and Swinhoe's pheasant, SWP). On the other hand, Bulwer's pheasant, BLP, differs from these by a single Ala¹⁵ → Thr¹⁵ substitution. Kalij pheasant differs from the sequences in the large identity cluster by two changes Met¹⁸ → Ile¹⁸ and Gly³² → Asp³². The newly sequenced Siamese fireback pheasant, SFP, third domain is identical to that of crested fireback pheasant, CFP (i.e., SFP = CFP in Table III). They differ from the cluster sequences by Thr³⁰ → Ser³⁰ and Gly³² → Asp³². Thus, they differ from OMKAP3 and by two replacements Ile¹⁸ → Met¹⁸ and Thr³⁰ → Ser³⁰. Note that not only (Table IV) is Thr³⁰ very strongly conserved in avian third domains, it is strongly conserved in all Kazal type domains. Replacement of Thr³⁰ leads to a drop in T_m (Otlewski and Laskowski, unpublished), but has no effect upon the equilibrium constants, K_a , for association with various enzymes (Park, 1985). It is closely analogous to the Asp⁷ → Asn⁷ replacement discussed in the case of OMRED3 = OMRPO3 above.

Among the previously known sequences, only one avian third domain sequence is missing the Asp²⁷...Tyr³¹ side chain...side chain hydrogen bond. This is grey peacock pheasant, *Polypsectron bicalcaratum*, third domain OMGPP3. The Asp²⁷...Tyr³¹ hydrogen bond is strongly conserved, not only in avian ovomucoid third domains, but in all Kazal domains. As the genus, *Polypsectron*, contains several species, it was of interest to obtain other sequences. Unfortunately, aside from the difficulty of obtaining various *Polypsectron* eggwhites, once the ovomucoids are isolated third domains are difficult to obtain. We succeeded here with Palawan peacock pheasant, OMPPP3. Its sequence is quite similar to the earlier reported OMGPP3, but differs from it at four positions. Both share Asn²⁷ and Tyr³¹ and thus both are missing the Asp²⁷...Tyr³¹ interaction. It seems likely that the third domains of ovomucoids from all or most species of *Polypsectron* are missing the Asp²⁷...Tyr³¹ hydrogen bond. The resultant low stability of these domains may be an explanation for the difficulty in their isolation.

Hoatzin, HTZ, is the only bird where extensive bacterial fermentation of plant food in the foregut is a part of the digestive process. The "flying cow" has

long been a phylogenetic puzzle (Sibley and Ahlquist, 1990). The most often suggested affinities are to cracid birds and among them specifically to chachalaca and to cuckoos and among these specifically to guira cuckoo. The sequence of four cracid third domains, among them chachalaca are known. Prior to this work, so were the sequences of two cuckoos, but this work extends it to four and includes guira cuckoo (see below). Therefore, obtaining the OMHTZ3 sequence appeared exciting. As might have been anticipated, OMHTZ3 appears quite distant from both cuckoos and cracids, and more detailed analysis will be required. It is unlikely that the ovomucoid third domain and first domain sequences can resolve the question in a decisive fashion. The curious characteristics of the OMHTZ3 sequence Leu¹⁷ at P₂ position and Ser²⁶ correspond to residues in cracid third domains.

The two cranes, *Grus japonica*, OMJPC3 and *Balearica pavonina*, OMCOC3, proved disappointing as their third domain sequences are identical and add to the large crane identity cluster starting with WTC in Table III. Note that a residue in the connecting peptide, not in the third domain proper of OMCOC3 differs by Thr³ → Met³ substitution. This is the first finding of Met in any position in the connecting peptide between second and third domains. As OMCOC3 contains an additional internal Met²¹, Met³ is not useful preparatively to obtain intact OMCOC3 by CNBr cleavage.

Two cuckoo ovomucoid third domain sequences were already reported: roadrunner, OMRDR3, and coral-billed ground cuckoo, OMCGC3. The two sequences differ at nine positions, but even more strikingly contain within them several residues that are unique or very rare in the entire avian third domain set. These are in OMRDR3 Ash⁷ (this occurs only in three cases, the other two OMRED3 and OMRPO3 were discussed previously) and Leu¹⁹ (the P₁' position) — this is unique, furthermore all other P₁' residues are either Glu or Asp. In OMCGC3, the unusual residues are the unique His³², Tyr³⁶, and Phe⁵⁰. It was of interest to add some cuckoo sequences to see whether these rare residues might occur in them. An additional motivation was that some cuckoos are parasitic. We were contemplating a study of possible adaptation of a particular cuckoo ovomucoid third domain sequence to that of the host. Finally, more cuckoo sequences were needed for comparison to the hoatzin sequence, especially that of guira cuckoos, that one is mentioned by Sibley and Ahlquist (1990) as particularly

similar to hoatzin. We report here two new cuckoo sequences OMCBC3 and OMGGC3. OMGGC3, guira cuckoo third domain is quite similar to OMRDR3. It differs from it only at two positions the Asp⁷ → Asn⁷ replacement and the Glu¹⁹ → Leu¹⁹. The OMCBC3 is more distant from both OMRDR3 and OMGGC3, but it does not share any of the rare (unique) residues nor does it introduce any of its own. The one closest to hoatzin is guira cuckoo.

Sibley and Monroe (1990) recognize 9672 species of birds. Of these, 5712 are passerine (i.e., belong to the single order *Passeriformes*). The remaining 3960, called nonpasserine, make up the 21 remaining orders. Thus far, ovomucoid third domain sequences were reported only from the nonpasserine birds. To redress the balance, the last four species in Tables I and II are all passerine. The four sequences are not strikingly unusual, they are rather typical avian ovomucoid third domains. However, all four share a unique substitution. Leu⁴⁸ (which is unvaried in all nonpasserines except the tinamous where it is deleted) is replaced by Ile⁴⁸. In addition, all four passerine sequences have Asp¹⁹ a moderately rare residue, but present in several other orders.

The first two sequences are from birds belonging to the Parvorder *Passerida*. In addition to the general passerine characteristics, they also share the very rare Lys²⁹ → Thr²⁹ replacement (the only other sequence with this replacement is OMWPA3 described above). The house finch third domain shows a unique Ile³⁶. The next two sequences are from crows, both belonging to the genus *Corvus*. The sequences differ from one another by the single replacement Lys¹³ → Arg¹³. In OMCP3. Both crow sequences share thus far undescribed Lys⁵⁵ → Arg⁵⁵ replacement.

It is clear that the set from 153 species of birds is far from saturating all that we would like to know about bird ovomucoid third domains.

ACKNOWLEDGMENTS

We are grateful to all the eggwhite donors listed in Table I, whose gifts were decisive in this work. Special thanks are due to Prof. C. G. Sibley, who donated the Sibley Collection of eggwhite and to Mrs. Barbara Corbin, who completed the indexing of this large collection. Prof. Ying-Ming Lin of Tennessee State University, Ms. Sreenganga Chandra and Ms. Bornadata Evans isolated OMWEP3, OMEMG3, and OMEGG3 during short working visits to our laboratory.

Supported by National Institutes of Health Grant GM10831.

REFERENCES

- Ardelt, W., and Laskowski, M. Jr. (1991). *J. Mol. Biol.* **220**, 1041–1053.
- Barkholt, V., and Jensen, A. L. (1989). *Analytical Biochem.* **177**, 318–322.
- Bigler, T. L. (1991). Ph.D. thesis, Purdue University.
- Bigler, T. L., Lu, W., Park, S. J., Tashiro, M., Wieczorek, M., Wynn, R., and Laskowski, M., Jr. (1993). *Protein Science* (in press).
- Bode, W., Wei, A., Huber, R., Meyer, E., Travis, J., and Neumann, A. (1986). *EMBO J.* **5**, 2453–2458.
- Bode, W., Epp, O., Huber, R., Laskowski, M. Jr., and Ardelt, W. (1985). *Eur. J. Biochem.* **147**, 387–395.
- Bolander, M. E., Young, M. F., Fisher, L. W., Yamada, Y., and Termine, J. D. (1988). *Proc. Natl. Acad. Sci. USA* **85**, 2919–2923.
- Creighton, T. E., and Darby, N. J. (1989). *Trends in Biochem. Sci.* **14**, 319–324.
- Empie, M. W., and Laskowski, M. Jr. (1982). *Biochemistry* **21**, 2274–2284.
- Feeney, R. E., and Allison, R. G. (1969). In *Evolutionary Biochemistry of Proteins*, Wiley-Interscience, New York, p. 30.
- Forwith, K. D. (1992). Ph.D. thesis, Purdue University.
- Fujinaga, M., Sielecki, A., Read, R., Ardelt, W., Laskowski, M., Jr., and James, M. N. G. (1987). *J. Mol. Biol.* **195**, 397–418.
- Gruson, E. S. (1976). *Checklist of the WORLD'S BIRDS*, Quadrangle/The New York Times Book Co., New York.
- Hoy, T. G., Ferdinand, W., Harrison, P. M. (1974). *Int. J. Pept. Protein Res.* **6**, 121–140.
- Johnsgard, P. A. (1978). In *Ducks, Geese and Swans of the World*, University of Nebraska Press, Lincoln, pp. xviii.
- Kato, I., Schrode, J., Wilson, K. A., and Laskowski, M., Jr. (1976). *Protides Biol. Fluids* **23**, 235–243.
- Kato, I., Schrode, J., Kohr, W. J., and Laskowski, M., Jr. (1987). *Biochemistry* **26**, 193–201.
- Kimura, M. (1983). In *The Neutral Theory of Molecular Evolution*, Cambridge University Press, Cambridge, pp. 15–31.
- Komiyama, T., Bigler, T. L., Yoshida, N., Noda, K., and Laskowski, M., Jr. (1991). *J. Biol. Chem.* **266**, 10,727–10,730.
- Laskowski, M., Jr., and Fitch, W. M. (1989). In *The Hierarchy of Life* (Fernholm, B., Bremer, K., and Jornvall, H., eds.), Elsevier Science Publishers B.V., Karlskoga, Sweden, pp. 371–387.
- Laskowski, M., Jr., and Kato, I. (1980). *Ann. Rev. Biochem.* **49**, 593–626.
- Laskowski, M., Jr., Empie, M. W., Kato, I., Kohr, W. J., Ardelt, W., Bogard, W. C., Jr., Weber, E., Papamokos, E., Bode, W., and Huber, R. (1981). *Colloq. Ges. Biol. Chem.* **32**, 136–152.
- Laskowski, M., Jr., Tashiro, M., Empie, M. W., Park, S. J., Kato, I., Ardelt, W., and Wieczorek, M. (1983). In *Proteinase Inhibitors: Medical and Biological Aspects* (Katanuma, N., Umezawa, H., and Holtzer, H., eds.), Japan Scientific Societies Press, Tokyo/Springer-Verlag, Berlin, pp. 55–68.
- Laskowski, M., Jr., Kato, I., Ardelt, W., Cook, J., Denton, A., Empie, M. W., Kohr, 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**, 202–221.
- Laskowski, M., Jr., Kato, I., Kohr, W. J., Park, S. J., Tashiro, M., and Whatley, H. E. (1988). *Cold Spring Harbor Symp. Quant. Biol.* **52**, 545–553.

- Laskowski, M., Jr., Park, S. J., Tashiro, M., and Wynn, R. (1989). In *Protein Recognition of Immobilized Ligands* (T. W. Hutchen, ed.), Alan R. Liss, New York, pp. 149–168.
- Laskowski, M., Jr., Apostol, I., Ardelt, W., Cook, J., Giletto, A., Kelly, C. A., Lu, W., Park, S. J., Qasim, M. A., Whatley, H. E., Wiczorek, A., and Wynn, R. (1990). *J. Protein Chemistry* **9**, 715–725.
- Lu, W., Komiyama, T., Warne, N., Laskowski, M., Jr., Chiang, Y., Rothberg, I., Ryan, K., Anderson, S. (1992). Poster Session at Sixth Annual Symposium of the Protein Society, July 25–29, 1992.
- March, C. J. (1980). Ph.D. thesis, Purdue University.
- Park, S. J. (1985). Ph.D. thesis, Purdue University.
- Read, R., Fujinaga, M., Sielecki, A., and James, M. N. G. (1983). *Biochemistry* **22**, 4420–4433.
- Reisinger, P. W. M., Hochstrasser, K., Göttlicher, I., Eulitz, M., and Wachter, E. (1987). *Biol. Chem. Hoppe-Seyler* **368**, 717–726.
- Rhodes, B. M., Bennett, N., and Feeney, R. E. (1960). *J. Biol. Chem.* **235**, 1686–1693.
- Rupp, F., Payan, D. G., Magill-Solc, C., Cowan, D. M., and Scheller, R. H. (1991). *Neuron* **6**, 811–823.
- Scott, K., Cannell, G. R., and Zerner, B. (1975). *Anal. Biochem.* **69**, pp 474–484.
- Scott, M. J., Huckaby, C. S., Kato, I., Kohr, W. J., Laskowski, M., Jr., Tsai, M.-J., and O'Malley, B. W. (1987). *J. Biol. Chem.* **262**, 5899–5907.
- Sealock, R. W., and Laskowski, M., Jr. (1969). *Biochemistry* **8**, 3703–3710.
- Sibley, C. G., and Monroe, B. L., Jr. (1990). *Distribution and Taxonomy of Birds of the World*, Yale University Press, New Haven.
- Sibley, C. G., and Ahlquist, J. E. (1990). *Phylogeny and Classification of Birds*, Yale University Press, New Haven.
- Warne, N. (1990). Ph.D. thesis, Purdue University.
- Wiczorek, M., Park, S. J., and Laskowski, M., Jr. (1987). *Biochem. Biophys. Res. Comm.* **144**, 499–504.
- Wiczorek, M., and Laskowski, M., Jr. (1982). *Fed. Proc.* **41**, 762.
- Wynn, R. A. (1990). Ph.D. thesis, Purdue University.