Ovomucoid Third Domains from 100 Avian Species: Isolation, Sequences, and Hypervariability of Enzyme-Inhibitor Contact Residues†

Michael Laskowski, Jr.,* Ikunoshin Kato,‡ Wojciech Ardelt, James Cook,§ Andrew Denton,‖ Mark W. Empie,⊥ William J. Kohr,# Soon Jae Park,○ Kevin Parks,× Barbara L. Schatzley, Oeyvind L. Schoenberger,% Misao Tashiro,& Gisela Vichot, Harry E. Whatley, Anna Wieczorek, and Maciej Wieczorek

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

Received July 14, 1986; Revised Manuscript Received August 25, 1986

ABSTRACT: Ovomucoids were isolated from egg whites of 100 avian species and subjected to limited proteolysis. From each an intact, connecting peptide extended third domain was isolated and purified. These were entirely sequenced by single, continuous runs in a sequencer. Of the 106 sequences we report (five polymorphisms and chicken from the preceding paper [Kato, I., Schrode, J., Kohr, W. J., & Laskowski, M., Jr. (1986) *Biochemistry* (preceding paper in this issue)]), 65 are unique. In all cases except ostrich (which has Ser^{45}), the third domains are either partially or fully glycosylated at Asn^{45}. The majority of the third domain preparations we isolated are carbohydrate-free. Alignment of the sequences shows that their structurally important residues are strongly conserved. On the other hand, those residues that are in contact with the enzyme in turkey ovomucoid third domain complex with *Streptomyces griseus* proteinase B [Read, R., Fujinaga, M., Sielecki, A. R., & James, M. N. G. (1983) *Biochemistry 22*, 4420-4433] are not conserved but instead are by far the most variable residues in the molecule. These findings suggest that ovomucoid third domains may be an exception to the widely accepted generalization that in protein evolution the functionally important residues are strongly conserved. Complete proof will require better understanding of the physiological function of ovomucoid third domains. This large set of variants differing from each other in the enzyme-inhibitor contact area and augmented by several high-resolution structure determinations is useful for the study of our sequence to reactivity (inhibitory activity) algorithm. It is also useful for the study of several other protein properties. In the connecting peptide fragment most phasianoid birds have the dipeptide Val^4-Ser^5, which is absent in most other orders. This dipeptide is often present in only 70-95% of the molecules and appears to arise from ambiguous excision at the 5′ end of the F intron of ovomucoid. Connecting peptides from the ovomucoids of cracid birds contain the analogous Val^4-Asn^5 peptide. In laughing kookaburra ovomucoid third domain we found (in 91% of the molecules) Gln^{5A}, which we interpret as arising from ambiguous intron excision at the 3′ end of the F intron.

In their classic paper [later augmented by many other papers from the Feeney group summarized in Feeney and Allison (1969) and Feeney (1970)] Rhodes et al. (1960) made two major observations. First, they found that ovomucoids from various avian species can be single, double, or triple headed. This observation is explained in the preceding paper (Kato et al., 1986), which shows that avian ovomucoids consist of three tandem, homologous, Kazal-type domains, each with an actual or putative reactive site. Their second observation was that ovomucoids from closely related species (e.g., chicken, turkey, golden pheasant) show very different inhibitory activity. This was a clear suggestion of unusual variation of residues responsible for inhibitory activity.

After our first ovomucoid sequence (Kato et al., 1976), we went on to sequence several ovomucoid third domains (chicken, turkey, ring-necked pheasant, Japanese quail, and chachalaca). We were startled to find that, in the five species, there were five different residues (Ala, Leu, Met, Lys, and Gln, respectively) at the reactive site position P_1. This extreme hypervariability of the P_1 position was difficult to understand. At that point we did not know whether this type of variation was expected on the race, species, or genus level and, furthermore, whether it was limited to P_1 or to all of the residues involved in enzyme-inhibitor contact. The answer to the first question could only come from sequencing ovomucoid third domains from many more closely related species, and this paper, after 101 sequences, provides a complex answer. The answer to the second question had to await X-ray crystallography of an enzyme-ovomucoid third domain complex (Fujinaga et al., 1982; Read et al., 1983) showing which positions are in contact. When these positions were examined in the large sequence set, we saw that 8 of the 11 enzyme-inhibitor contact positions are strongly hypervariable, casting doubt on the universality of the generalization that in orthologous proteins of closely related species the functional residues are conserved. The complete falsification of this generalization must await direct proof that for some avian ovomucoid third domains inhibition of serine proteinases is a biological function.

As the sequencing effort to prove and define hypervariability went on, we realized that the set of sequenced ovomucoid third domains with accurately determined three-dimensional structures (Weber et al., 1981; Papamokos et al., 1982; Fujinaga et al., 1982; Read et al., 1983; Bode et al., 1985) was

† Supported by National Institutes of Health Grant GM10831.

‡ Present address: Centocor, Malvern, PA 19355.

§ Present address: Eli Lilly and Co., Indianapolis, IN 46285.

‖ Present address: Chemical Abstracts Service, Columbus, OH 43210.

⊥ Present address: Dow Chemical, New England Lab, Wayland, MA 01778.

Present address: Genentech, Inc., South San Francisco, CA 94080.

○ Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.

× Present address: Hazeltine, Reston, VA 22091.

% Present address: Krankenhaus Rohrbach, 6900 Heidelberg 1, West Germany. Deutsche Forschungsgemeinschaft Postdoctoral.

& Present address: Department of Food Science, Kyoto Prefectural University, Shimogamo Kyoto 606, Japan.

0006-2960/87/0426-0202$01.50/0 © 1987 American Chemical Society

a superb resource for sequence to reactivity (inhibitory activity) studies. Even later we realized that it was an excellent set for sequence to protein property studies including inhibition, antigenic activity, stability, NMR and optical spectra, and chromatographic behavior. Once this realization was made, the acquisition of new third domains went on with the twin goals of expanding our understanding of hypervariability and of generating the best set for various protein chemical studies. After completing the sequences from 100 avian species (101 including chicken from the previous paper), it is time to report.

Materials and Methods

Biochemicals. Early samples of *Staphylococcus aureus* proteinase V8 (Drapeau, 1976) were kindly provided by Dr. G. Drapeau; later, we bought this enzyme from Miles Scientific. Thermolysin (from *Bacillus thermoproteolyticus*) was purchased from Calbiochem. Porcine pancreatic elastase I (Satoch et al., 1979) was a gift from the late Prof. M. Laskowski, Sr. Bio-Gel P-10 (200–400 mesh) was obtained from Bio-Rad. DEAE-Sepharose[1] CL-6B, CM-Sepharose CL-6B, and Sephadex G-15 were products of Pharmacia.

Nomenclature and Acquisition of Egg Whites. Eggs from various species of birds were obtained from numerous donors or purchased from breeders (see Table I and the Appendix). When the eggs arrived we first cataloged them by scientific name as given in Gruson (1976). We then coined a three-letter abbreviation based on the English name. Gruson was chosen as our standard because it lists "all" birds and gives each only a single English and a single scientific name. Abbreviations were used to index the computer files of sequences and data on egg supply. Gruson's (1976) arrangement of families is used throughout this paper. It can be viewed by nonornithologist readers as the conventional order. Gruson lists only families, not orders. Orders are listed in Sibley and Ahlquist (1972) from which we have adopted the following groupings (to follow this the reader should look at Table I or Table II): Ratitae, OST[1] to EMU; Tinamiformes, BTI and CTI; Pelecaniformes, LBC and ADR; Anseriformes, MPG to LKD; Falconiformes (but since all the species we list are members of the family Accipitridae we use the family name), BRK to WTE; Megapodiidae, [here we should talk about Galliformes, but because the assignment of megapodes, cracids, and phasianoids to the same order is controversial (Prager & Wilson, 1976), we split the three groups], CSH and MLF; Cracidae, CHA to PGN; Phasianoidea, RFG to OTK; Gruiformes, BUQ to COO; Charadriiformes, SPL and BGL; Cuculiformes, RDR and CGC.

Once the eggs are identified they are opened, and if possible, the egg whites are separated from the yolks. The egg whites are then frozen and stored at −20 °C. They appear to be indefinitely stable in this form. Unfortunately, many eggs we receive have been extensively incubated to ascertain that they are not fertile. In such eggs the yolk membrane is frequently lysed, and the separation of egg white is no longer possible. Unless the mixture is judged as extensively rotten, it is still frozen and saved. We have been frequently successful in obtaining ovomucoid and later ovomucoid third domain from such mixtures, but generally the yields both at the crude ovomucoid isolation step and especially at the third domain isolation step (see below) were much lower than from cleanly separated egg whites.

Occasionally we were given already separated, frozen egg whites. In cases involving gifts from other countries, egg white was provided in lyophilized form. This works as well as "fresh" egg white from our freezer.

Finally, at the start of this study we were greatly aided by Prof. A. C. Wilson, who had saved egg white solutions that were passed through cation exchangers to retain cationic proteins (lysozyme). Anionic proteins were eluted and served us well as an ovomucoid source.

Preparation of Avian Ovomucoids. This was carried out as described for chicken ovomucoid in the preceding paper, except that the final purification step on CM-Sepharose was frequently omitted.

Preparation of Ovomucoid Third Domains. Connecting peptide extended third domains were generated by limited proteolysis of entire ovomucoids with one of the following proteinases: *Staphylococcus aureus* proteinase V8 (Spase V8), thermolysin, or porcine pancreatic elastase I (see Table I). The proteolysis was routinely carried out at 21 °C in 0.2 M Tris-HCl buffer (pH 8.1) containing 0.02 M $CaCl_2$. The concentration of ovomucoids in the incubation mixture varied from 20 to 60 mg/mL. The proteinases were added in amounts equivalent to 0.5–3% (Spase V8), 1–2% (thermolysin), or 1% (elastase) of the ovomucoid treated, and the reaction proceeded for 24–72 h. In several cases in which generation of the third domain was very difficult, the standard conditions of the proteolysis were modified. Lithium bromide (a chaotropic agent) (1 M) was used for proteolysis of the ovomucoid from grey peacock-pheasant (GPP). Limited proteolysis of whistling tree duck (HTD) ovomucoid was performed with 2% thermolysin at 21 °C for 24 h and continued for 3 h more at 50 °C. Incubation time was extended to over 100 h in the case of black-crowned night heron (BNH), and the concentration of thermolysin was increased to 6% for the generation of common hill patridge (CPA) and of bare-throated francolin (BTF) ovomucoid third domains.

In all cases generation of third domains was monitored by analytical size-exclusion HPLC. Aliquots of the incubation mixture were withdrawn at different times and run on a Varian MicroPak (G-2000SW) column (Figure 1). In most cases, the appearance of at least two peaks (in addition to the initial ovomucoid peak) could be observed, e.g., peaks 2 and 3 in Figure 1A. They contained glycosylated and unglycosylated ovomucoid third domains, respectively. In many cases (depending on the ovomucoid used) we found various amounts of ovomucoid second domain in peak 2 in addition to the glycosylated third domain. In other cases, e.g., ERF and CHA, a glycopeptide cut out from the ovomucoid first domain was also found in peak 3 in addition to the unglycosylated third domain. The glycopeptide consists of eight residues corresponding to the 50–57 fragment of the first domain [see Kato

[1] Abbreviations: Spase V8, *Staphylococcus aureus* proteinase V8; HPLC, high-performance liquid chromatography; DEAE, diethylaminoethyl; CM, carboxymethyl; NMR, nuclear magnetic resonance; Tris, tris(hydroxymethyl)aminomethane; Hse, homoserine. The bird species are abbreviated by a three-letter abbreviation, e.g., OST = ostrich. These are defined in Table I and are generally based on the English names of birds appearing in Gruson (1976). The abbreviation OM means ovomucoid. Thus, OMOST means ostrich ovomucoid. The designator 3 indicates the third domain or connecting peptide extended third domain (which of the two is meant should be clear from the context). Thus, OMOST3 is ostrich ovomucoid third domain. In cases of polymorphism, both forms are designated by a single letter following the 3 and indicating the varied amino acid; thus, OMMNQ3S is the Ser^{18} form of Montezuma quail ovomucoid third domain. Three numbering systems are employed: entire chicken ovomucoid (see the preceding paper), third domain numbering system [listed on the top and bottom of Table II and first employed by Weber et al. (1981)], and reactive site based system (Schechter & Berger, 1967). These systems are compared in Table VI. In the polymorphism designation, in Table II and in Figure 5, the single Roman capital letters denote amino acid residues. In all other places, three-letter amino acid abbreviations are used.

Table I: Names of Birds Whose Ovomucoid Third Domains Were Sequenced: Sources of Egg Supply and Isolation Data

	common name	Latin name	supplier code[a]	entire ovomucoid yield (mg/mL of egg white)	proteinase used[b]	third domain yield (mg/mL of egg white)	glycosylation[c]
OST	ostrich	*Struthio camelus*	s2 (G)	7	Spase V8	1.4	(−)
REA	greater rhea	*Rhea americana*	s3 (P)	6	Spase V8	0.4	(−)
LER	lesser rhea	*Pterocnemia pennata*	s4 (G)	6	Spase V8	0.4	(−)
ASC	Australian cassowary	*Casuarius casuarius*	s5 (G)	8	Spase V8	0.5	(+)
EMU	emu	*Dromaius novaehollandiae*	s6 (G), s7 (P), s8 (G)	6	thermolysin	0.4	(−)
BTI	brushland tinamou	*Nothoprocta cinerascens*	s3 (P), s9 (G)	4	thermolysin	0.1	(−)
CTI	elegant crested-tinamou	*Euotromia elegans*	s3 (P), s4 (G), s10 (G)	5	thermolysin	0.25	(−)
HPE	Humboldt penguin	*Spheniscus humboldti*	s11 (G)	4	thermolysin	<0.1	(−)
LBC	little black cormorant	*Phalacrocorax sulcirostris*	s5 (G)	2.5	Spase V8	0.2	(−)
ADR	Australian darter	*Anhinga novaehollandiae*	s5 (G)	7.5	Spase V8	0.15	(−)
BNH	black-crowned night-heron	*Nycticorax nycticorax*	s12 (G)	2	thermolysin	<0.1	(−)
MPG	magpie goose	*Anseranas semipalmeta*	s4 (G)	3.5	thermolysin	0.3	(−)
WTD	West Indian tree duck	*Dendrocygna arborea*	s3 (G), s4 (G)	11.5	thermolysin	1	(−)
HTD	whistling tree duck	*Dendrocygna arcuata*	s2 (G), s13 (G)	6	thermolysin	<0.1	(−)
BTD	black-bellied tree duck	*Dendrocygna autumnalis*	s14 (P)	9.5	thermolysin	0.25	(−)
FTD	fulvous tree duck	*Dendrocygna bicolor*	s15 (G)	12	thermolysin	0.7	(−)
PWD	plumed whistling duck[d]	*Dendrocygna eytoni*	s13 (G)	8.5	thermolysin	0.7	(−)
WFD	white-faced tree duck	*Dendrocygna viduata*	s13 (G)	10	thermolysin	0.3	(−)
CSN	coscoroba swan	*Coscoroba coscoroba*	s16 (G)	14	thermolysin	0.6	(−)
SWN	black swan	*Cygnus atratus*	s17 (G)	8	thermolysin	0.7	(−)
MSN	mute swan	*Cygnus olor*	s18 (G)	9	thermolysin	0.7	(−)
GOO	domestic goose[e]	*Anser anser*	s1 (P), s19 (P), s20 (P)	10	thermolysin	1	(−)
BHG	bar-headed goose	*Anser indicus*	s21 (P), s22 (G), s23 (G)	10	thermolysin	0.9	(−)
CNG	Canada goose	*Branta canadensis*	s2 (G), s21 (P)	7.5	thermolysin	0.5	(−)
CBG	Cape Barren goose	*Cereopsis novaehollandiae*	s3 (P)	10	thermolysin	0.9	(−)
UPG	upland goose	*Chloephaga picta*	s22 (G)	6	thermolysin	0.15	(−)
BWG	blue-winged goose	*Cyanochen cyanoptera*	s24 (G)	10	thermolysin	0.5	(−)
ORG	Orinoco goose	*Neochen jubata*	s25 (G)	6	thermolysin	0.5	(−)
WSD	white-headed sheld duck	*Tadoma radjah*	s4 (G), s22 (G)	15	thermolysin	1	(−)
CRD	crested duck	*Lophonetta specularioides*	s2 (G), s26 (G)	8	thermolysin	0.1	(−)
CAW	Cape widgeon	*Anas capensis*	s22 (G)	5	thermolysin	0.2	(−)
DUK	white Pekin duck[e] (mallard duck)	*Anas platyrhynchos*	s17 (G), s21 (P)	10	thermolysin	0.3	(−)
WOD	wood duck	*Aix sponsa*	s4 (G), s10 (G), s20 (P), s21 (P), s22 (G)	10	thermolysin	0.8	(−)
HDM	hooded merganser	*Mergus cucullatus*	s2 (G), s13 (G), s22 (G), s24 (G)	9	thermolysin	0.7	(−)
RUD	ruddy duck	*Oxyura jamaicensis*	s4 (G), s13 (G), s22 (G)	13	thermolysin	1	(−)
LKD	lake duck	*Oxyura vittata*	s13 (G)	11	thermolysin	0.3	(−)
BRK	brahminy kite	*Haliastur* indus	s4 (G)	3	thermolysin	<0.1	(−)
CPV	Cape vulture	*Gyps coprotheres*	s5 (G)	1.5	Spase V8	<0.1	(−)
AHH	African harrier-hawk	*Polyboroides radiatus*	s5 (G)	6	thermolysin	<0.1	(−)
MAH	marsh harrier	*Circus aeruginosus*	s18 (G)	2	thermolysin	0.1	(+)
WTE	wedge-tailed eagle	*Aquila audax*	s5 (G)	4.5	thermolysin	0.1	(−)
CSH	common scrub hen	*Megapodius freycinet*	s5 (G)	1	thermolysin	<0.1	(+)
MLF	Mallee fowl	*Leipoa ocellata*	s5 (G)	12	thermolysin	<0.1	(+)
CHA	plain chachalaca	*Ortalis vetual*	s6 (G)	9	Spase V8	<0.1	(−)
GUA	Spix's guan	*Penelope jacquacu*	s3 (P), s10 (G)	7	Spase V8	0.2	(−)
PGN	common piping guan	*Aburria pipile*	s23 (G)	6	Spase V8	0.1	(−)
RFG	ruffed grouse	*Bonasa umbellus*	s24 (G), s28 (P)	8	Spase V8	0.6	(−)

Table I (Continued)

	common name	Latin name	supplier code[a]	entire ovomucoid yield (mg/mL of egg white)	proteinase used[b]	third domain yield (mg/mL of egg white)	glycosylation[c]
GPC	prairie chicken	*Tympanuchus cupido*	s4 (G), s29 (P)	10	thermolysin	1	(-)
SGG	sage grouse	*Centrocercus urophasianus*	s30 (G)	10	Spase V8	0.3	(-)
MTQ	mountain quail	*Oreortyx pictus*	s31 (P), s32 (P), s33 (P)	10	Spase V8	0.6	(-)
SCQ	scaled quail[f]	*Callipepla squamata*	s6 (G)				
SCQS	blue scaled quail[g]	*Callipepla squamata pallida*	s34 (P), s35 (P), s36 (P), s37 (P)	6	Spase V8	0.2	(-)
SCQN	chestnut bellied scaled quail[g]	*Callipepla squamata castanogastric*	s35 (P), s37 (P)	6.5	Spase V8	0.2	(-)
CAQ	California quail	*Lophortyx californicus*	s6 (G)	7	Spase V8	0.35	(-)
GMQ	Gambel's quail[h]	*Lophortyx gambelii*	s37 (P), s38 (P)	9	Spase V8	0.5	(-)
BWQ	bobwhite quail	*Colinus virginianus*	s39 (P), s40 (P), s41 (P), s42 (G)	4	Spase V8	0.1	(-)
MNQL	Montezuma quail[g]	*Cyrtonyx montezumae*	s43 (G)	5.5	Spase V8	0.2	(-)
MNQS	Montezuma quail[g]	*Cyrtonyx montezumae*	s44 (G, P)	3.5	Space V8	0.1	(-)
CHU	chukar partridge	*Alectoris chukar*	s45 (P)	8	Space V8	0.3	(-)
RPA	red-legged partridge	*Alectoris rufa*	s46 (G)	12	thermolysin	0.2	(-)
BTF	bare-throated francolin	*Francolinus afer*	s3 (P), s4 (G)	5	thermolysin	<0.1	(-)
COF	coqui francolin	*Francolinus coqui*	s46 (G, P), s47 (G)	4.5	thermolysin	0.3	(+)
						<0.1	(-)
ERF	Erckel's francolin	*Francolinus erckelii*	s28 (P), s46 (G, P)	2.5	Spase V8	<0.1	(+)
BPA	black partridge	*Francolinus francolinus*	s3 (P), s46 (P)	9	thermolysin	0.5	(-)
GRF	gray francolin	*Francolinus pondicerianus*	s6 (G), s10 (G), s46 (G)	5	Space V8	0.1	(-)
HPA	gray partridge	*Perdix perdix*	s28 (P), s37 (P), s46 (G)	16	Spase V8	1	(-)
HRQ	harlequin quail	*Coturnix delegorguei*	s32 (P)	5	elastase	0.2	(-)
JPQ	Japanese quail[i]	*Coturnix coturnix japonica*	s48 (P), s49 (P)	3	Spase V8	0.2	(-)
CPA	common hill-partridge	*Arborophila torqueola*	s4 (G)	4	thermolysin	<0.1	(-)
STP	satyr tragopan	*Tragopan satyra*	s10 (G), s50 (P)	6	Spase V8	0.4	(-)
TTP	Temminck's tragopan	*Tragopan temminckii*	s4 (G), s50 (P), s51 (P)	7	Spase V8	0.5	(-)
HMP	Himalayan monal pheasant	*Lophophorus impejanus*	s2 (G), s10 (G), s23 (G), s28 (P), s50 (P), s51 (P)	7	Spase V8	0.6	(-)
BEP	blue eared-pheasant	*Crossoptilon auritum*	s4 (G), s28 (P), s52 (P)	7	Space V8	0.4	(-)
EDP	Edwards's pheasant	*Lophura edwardsi*	s10 (G), s51 (P), s53 (G)	8	Space V8	0.6	(-)
CFP	crested fireback pheasant	*Lophura ignita*	s10 (G)	6	Space V8	<0.1	(-)
						0.1	(+)
SVP	silver pheasant	*Lophura nycthemera*	s37 (P), s42 (G), s54 (G)	9	Space V8	0.7	(-)
						0.7	(+)
CJF	Ceylon junglefowl	*Gallus lafayettii*	s3 (G)	6	Space V8	<0.1	(-)
GJF	gray junglefowl[d]	*Gallus sonneratii*	s3 (G)	7	Space V8	0.1	(-)
KOP	Koklass pheasant	*Pucrasia macrolopha*	s28 (P), s51 (P)	7	Spase V8	0.4	(-)
CHP	cheer pheasant	*Catreus wallichii*	s28 (P), s52 (P)	2	Spase V8	0.1	(-)
RNP	ring-necked pheasant	*Phasianus colchicus*	s42 (G), s45 (P)	8	Spase V8	0.5	(-)
VCP	green pheasant	*Phasianus versicolor*	s20 (P)	6	Spase V8	0.4	(-)
ELP	Elliot's pheasant	*Syrmaticus ellioti*	s55 (P)	14	Spase V8	1	(-)
RVP	Reeves's pheasant	*Syrmaticus reevesii*	s4 (G), s5 (G), s6 (G), s20 (P), s42 (G), s53 (G)	8	Space V8	0.6	(-)
LAP	Lady Amherst's pheasant	*Chrysolophus amherstiae*	s4 (G), s6 (G)	7	Space V8	0.6	(-)
GLP	golden pheasant	*Chrysolophus pictus*	s4 (G), s5 (G), s6 (G), s42 (G), s50 (P)	8	Spase V8	0.7	(-)
GPP	peacock-pheasant	*Polyplectron bicalcaratum*	s3 (G), s13 (G), s50 (P), s56 (P)	3	thermolysin	<0.1	(-)
						<0.1	(+)
MGA	great argus	*Argusianus argus*	s23 (G), s50 (P)	5	Spase V8	0.5	(+)
IPF	peafowl	*Pavo cristatus*	s7 (P), s20 (P), s57 (P)	8	Spase V8	0.4	(-)
						0.5	(+)
GUI	helmet guineafowl	*Numida meleagris*	s58 (P), s59 (P)	5	Spase V8	0.35	(-)
VGU	vulturine guineafowl	*Acryllium vulturinum*	s4 (G)	4.5	Space V8	0.35	(-)
TKY	turkey[e]	*Meleagris gallopavo*	s8 (G), s60 (P), s61 (P)	7	Spase V8	0.6	(-)
						0.7	(+)

Table I (Continued)

	common name	Latin name	supplier code[a]	entire ovomucoid yield (mg/mL of egg white)	proteinase used[b]	third domain yield (mg/mL of egg white)	glycosylation[c]
OTK	ocellated turkey	*Agriocharis ocellata*	s3 (P), s4 (G), s50 (P)	4	Spase V8	<0.1	(−)
BUQ	little button-quail	*Turnix sylvatica*	s32 (P), s62 (P), s63 (P)	6	thermolysin	0.15	(+)
WTC	wattled crane	*Grus carunculatus*	s10 (G)	6	Space V8	<0.1	(+)
WNC	white-naped crane	*Grus vipio*	s27 (G)	5	Space V8	0.15	(+)
DEC	demoiselle crane	*Anthropoides virgo*	s24 (G)	7	Space V8	0.1	(+)
COO	coot	*Fulica atra*	s18 (G)	9	Space V8	0.1	(−)
SPL	spur-winged lapwing	*Vanellus spinosus*	s2 (G), s11 (G), s13 (G)	2	Spase V8	<0.1	(−)
BGL	black-headed gull	*Larus ridibundus*	s18 (G)	6	thermolysin	0.1	(+)
RDR	roadrunner	*Geococcyx californianus*	s4 (G)	1	thermolysin	<0.1	(+)
CGC	coral-billed ground-cuckoo	*Carpococcyx renauldi*	s4 (G)	2	Space V8	<0.1	(+)
LGK	laughing kookaburra	*Dacelo novaeguineae*	s10 (G)	9	thermolysin	<0.1	(+)

[a] A list of suppliers is given in the Appendix. (G) indicates a gift; (P) indicates a purchase. [b] Spase V8, *Staphylococcus aureus* proteinase V8; elastase, porcine pancreatic elastase I. [c] The domain we isolated is (−) unglycosylated or (+) glycosylated. [d] Ovomucoid third domains obtained from this sample is polymorphic. [e] Gruson (1976) does not list domesticated species. [f] Obtained as a 2:1 mixture of *C. squamata pallida* (SCQ3S) and *C. squamata castanogastrix* (SCQ3N) from Prof. A. C. Wilson. [g] See Polymorphism in the text. [h] Former abbreviation was GAM (Kato et al., 1978). [i] All samples are polymorphic (Bogard et al., 1980).

et al. (1986)]; the presence of the carbohydrate chain causes its coelution with the 56 residue long third domain.

Sometimes we generated only a single new protein peak as a result of limited proteolysis of various ovomucoids. In 12 cases, only the glycosylated form of ovomucoid third domain could be obtained (Figure 1B); in the single case of ostrich ovomucoid (Figure 1C), only unglycosylated third domain was found in the digest. Usually, we continued the limited proteolysis until no further increase of the third domain peak(s) was detected in the aliquots of the incubation mixture. Such judgment may be sometimes improper because of the appearance of materials other than the third domain under the presumed third domain peak. The reaction was stopped by acidification of the mixture to a pH of about 2 with formic acid. The third domain was isolated from the digest by size-exclusion chromatography (Figure 2A) and was further purified by ion-exchange chromatography (Figure 2B,C). The resulting preparation was desalted on a Sephadex G-15 column in 0.05 M ammonium bicarbonate or in 1% acetic acid and lyophilized.

Amino Acid Analyses. Aliquots (~50 μg) of all newly prepared connecting peptide extended ovomucoid third domains were hydrolyzed in 6 N HCl at 110 °C (most frequently for 20 h although for some samples times were varied and for some glycosylated samples shorter times were employed). The results of the hydrolyses were highly satisfactory for nonglycosylated samples and less satisfactory for glycosylated ones where the losses of Ser, Thr, Cys, and sometimes Tyr were large, and the degradation products tended to increase the Pro and Ala peaks. The hydrolysates were analyzed on a Durrum D-500 amino acid analyzer and the results converted to residues per molecule by an integer-fit program based on Hoy et al. (1974). The objectives of these amino acid analyses were 3-fold: (a) to determine whether the material is indeed a third domain and to ascertain that the sample was pure enough to merit sequencing, (b) to serve as an aid in guessing the appropriate sequencing program for each position, and (c) to compare to the composition determined by sequence analysis. Unless an exact agreement was found (allowances were made for the amino acid residues mentioned above), either amino acid analysis, sequencing, or both were repeated until an exact agreement was reached. The need to make such repeats was small. Amino acid analysis is also routinely used to test new batches of already sequenced third domains prior to any further studies on our sequence to reactivity algorithm. This is essential. For example, a difference was detected by amino acid analysis in the second batch of OMMNQ3; polymorphism was suspected and later confirmed by sequencing. Thus far, we have not encountered any third domains that have identical amino acid composition but differ in sequence except in Asn → Asp differences [e.g., OMGLP3–OMHPA3 (see below)].

Sequencing. As this project lasted over a decade, the method of sequencing and identification of the phenylthiohydantoin derivatives of the amino acids changed dramatically. Only the current method is described. We believe after extensive checking of all the sequence data files that all of the older data are of the same quality as the current data. In all cases of doubt (10 sequences), we have repeated the work in 1985. In none of the cases have we found discrepancies. However, during this rechecking period one typographical error was found.

Our Beckman-890C automatic sequencer is equipped with a Sequemat P-6 autoconverter that employs a methanolic conversion program suggested by the manufacturers and with a Sequemat SC-510 programmer. Detection of the phenylthiohydantoin derivatives of the amino acids is by a local modification of the method of Lottspeich (1980). The high-performance liquid chromatograph is by Waters (M6000-A pump and 440 detector at 254 nm); the column is Waters "Resolve" C_{18}, 15 cm. The automatic injector is from a Durrum D500 amino acid analyzer and was obtained from the company as a specialty item. The data acquisition and injector control are by a Hewlett-Packard 3385A automation system.

The approach of this laboratory differs from that of many of our colleagues because our sequencing goals are different. The samples we sequence tend to be highly similar, and since we need to have at least 1 mg of sample for later use in sequence to reactivity studies, we adopted an approach that stresses (1) high reliability of data, (2) very long and reliable reach of sequencing runs, and (3) high utilization of equipment and automation. As a result, in the last 40 attempts to sequence third domains (51–56 residues long) we got to the end 36 times. All four of the failures involved relatively impure samples.

Only seven of the sequences were run on reduced and carboxymethylated proteins. They were OMBWQ3,

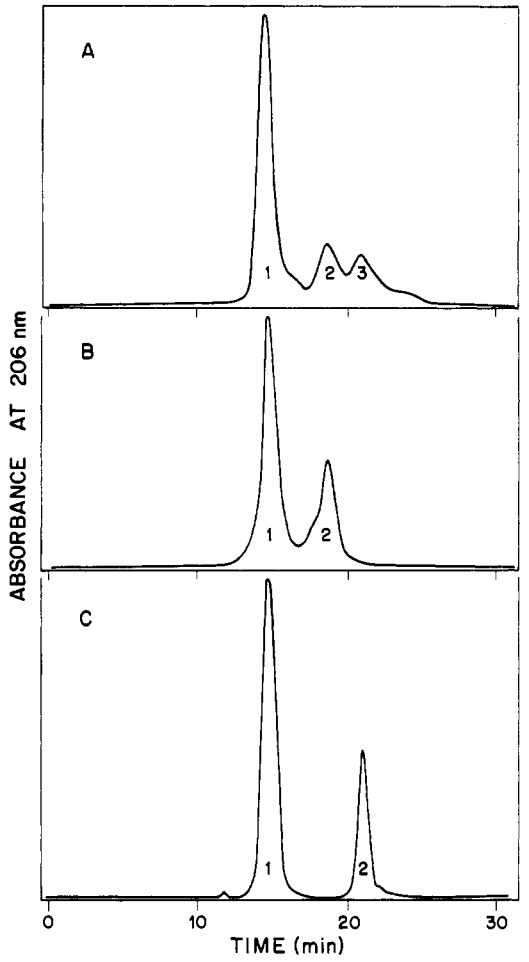


FIGURE 1: Limited proteolysis of avian ovomucoids as traced by size-exclusion HPLC. Entire ovomucoids of Gambel's quail (Panel A), great argus (panel B), of ostrich (panel C) were treated with *Staphylococcus aureus* proteinase V8 for 48 h (see text for details). Aliquots of the incubation mixtures corresponding to 30–40 μg of the initial ovomucoid were run on a Varian Micropak column, grade G2000SW (Toya-Soda column), in 0.2 M sodium phosphate buffer, pH 6.5. The runs were monitored at 206 nm. Peak 1 (panels A–C) contains mainly undigested ovomucoid and ovomucoid from which the third domain has been removed, as well as the proteinase. Peak 2 (panels A and B) contains mainly ovomucoid third domain, glycosylated; peak 2 (panel C) and peak 3 (panel A) contain mainly ovomucoid third domain, unglycosylated (see text for more detailed information).

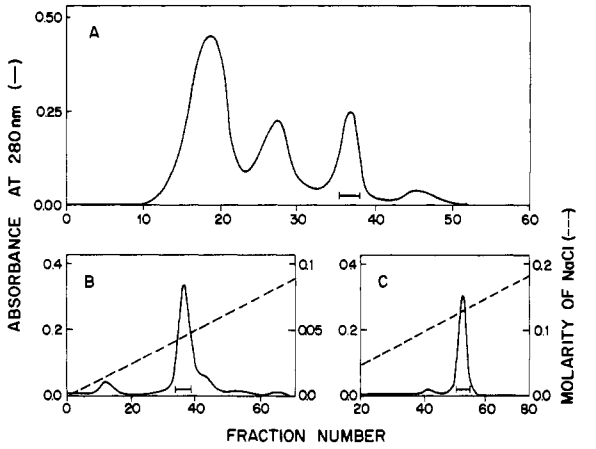


FIGURE 2: Isolation and purification of Gambel's quail ovomucoid third domain. The entire ovomucoid was treated with Staphylococcal proteinase V8 for 48 h (see text), and the unglycosylated third domain was isolated from the incubation mixture in three consecutive chromatographic steps. (A) Size-exclusion chromatography on Bio-Gel P-10. The incubation mixture was introduced into the column (2 × 92 cm) in 5% formic acid and run (20 mL/h) with continuous monitoring at 280 nm. Seven-milliliter fractions were collected. The fractions containing unglycosylated third domain (indicated by a solid bar) were combined and freeze-dried. (B) DEAE-Sepharose CL-6B chromatography. Crude third domain isolated in the previous step was dissolved in 0.04 M Tris-HCl buffer, pH 8.8, and loaded onto the column (2.5 × 9 cm) equilibrated with the same buffer. The column was developed (50 mL/h, 5-mL fractions) with a 0–0.1 M sodium chloride gradient (total volume 400 mL) in the Tris buffer. The solid bar indicates the fractions saved. (C) CM-Sepharose CL-6B chromatography. The fractions isolated in the previous step were combined, diluted 4 times with water, and acidified to pH 4.0 with acetic acid. The solution was then loaded onto the column (2.5 × 9 cm) in 0.05 M sodium acetate buffer, pH 4.0. The column was developed (50 mL/h, 5-mL fractions) with a linear sodium chloride gradient (0–0.2 M, total volume 400 mL). Fractions indicated by the solid bar were isolated.

OMCHI3, OMMTQ3, OMSCQ3, OMJPQ3, OMTKY3, and OMDUK3. The remaining samples were run with disulfide bonds intact. There was no consistent difference in repetitive yield or quality of the runs except that in the disulfide-intact runs the Cys positions were blank. The only other blank position was Asn[45], the glycosylation position, when glycosylated third domains were run. The Cys locations were identified by (1) being blank, (2) homology, and (3) amino acid analysis.

The amount of sample used per run is about 500 μg (50–80 nmol). Polybrene (Tarr et al., 1978; Klapper et al., 1978; Hunkapiller & Hood, 1978) is used in all runs to reduce sample washout. Four sequencer programs are used: a Startup program and three different programs based on Beckman Quadrol program 123974, which differ from one another only by the length of time devoted to cleavage (10, 60, or 400 s). The objective of the program variations is to reduce the total cleavage time and thus background; only Pro residues appear to require long cleavage time. The middle time is used for Gly, Ile, Val, and glycosylated Asn and the short time for all other residues. As is seen in Tables VI–VIII, Pro is exceptionally well conserved in ovomucoid third domains, and its positions in Table II can be anticipated in advance. However, the length of the connecting peptide fragment varies erratically (see Table II) and is, therefore, inferred from amino acid analysis prior to the run. In the relatively rare cases where the third domain contains four Pro residues, we occasionally misguess the fourth position. In all such cases enough information was obtained from the first misprogrammed run to program a second run correctly and to identify all of the residues.

Once a third domain sequence is determined, it is entered into Purdue University Computer Center memory files via a microcomputer in our laboratory. A program for doing this was originally written by W. R. Finkenstadt. Once a sequence was entered, it was never retyped or rewritten without computer assistance and thus both Tables II and IV as well as many alignments that were tried on these data were all computer generated. We feel that this is clearly needed to decrease the probability of typographical errors.

Other Characterization Techniques. Three analytical chromatographic techniques were used in this study for the characterization of third domains with respect to purity, polymorphism, glycosylation, or molecular charge. These were: size-exclusion HPLC, reverse-phase HPLC, and analytical ion-exchange liquid chromatography.

Reverse-phase HPLC was performed according to Lin et al. (1985) on a Waters μBondapak C_{18} column with a binary gradient (water, trifluoroacetic acid, and 2-propanol). The runs were monitored at 214 nm. Since this technique can

Table II: Amino Acid Sequences of 106 Ovomucoid Third Domains (100 Species plus Chicken from Preceding Paper plus 5 Polymorphisms)[a]

	1 2 3 4 5 5A	6 7 8 9 10 11 12 13 14 15 16 17 18	19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56
OMOST3	F A T - - -	V D C S D Y P K P V C P L	D Y M P L C G S D S K T Y S N K C N F C N A V V E S S G T L T L R H F G K C
OMREA3	F A T - - -	V D C S D H 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 E S N G T L T L S H F G K C
OMLER3	F A T - - -	V D C S D H 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 E S N G T L T L S H F G K C
OMASC3	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
OMEMU3	F A T - - -	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 N N K C D F C N A V V E S N G T L T L G H F G K C
OMBTI3	V T - - - -	V D C S G Y P K P A C T L	E Y F P L C G S D N Q T Y A N K C A F C N A V V E K N V - - T L R H L G K C
OMCTI3	V S - - - -	I D C S G Y P K P A C T L	E 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 N H I G E C
OMHPE3	I A T - - -	I 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 N F C N A V V D S N G T L I L S H F G K C
OMLBC3	I A T - - -	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 N F C N A V A D S N G T L T L S H F G K C
OMADR3	I A S - - -	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 N F C N A V V D S N G T L T L S H F G K C
OMBNH3	I V T - - -	V D C S D Y P K P V C S P	E N M P V C G S D S K T Y S N K C D F C N A V A D S N G T L T L S H F G K C
OMMPG3	V A T - - -	V D C S E 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 N F C N A V V D S N G T L T L S H F G K C
OMHTD3	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 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
OMHTD3	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 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
OMBTD3	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 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
OMFTD3	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 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
OMPWD3D	V A T - - -	V D C S D Y P R P D C T L	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
OMPWD3E	V A T - - -	V D C S E Y P R P D C T L	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
OMWFD3	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 K T Y G N R C N F C N A V V D S N G T L T L S H F G K C
OMCSN3	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 K C
OMSWN3	V A T - - -	V D C S D 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 K C
OMMSN3	V A T - - -	V D C S D 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 K C
OMGOO3	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 D N K C N F C N A V V D S N G T L T L S H F G K C
OMBHG3	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 D N K C N F C N A V V D S N G T L T L S H F G K C
OMCNG3	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
OMCBG3	V A T - - -	V D C S D Y P K P A C R 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 K C
OMUPG3	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
OMBWG3	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
OMORG3	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
OMWSD3	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
OMCRD3	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
OMCAW3	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
OMDUK3	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
OMWOD3	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
OMHDM3	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
OMRUD3	V A T - - -	V D C S D 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 K C
OMLKD3	V A T - - -	V D C S D 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 K C
OMBRK3	I T 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 N L S H F G K C
OMCPV3	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
OMAHH3	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
OMMAH3	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 F V D S N G T L S L S H F G K C
OMWTE3	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
OMCSH3	L V T - - -	V D C S D Y P K P S C T L	E D K 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
OMMLF3	V V T - - -	V D C S G Y P T H A C T L	E L K P L C G S D N Q T Y S N K C G F C N A V A Q S N G T L T L S H F G K C
OMCHA3	F A P V N -	V D C S D H P K P A C L Q	E Q K P I 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 T L S H F G K C
OMGUA3	F A P V N -	V D C S D H P K P A C L Q	E Q K 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 T L S H F G K C
OMPGN3	L A P V N -	V D C S D H P K P A C L Q	E Q K 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 T L S H F G K C
OMRFG3	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 N L S H F G T C
OMGPC3	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
OMSGG3	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
OMMTQ3	F 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 S K T Y A N K C N F C N A V V E S N G T L T L S H F G K C
OMSCQ3S	F A A V S -	V D C S E Y P K P D C T L	E Y R P L C G S D N K T Y A S K C N F C N A V V E S N G T L T L S H F G K C
OMSCQ3N	F A A V S -	V D C S E Y P K P D C T L	E Y R P L C G S D N K T Y A N K C N F C N A V V E S N G T L T L S H F G K C
OMCAQ3	F 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 A N K C N F C N A V V E S N G T L T L S H F G K C
OMGMQ3	F 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 A N K C N F C N A V V E S N G T L T L S H F G K C
OMBUQ3	F A A V S -	V D C S E Y P K P A C M S	E Y R P L C G S D N K T Y V N K C N F C N A V V E S N G T L T L S H F G K C
OMMNQ3L	F A A V S -	V D C S E Y P K P A C T L	E Y V P I 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
OMMNQ3S	F A A V S -	V D C S E Y P K P A C T S	E Y V P I 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
OMCHU3	L A A - - -	V D C S E Y P K P A C T L	E Y R P L C G S D S 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
OMRPA3	L A A - - -	V D C S E Y P K P A C T L	E Y R P L C G S D S 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
OMBTF3	V S -	V D C S E Y P K P A C T M	E Y R P V C G S D N I T Y G N K C N F C N A V V K S N G T L T L S H F G K C
OMCOF3	V S -	V D C S E Y P K P G C T M	E Y R P V C G S D N I T Y G N K C N F C N A V V K S N G T L T L S H F G K C
OMERF3	P A A V S -	V D C S E Y P K P A C T M	E Y R P V C G S D N I T Y G N K C N F C N A V V K S N G T L T L S H F G K C
OMBPA3	V S -	V D C S G Y P K P E C T A	E E T P L C G S D N K T Y R N K C N F C N A V V E S N G T L T L R H F G K C
OMGRF3	L P A V S -	V D C S E Y P K P D C T T	E E 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
OMHPA3	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 D F C N A V V E S N G T L T L N H F G K C
OMHRQ3	S -	V D C S E Y P K P A C P K	D Y R P V 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 N R F G K C
OMJPQ3S	L A A V S -	V D C S E Y P K P A C P K	D Y 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 L N H F G K C
OMJPQ3G	L A A V S -	V D C S E Y P K P A C P K	D Y R P V 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 N H F G K C
OMCPA3	V S -	V D C S D Y P M P A C T M	E Y R P V 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
OMSTP3	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
OMTTP3	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
OMHMP3	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
OMBEP3	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
OMEDP3	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
OMCFP3	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 S Y D N K C N F C N A V V E S N G T L T L S H F G K C
OMSVP3	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
OMCHI3	L A A V S -	V D C S E Y P K P D C T A	E D 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
OMCJF3	L A A V S -	V D C S E Y P K P D C T A	E D 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
OMGJF3A	L A A V S -	V D C S E Y P K P D C T A	E D 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
OMGJF3V	L A A V S -	V D C S E Y P K P D C T V	E D 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
OMKOP3	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
OMCHP3	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
OMRNP3	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 N R F G Q C
OMVCP3	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 N R F G Q C
OMELP3	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
OMRVP3	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 R F G E C
OMLAP3	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 N H F G K C
OMGLP3	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 N H F G K C
OMGPP3	V S -	V D C S E Y P Q P T C T T	E H R P V C G S N N E T Y G N K C N F C N A V V K S N G T L T V S H F G K C
OMMGA3	L A A - - -	V D C S E H P K P A C T L	E D R P L C G S D N K I Y S N K C D F C N A V L E S N G T L T L S H F G K C
OMIPF3	L A A V S -	V D C S E Y P K P A C T L	E H 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
OMGUI3	L A A - - -	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 D N K C N F C N A V V E S N G T L T L S H F G K C
OMVGU3	L A A - - -	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 D N K C N F C N A V V E S N G T L T L S H F G K C
OMTKY3	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
OMOTK3	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
OMBUQ3	V S - - -	V D C S E Y P K P G C M M	E R L P L C G S D N K T Y N D K C N F C N A V V E S N G T L T L N H F G E C
OMWTC3	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
OMWNC3	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
OMDEC3	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
OMCOO3	T A T - - -	V D C S G Y P Q P A C S L	E Y V P F C G S D N K T Y S N K C D F C N A V A D S N G T L T L S H F E K C
OMSPL3	I A T - - -	V D C S D Y P K P A C S L	D Y M P L 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 E K C
OMBGL3	I A T - - -	V D C S D Y P K P A C S L	D Y M P L 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 E K C
OMRDR3	I A S - - -	V N C S D Y P K P V C S L	L Y M P L C G S D N K T Y G N K C N F C N A V A D S N G T L T L S H F G K C
OMCGC3	I A T - - -	V D C S D Y P K P V C T L	E D M P L C G S D N I T Y H N K C Y F C N A V A H S N G T L T F S H F G K C
OMLGK3	I A T - - Q	V D C S E H P K P A C T L	D Y R P I C G S D S K T Y S N K C D F C N A V M D S N G T L T L S H F G K C

[a] For abbreviations of bird names, see Table I. Three numbering systems can be used for these sequences; they are compared in Table VI; the one used here is the third domain system. The left-hand side vertical line divides the connecting peptide fragment (residues 1–5A) from ovomucoid third domain proper (residues 6–56). The alignment of connecting peptide fragments for OMBT13, OMCTI3, and OMBUQ3 is arbitrary. The second vertical line near the center divides the P_1 residue from the P'_1 residue of the reactive site.

separate proteins differing by a single amino acid residue, it is very useful for studies of polymorphism. It was used in this work for resolving several polymorphic third domain pairs, e.g., plumed whistling duck (PWD), Japanese quail (JPQ), and gray junglefowl (GJF) representing Glu/Asp, Ser/Gly, and Ala/Val polymorphisms, respectively, as well as for separating domains with the Val^4-Ser^5 deletion [see Connecting Peptide Fragment (Residues 1–5A) and see Polymorphisms].

A modification of the method of Ardelt and Laskowski (1982) was used for analytical ion-exchange chromatography. Protein samples (2–5 μg) were run on either Mono Q or Mono S columns in the Fast Protein Liquid Chromatography System of Pharmacia at the flow rate of 72 mL/h. The runs were monitored at 206 nm (LKB 2138 Uvicord S detector). The Mono Q column was developed isocratically in 0.02 M ammonium chloride buffer, pH 9.1, which contained 0–0.1 M NaCl depending on the overall charge of the third domain. Occasionally, a linear NaCl gradient was used. The Mono S column was developed in 0.02 M sodium acetate buffer, pH 3.5–4.0, containing 0–0.2 M NaCl. This technique allowed us "down to the base line" separation of third domain variants different by only one unit charge. In addition to its use as a purity check, we could, by calculating the charge from the sequence, use it as a check of sequence accuracy. The chromatographic behavior of 44 of the 45 third domains tested here was in agreement with our expectations, i.e., consistent with the sequences. The detailed report on this study will be given in another publication.

Results and Discussion

Table I lists the abbreviations and names for all the species of birds employed in this work, the suppliers of their eggs, the yield of ovomucoid, the enzyme used for proteolysis, and the yield of connecting peptide fragment extended third domain. Table II lists the sequences of the connecting peptide fragment extended third domains. These two tables are the results of this paper; the remainder is commentary, analysis of data, and discussion.

Sequence Reliability. Since the proteins whose preparation and sequencing are reported here are being used by many laboratories in sequence to reactivity studies, great effort was made to ensure that the sequences are correct. The question of sequence reliability, however, consists of two parts, and our level of confidence about these parts differs greatly.

(*1*) *Are the Sequences Those of the Substances Put in the Sequencer*? While all scientific procedures involve experimental errors, we believe that there are very few, if any, errors in the data given in Table II. The straight-through sequencing method completely eliminates any errors due to peptide alignment. The domains are quite small and, in the case of carbohydrate-free third domains, so highly purified that heterogeneities involving even a single amino acid residue (polymorphisms, presence and absence of the Val^4-Ser^5 dipeptide, and enzymatic jagged cuts) can be chromatographically detected. The agreement between the sequence and amino acid analysis is a very strong criterion for correctness of the sequence. Since we have so many homologous sequences, any aberrant substitution is originally doubted and especially strongly scrutinized. Three of the third domains, OMJPQ3G and OMJPQ3S mixture (Weber et al., 1981; Papamokos et al., 1982), OMSVP3 (Bode et al., 1985), and OMTKY3 in complex with SGPB (Fujinaga et al., 1982; Read et al., 1983), were subject to high-resolution crystallographic studies. In each case the sequence was verified. Several third domains were subjected to comparative nuclear magnetic resonance studies, and again, no error in sequence was detected. Enzyme–inhibitor equilibrium constants, K_a, were determined for most of the third domains studied here, and the results seem to form a self-consistent set (Empie & Laskowski, 1982; Laskowski et al., 1983; Park, 1985). The sequences are also consistent with chromatographic behavior (see Other Characterization Techniques).

Some of the sequences listed here were published without experimental details before. All those later found to contain errors were labeled preliminary. In Kato et al. (1976), the entire sequence of Japanese quail ovomucoid is given. The third domain sequence given there differs from that reported in Table II by (old → new) $Glx^{10} \rightarrow Glu^{10}$, $Asp^{28} \rightarrow Asn^{28}$, and $Ser^{32} \rightarrow Ser^{32}/Gly^{32}$ mixture, i.e., polymorphism. In Kato et al. (1978), sequences of 17 avian ovomucoid third domains are listed. These were not only labeled as preliminary, but most of the errors given below occur in parentheses. They are as follows: OMJPQ3, polymorphism is still not explicitly recognized but the OMJPQ3G rather than OMJPQ3S form is listed; OMCAQ3, $Leu^1 \rightarrow Phe^1$; OMGMQ3, $Leu^1 \rightarrow Phe^1$; OMSCQ3, the polymorphism was not yet recognized and only the major form OMSCQ3S was listed, also $Asn^{51} \rightarrow Ser^{51}$; OMRNP3, $Glx^{55} \rightarrow Gln^{55}$; OMRVP3, $Glx^{55} \rightarrow Glu^{55}$; OMDUK3, $Glx^{55} \rightarrow Glu^{55}$; OMCHA3, $Leu^1 \rightarrow Phe^1$, and $Phe^{50} \rightarrow Leu^{50}$; OMEMU3, $Glu^{10} \rightarrow Asp^{10}$ and $Asn^{36} \rightarrow Asp^{36}$. These are mostly corrections of ambiguities caused by incorrect identification of phenylthiohydantoins of amino acids by gas chromatography, e.g., Glx → Gln or Glx → Glu, or errors arising from an incorrect inference in the COOH-terminal part made on the basis of homology and amino acid analysis. Neither of these techniques is now used in our laboratory. More recently we have published a set of 17 sequences in Empie and Laskowski (1982). Several of them were extensively rechecked since then, and all are in complete accord with what is listed here. Similarly, we believe that all of the partial sequences listed in our various papers on sequence to reactivity algorithm, three-dimensional structure determination, nuclear magnetic resonance, etc. are consistent with what is given here.

(*2*) *Were Egg Whites Correctly Labeled with the Bird's Name*? Here our confidence is lower as we are at the total mercy of our suppliers, who occasionally have serious problems with the correct identification (see Table I).

Connecting Peptide Fragment (Residues 1–5A). Avian ovomucoids consist of three homologous, tandem Kazal family domains. Domain 1 is connected to domain 2 by the first connecting peptide; domain 2 is connected to domain 3 by the second connecting peptide. Alternately, we can speak about ovomucoid consisting of just the three domains and thus incorporate the connecting peptides into the domains. In either case, the exact boundaries between the domains or between domains and connecting peptides are somewhat arbitrary. The work of Stein et al. (1980) on the structure of the chicken ovomucoid gene provided an additional criterion, and in the preceding paper (Kato et al., 1986) we have decided to call the translated product of the last two exons the third domain. This corresponds to residues 136–186 in chicken ovomucoid or to residues 6–56 in the notation used in Table II and in the rest of this paper. Our work would certainly be simpler if we could for every ovomucoid hydrolyze the bond between residues 5 and 6 and thus liberate the third domain proper. We have not succeeded in doing that with any ovomucoid/proteinase combination. Instead, in every case we obtain a larger fragment that, in laboratory slang, we call "the third domain". It is, in fact, "connecting peptide fragment extended third domain", and it is a set of such sequences that is reported in

Table II. For our application in correlating inhibitory activity with amino acid sequence, the connecting peptide fragments are only a nuisance. We have a good deal of strong evidence that they exert no effect on enzyme–inhibitor association constants. Furthermore, the X-ray crystallographic studies of two free "connecting peptide fragment extended ovomucoid third domains", OMJPQ3 (Weber et al., 1981; Papamokos et al., 1982) and OMSVP3 (Bode et al., 1985), show that the connecting peptide fragment is located far from the residues involved in enzyme–inhibitor contact. In both of the complexes of OMTKY3 with enzymes (Fujinaga et al., 1982; Read et al., 1983, 1984) the connecting peptide fragment is only barely visible.

As can be seen in Table II, the connecting peptides vary in length. The three known causes of this variation are (1) the specificity of the enzyme used for cutting, (2) the presence or absence of the dipeptide Val[4]-Ser[5] (or Val[4]-Asn[5]), and (3) the presence or absence of Gln[5A].

(1) Enzyme Used for Hydrolysis. This was already considered in Kato et al. (1978). The single case where porcine pancreatic elastase I was tried produced hydrolysis of the putative Val[4]–Ser[5] bond in OMHRQ and yielded Ser[5]–third domain. Spase V8 hydrolyzes after Glu[130] (in entire chicken ovomucoid numbering system). The Glu[130] residue is highly conserved, and thus all Spase V8 generated third domains start at residue 131 (residue 1 in the third domain numbering). Thermolysin acting on known sequences with long connecting peptides cuts primarily at Ala[133] to produce ...Val[4]-Ser[5] starts but also at Glu[130]-Leu[131] and at Ala[132]-Ala[133]. In the cases of CTI, BTI, and BUQ we do not have any information about the second domain sequence; in these cases the decision about placement of the connecting peptide residues was arbitrary.

(2) Ambiguous Exon VI/Intron F Junction. In the preceding paper [Kato et al. (1986), especially in Figure 2) the deletion polymorphism involving the dipeptide Val[4]-Ser[5] is described for chicken, a phasianoid bird. In Table II the dipeptide Val[4]-Ser[5] is present in most of the phasianoid sequences listed. It is missing in five phasianoids: OMCHU3, OMRPA3, OMGUI3, OMVGU3, and OMMGA3. CHU and RPA are congeneric (see Table I); their third domain sequences are identical, and their first domain sequences (unpublished) differ only by a single replacement. It seems clear that the Val[4]-Ser[5] deletion occurred in their common ancestor. A similar argument can be made for the pair OMGUI3–OMVGU3, which again show the same third domain sequences and differ by a single replacement in their first domains. On the other hand, neither sequence data nor ornithology suggest a particularly close relationship (within the superfamily Phasianoidea) of the sets OMCHU3–OMRPA3, OMGUI3–OMVGU3, and OMMGA3. It is probable that the mutations that led to Val[4]-Ser[5] deletion in these sets occurred independently. In view of the explanation advanced for the Val[4]-Ser[5] polymorphism by Stein et al. (1980), it is easy to imagine that the deletion could arise by a single nucleotide mutation within the F intron.

In most of the remaining phasianoid species where the Val[4]-Ser[5] sequence is present, it appears to be polymorphic; a majority of the molecules (100–70%) contain the dipeptide sequence. The fraction with Val[4]-Ser[5] varies not only from species to species but from preparation to preparation. Since our yields of connecting peptide extended third domains are often quite low (see Table I), the variation in the fraction of molecules with Val[4]-Ser[5] deleted may be a consequence of biasing due to differential enzymatic digestion or due to differential isolation. The problem appears amenable to study

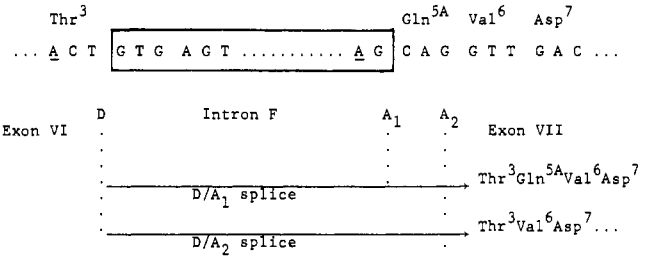


FIGURE 3: Probable mechanism for deletion polymorphism in laughing kookaburra ovomucoid, OMLGK3 (compare to Figure 2 in preceding paper). Adapted from Stein et al. (1980) except that alternative acceptor sites rather than donor sites are involved. The nucleotide sequence is that of chicken (Stein et al., 1980) except that two mutations (underlined) were postulated. At the 5′ end we have ...*A*CT rather than ...*G*CT to account for the change from Ala[3] in chicken to Thr[3] in laughing kookaburra and *T* in chicken to *A* in LGK to produce a minimal concensus splice site in LGK. It is possible that a greater number of mutations are actually required to account for our result. This scheme seems more plausible since the last three nucleotides of the F intron in chicken (Stein et al., 1980) are in fact a codon for Gln[5A].

since the connecting peptide extended third domains with Val[4]-Ser[5] present and Val[4]-Ser[5] missing can be separated by reverse-phase HPLC (Lin et al., 1985).

All three cracid bird sequences in Table II, OMCHA3, OMGUA3, and OMPGN3, contain the sequence Val[4]-Asn[5] (Asn[5] may have arisen from Ser[5] by a single nucleotide mutation). This may indicate that the cracids are closely related to the phasianoids. Ornithology puts cracids and phasianoids into the single order Galliformes, but this placement has been questioned on the basis of amino acid sequences of lysozymes (Prager & Wilson, 1976). On the other hand, the megapodes, sometimes classified as close relatives of the cracids, do not have the Val[4]-Ser[5] sequence or its Val[4]-Asn[5] analogue.

All other orders of birds appear not to have the dipeptide Val[4]-Ser[5]. This conclusion seems quite firm for all sequences except three—OMCTI3, OMBTI3, and OMBUQ3.

(3) Possible Ambiguous Boundary between Intron F and Exon VII. The laughing kookaburra ovomucoid third domain, OMLGK3, is the only one that has a residue at a position we have chosen to name 5A. The Gln at this position is present in 91% of the OMLGK3 molecules we isolated. It seems difficult to explain this polymorphism as a direct extension of the Stein et al. (1980) model (Figure 2 of the preceding paper) involving alternative donors. Instead, in Figure 3 we propose that alternative acceptors explain the phenomenon seen in laughing kookaburra ovomucoid.

Ovomucoid Third Domain Proper (Residues 6–56). (1) Glycosylation. To the best of our knowledge, avian ovomucoids are glycosylated only at Asn residues. The Asn-X-Thr/Ser sequence that is a necessary but not sufficient condition for glycosylation occurs in Table II at three different positions. The first two are not glycosylated. They are ...Asn[6]-Cys[7]-Ser[8]..., which occurs only in OMRDR3, and ...Asn[28]-Lys[29]-Thr[30]..., which occurs in most sequences. On the other hand ...Asn[45]-Gly[46]-Thr[47]... occurs in 103/106 sequences; the two tinamous have ...Asn[45]-Val[46]——Thr[49]..., which still serves as a glycosylation signal. OMOST3 has Ser[45]-Gly[46]-Thr[47] and is not glycosylated. All the others are partially or completely glycosylated. Whenever we could, we isolated the non-glycosylated domain as it is easier to characterize and to crystallize. In some cases (see Table I), both (−) and (+) forms were isolated, giving us the opportunity of making comparisons and of using the (+) domains in competitive assays (Empie & Laskowski, 1982; Wieczorek & Laskowski, 1983). Thus far, all properties other than the

chromatographic behavior appear to be the same for (−) and (+) domains.

(*2*) *Polymorphisms.* Among the third domain sequences from 101 species given in Table II, we found 5 to be polymorphic.[1] They are plumed whistling duck (OMPWD3D and OMPWD3E), scaled quail (OMSCQ3S and OMSCQ3N), Montezuma quail (OMMNQ3L and OMMNQ3S), Japanese quail (OMJPQ3S and OMJPQ3G), and gray junglefowl (OMGJF3A and OMGJF3V). As expected in each polymorphism, all of which are consistent with two allelic genes, there is only one replacement in the third domain, and each replacement is consistent with a single nucleotide change. Interestingly, two of the five polymorphisms (MNQ and GJF) occur at the P_1 (18) position (one of the two most variable positions in the molecule); one (JPQ) occurs at position P'_{14} (32) (third most variable position). All three of the above changes are "contact" positions in enzyme–inhibitor complexes.

The plumed whistling duck polymorphism, P_9 (10) Asp/Glu, was detected in a single batch of third domain obtained from a mixture of several eggs donated by Prof. S. DeKloet, Mergus Farms. The Asp to Glu ratio in this preparation was approximately 4:1.

The scaled quail third domain sequence on material obtained from Prof. A. Wilson showed predominantly Ser at position P'_{15} (33), but there was an indication of Asn as well. In our preliminary publication (Kato et al., 1978), we listed only Ser, but on rechecking a 70% Ser–30% Asn distribution was obtained. Prof. Wilson then informed us that his scaled quail material was a 2:1 mixture of two subspecies, blue scaled quail *Callipepla squamata pallida* and chestnut bellied scaled quail *Callipepla squamata castanogastris*. These two subspecies are separately available from breeders. When we obtained the material, we found *C. squamata pallida* to have only P'_{15} Ser and *C. squamata castanogastris* to have P'_{15} Asn. Thus, the P'_{15} Ser/Asn difference is a true breeding characteristic of the two subspecies.

The Montezuma quail sequence obtained from material given to us by R. Tybie (six infertile eggs, heavily incubated, egg white and yolks mixed) yielded the OMMNQ3 sequence with Leu at the P_1 position. The sequence is consistent with amino acid analysis and K_a studies on this material (Park, 1985). A gift of three infertile eggs from D. Gue yielded OMMNQ3S. Again the sequence, amino acid analysis, and K_a results (unpublished) are consistent with P_1 Ser. The sequences are otherwise identical with one another but differ by at least two amino acid replacements from any other North American quail. Furthermore, one of these replacements surprisingly requires two nucleotide changes (P'_3 Arg → Val). Neither R. Tybie nor D. Gue could help us in identifying the subspecies of Montezuma quail that they have.

The Japanese quail polymorphism, ca. 70% Ser–30% Gly at P'_{14} position, was the subject of extensive study (Bogard et al., 1980). In contrast to the two cases described above, we found that all populations of Japanese quail we have sampled, both from the U.S. and Japan, exhibit this polymorphism. It is a product of two allelic, autosomal genes at a single locus. We can obtain the two separate forms by identifying homozygous Ser and homozygous Gly eggs (or, better, hens that lay them) and isolating ovomucoid from such eggs. Recently, we learned to distinguish OMJPQ3S and OMJPQ3G by reverse-phase HPLC (Lin et al., 1985).

In the case of the gray junglefowl (*Gallus sonneratii*), ovomucoid was isolated from eggs donated by M. Ollson in 1979. Sequencing and high-performance liquid chromatographic separation (Lin et al., 1985) indicate a 50/50 mixture of OMGJF3A and OMGJF3V—the Ala/Val polymorphism

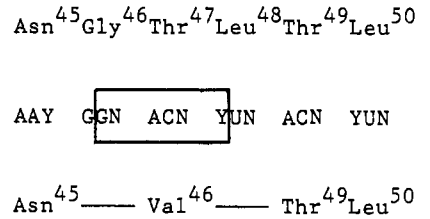


FIGURE 4: Scheme showing on top the amino acid sequence of a great majority of the ovomucoid third domains near the Asn^{45} glycosylation site, in the middle the most ambiguous (in case of Leu somewhat more ambiguous than the code) nucleotide sequence required to generate this sequence, and on the bottom the sequences of the two tinamou third domains OMBTI3 and OMCTI3. If we imagine that the deletion in the tinamou sequences occurred by the deletion of the six boxed nucleotides, we can account both for the two amino acid deletion and the Gly^{46} → Val^{46} mutation by a single event. It is clear that the loss of Thr^{47}-Leu^{48} in the two tinamou sequences was a deletion in tinamous rather than an insertion in the other sequences partly because of all the avian orders represented here only the tinamous show the deletion but more importantly because other Kazal domains do not have this deletion. On the other hand, it remains for speculation which nucleotides or residues, i.e., Thr^{47}-Leu^{48} or Leu^{48}-Thr^{49}, Thr^{49}-Leu^{53}, were deleted. The deletion proposed here seems most parsimonious.

is at the P_1 position. It is of interest that the third domains of chicken (*Gallus gallus*, OMCHI3) and of Ceylon junglefowl (*Gallus lafayettii*, OMCJF3) have the same sequence as OMGJF3A. Chicken ovomucoid third domain clearly is not polymorphic as we have obtained the same third domain from chicken eggs from numerous sources. The Ceylon junglefowl preparation we have is not polymorphic, but this may not necessarily be characteristic of the entire population (see below).

It is highly likely that several more significant polymorphisms exist in the populations of the species for which the monomorphic third domain sequence is given in Table II. We feel that this is primarily due to sampling errors. In most cases, the material was isolated from a small number of eggs, and there is a high probability that all of these were laid by a single female. However, even in an extensive dimorphism there is ≥50% probability of a single phenotype being homozygous. The second type of error, the failure to notice the minor amino acid compared to the major one, i.e., an experimental error, is likely to arise only when a mixture of many phenotypes (eggs laid by many females) is sampled since, with present techniques, such an error is not likely to be made on a 50/50 (heterozygous) mixture but only on less balanced mixtures arising from mixing of phenotypes.

(*3*) *Alignment.* As is immediately seen by inspection of Table II, the ovomucoid third domains align to each other very easily. Only in the case of the two tinamou species, OMBTI3 and OMCTI3, is it necessary to postulate a deletion[2] (see Figure 4).

Of the 106 sequences given in Table II, only 65 are unique. Table III is a listing of the duplicated ones. From the point of view of sequence to reactivity algorithm, this can be viewed as a "failure" list since these 41 sequence determinations did not yield any new variants. The long list is a consequence of examining closely related species in order to better define hypervariability and to obtain many variants that differ from one another by a single replacement rather than by several.

The clusters in Table III can be viewed as clusters of most closely related species provided that the sequence of ovomucoid third domain proper is used as the sole criterion for relatedness. It is gratifying that thus far members of two different orders

[2] It should be noted, however, that deletions need to be introduced within the third domains to align them with the other Kazal domains (see, for example, Figure 7 of the preceding paper).

Table III: Species Having Identical Ovomucoid Third Domain Amino Acid Sequences

REA = LER	Ratitae
WTD = HTD = BTD = FTD	Anseriformes
SWN = MSN = RUD = LKD	Anseriformes
GOO = BHG	Anseriformes
UPG = BWG = ORG = WSD = CRD = CAW = DUK = WOD = HDM	Anseriformes
CPV = AHH = WTE	Accipitridae
GUA = PGN[a]	Cracidae
GPC = SGG = STP = TTP = HMP = BEP = EDP = SVP = KOP = CHP = ELP	Phasianoidea
CAQ = GMQ	Phasianoidea
CHU = RPA	Phasianoidea
BTF = ERF[b]	Phasianoidea
CHI = CJF = GJF(A)[c]	Phasianoidea
RNP = VCP	Phasianoidea
LAP = GLP	Phasianoidea
GUI = VGU	Phasianoidea
TKY = OTK	Phasianoidea
WTC = WNC = DEC	Gruiformes
SPL = BGL	Charadriiformes

[a] Connecting peptide fragments are different. [b] Connecting peptides differ in length; different enzymes were used in cutting (see Table I). [c] GJF ovomucoid third domain is polymorphic. One of the two forms is identical with those of CHI and CJF.

do not belong to the same cluster. However, members of different suborders, SPL and BGL, are in the same cluster. Similarly, two (but not all, see RFG) members of the family Tetraonidae, GPC and SGG, belong to the same cluster as do many (but by far not all) members of the family Phasianidae (STP, TTP, HMP, BEP, EDP, SVP, KOP, CHP, and ELP). Indeed, our data here (as well as the data on first domain sequences) strongly suggest that the division of the superfamily Phasianoidea into its component families may require revisions.

Table IV is a difference matrix of the 65 unique sequences reported in Table II. For duplicated sequences only the first entry given in Table III is used. Note that the number of minimal base changes (upper right) is only slightly larger than the number of amino acid differences (lower left). This indicates that relatively few multiple base changes within codons were fixed during the divergence of the third domains. This is one of many indicators that the sequences are closely related. The maximal amino acid difference in Table IV is 20 (OMGPP3 vs. OMCTI3). Since OMCTI3 has only 49 residues, this corresponds to 29/49 or 59% identity. Most differences are much smaller. The largest minimal base change distance is 25. This time it is a tie: again OMGPP3 vs. OMCTI3 and also OMCHA3 vs. OMBTI3. Avian ovomucoid third domains appear to be more closely related to each other than to any of the about 150 other Kazal family domains in the Purdue sequence list.

(*4*) *Phylogeny*. It is clear that the two tinamou sequences are relatively distant from all of the other avian sequences both in having the deletion (see Figure 4) and in showing very large distances to all other sequences in Table IV.

It is gratifying for ornithologists that most large distances in Table IV are interordinal; on the other hand, not all interordinal distances are large, nor are all the intraordinal distances small. The largest intraordinal difference is OMBUQ3 vs. OMCOO3 of 15 amino acids. As OMBUQ3 is less distant from several sequences outside the order Gruiformes than it is to other members of this order, we question the already controversial assignment of BUQ to Gruiformes. On the other hand, the next highest intraordinal difference is within the superfamily Phasianoidea, OMGPP3 vs. OMBPA3, with 13. This is probably not a misassignment as both OMGPP3 and OMBPA3 show many smaller distances to other members of Phasianoidea. The set here is very large, and two very aberrant sequences within it are being compared.

We do not provide a phylogenetic tree, postponing that effort to the completion of a matching set of first domain sequences. However, we provide here for some "orders" (boxes in Table IV) a listing of striking characteristics, a classification key that might allow one to assign a third domain sequence to a particular group (Table V).

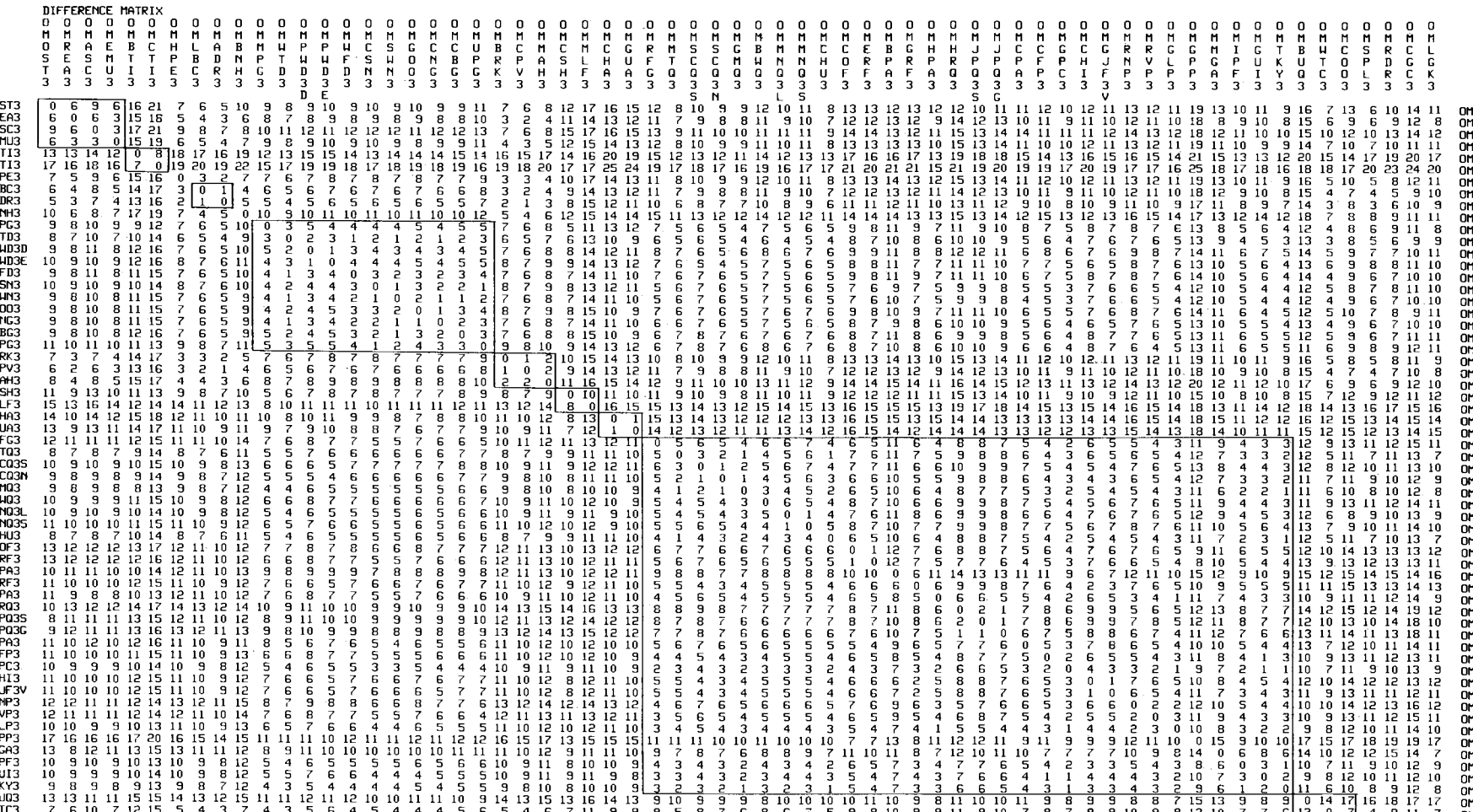
Which Positions Vary? A different type of analysis is to examine the entire set to see which positions accept many mutations and which remain unvaried. This is done in Tables VI and VII. Our interest here is both in the unvaried or almost unvaried positions and in those that vary the most.

(*1*) *Residues Which Determine the Folding Are Highly Conserved*. Table VIII is a listing of residues that are unvaried or only slightly varied in the entire sequence set. Most of these residues appear to be strongly conserved for structural reasons. We conclude this in part because of the chemical nature of many of the residues listed in Table VIII such as Pro, Gly, and disulfide-bridged Cys. In addition, on the basis of the extensive X-ray crystallographic studies on three of these third domains [OMJPQ3S and OMJPQ3G mixture (Weber et al., 1981; Papamokos et al., 1982), OMSVP3 (Bode et al., 1985), and OMTKY3 in complex with *Streptomyces griseus* proteinase B (Fujinaga et al., 1982; Read et al., 1983)], we can provide a plausible structural reason for almost all of the conserved residues without invoking any functional explanations [save for Asn^{45} (P'_{27}), which is the glycosylation site].

A particularly important column in Table VIII is the exceptions column. There are several cases where a residue is conserved in all but one or two of the 106 sequences and where a plausible reason for the conservation is often provided. Seldom, however, can we provide a reason for the exceptions. The simplest of these puzzling cases in OMOST3; all other third domains have P'_{27} Asn and are partially or completely glycosylated at that position. We are at a loss to explain why OMOST3 is the only domain that has Ser at that position. Other cases are more amenable to experimental study. All but one of the third domains have a Tyr^{31} to Asp^{27} hydrogen bond, but OMGPP3 changes this Asp^{27} to Asn^{27}, and we are doing spectroscopic and denaturation studies to ascertain the consequences of this replacement. Several cases seem to be good candidates for X-ray crystallography or for energy minimization calculations or for both. They are OMGPP3, OMSCQ3S, where the almost unvaried Asn^{33} that is hydrogen bonded to the reactive site residues is replaced by Ser^{33}, and OMSPL3 (also OMBGL3), where the otherwise unvaried Gly^{54} is replaced by Glu^{54} and presumably there is a conformational change in the main chain as a result. We surmise this since the Ramachandran angles at Gly^{54} in all the known three-dimensional structures are unfavorable for any other residue. OMCTI3 and OMBTI3, as well as OMMLF3, OMWFD3, and OMASC3, could be added to this list.

(*2*) *Hypervariability of Enzyme–Inhibitor Contact Region*. In contrast to the above, for some positions there are a great number of alternatives. A glance at Table VI demonstrates that there is a strong positive correlation between the number of alternatives per position and the position being in contact with the enzyme in an enzyme–inhibitor complex. This is brought into even sharper focus by Table VII. The 7 positions with the largest number of alternatives are all among the 11 that are in contact with the enzyme. There is a tie for the eighth and nineth largest number of alternatives, and one of these positions is also in contact with the enzyme. The im-

Table IV: Difference Matrix of Third Domains Proper (Residues 6–56 of Table II)[a]



[a] In order to reduce the size of this matrix, each cluster of identical third domains proper is represented here only by one sequence—the first one in the listing of such clusters in Table IV. The portion of the matrix below the diagonal is amino acid differences; the portion above the diagonal is minimal base changes. In neither section is there any penalty for deletion. This is important only in comparing the two tinamou sequences OMBT13 and OMCT13 with sequences of other species. From upper left to lower right, the rectangles isolate the following intragroup comparisons: Ratitae, Tinamiformes, Pelecaniformes, Anseriformes, Accipitridae, Megapodiidae, Cracidae, Phasianoidea, Gruiformes, and Cuculiformes. Groups represented by a single sequence are not boxed.

Table V: Characteristics of Sequences of Ovomucoid Third Domains Belonging to Various Ornithological Groups

classification	third domain	sequence characteristic	note
Ratitae (superorder)	OMOST3 OMREA3 OMLER3 OMASC3 OMEMU3	connecting peptide Phe^1-Ala^2-Thr^3-; P_4 (15), Val; P_2 (17), Ser or Pro; P'_3 (21), Met; P'_{25} (43), Glu	
Tinamiformes (order)	OMBTI3 OMCTI3	P_4 (15), Ala; P_2 (17), Thr; P'_3 (21), Phe; P'_{11} (29), Gln; P'_{18} (36), Ala; P'_{25} (43), Glu; P'_{26} (44), Lys; P'_{28} (46), Val; deletion at 47 and 48	residue 53 is not Phe
Anseriformes (order)	from OMMPG3 to OMLKD3 in Table II (26 sequences)	connecting peptide Val^1-Ala^2-Thr^3-; P_4 (15), mostly Ala; P_2 (17), Thr; P'_3 (21), Met; P'_{25} (43), Asp	P_4, no Val; P_2, exception in OMCBG3
Accipitridae (family)	OMBRK3 OMCPV3 OMAHH3 OMMAH3 OMWTE3	connecting peptide Ile^1-X^2-Ile^3-; P_9 (10), Asp; P_4 (15), Val; P_2 (17), Ser; P_1 (18), Leu; P'_3 (21), Met; P'_{14} (32), Ser; P'_{18} (36), Asp; P'_{25} (43), Asp	enzyme–inhibitor contact residues do not change within this group
Megapoiidae (family)	OMCSH OMMLF	connecting peptide X^1-Val^2-Thr^3-; P_2 (17), Thr; P'_3 (21), Lys	
Phasianoidea (superfamily)	from OMRFG3 to OMOTK3 in Table II (49 sequences)	connecting peptide contains Val^4-Ser^5; P_9 (10), Glu; P_4 (15), Ala; P_2 (17) Thr; P'_3 (21), Arg; P'_{25} (43), Glu	P_9, exceptions in OMBPA3 and OMCPA3; P_4, no Val; P_2, exceptions in OMBWQ3, OMJPQ3, and OMHRQ3; P'_3, exceptions in OMBPA3, OMMNQ3L, and OMNNQ3S; P'_{25}, exceptions in OMBTF3, OMCOF3, and OMERF3
Gruiformes (order)	OMBUQ3 OMWTC3 OMWNC3 OMDEC3 OMCOO3	connecting peptide Thr^1-Ala^2-Thr^3-; P_9 (10), Asp; P_4 (15), Ala; P_2 (17), Thr or Ser; P'_5 (23), Phe; P'_{25} (43), Asp	OMBUQ3 disregarded, see text
Charadiiformes (order)	OMSPL OMBGL	P'_{36} (54), Glu	this is an exceptional residue in the third domain, see Table VIII

Table VI: Amino Acid Sequence of Turkey Ovomucoid Third Domain (Shown Vertically) and Alternatives from 100 Additional Avian Species (Shown Horizontally)[a]

Val Ile	6	136	13	
Asp Asn	7	137	12	
Cys	8	138	11	
Ser	9	139	10	
Glu Asp Gly	10	140	9	
Tyr His	11	141	8	
Pro	12	142	7	
***Lys Arg Thr Met Gln	13	143	6	P_n
***Pro His	14	144	5	
***Ala Val Asp Ser Gly Glu Thr	15	145	4	
***Cys	16	146	3	
***Thr Pro Ser Arg Leu Met	17	147	2	
***Leu Pro Met Val Gln Ser Ala Thr Lys	18	148	1	
***Glu Asp Leu	19	149	1	
***Tyr Phe Asn Asp Leu Gln Glu His Arg	20	150	2	
***Arg Met Phe Lys Val Thr Leu	21	151	3	
Pro	22	152	4	
Leu Val Ile Phe	23	153	5	
Cys	24	154	6	
Gly	25	155	7	
Ser	26	156	8	
Asp Asn	27	157	9	
Asn Ser	28	158	10	
Lys Gln Ile Glu	29	159	11	
Thr Ser Ile	30	160	12	
Tyr	31	161	13	
***Gly Ser Asn Ala Asp Val Arg His	32	162	14	
Asn Ser Asp	33	163	15	
Lys Arg	34	164	16	$P_{n'}$
Cys	35	165	17	
***Asn Asp Ala Ser Gly Tyr	36	166	18	
Phe	37	167	19	
Cys	38	168	20	
Asn Ser	39	169	21	
Ala	40	170	22	
Val Ala Phe	41	171	23	
Val Ala Leu Met	42	172	24	
Glu Asp Gln Lys His	43	173	25	
Ser Lys	44	174	26	
Asn Ser	45	175	27	
Gly Val	46	176	28	
Thr -	47	177	29	
Leu -	48	178	30	
Thr Ile Asn Ser	49	179	31	
Leu Val Phe	50	180	32	
Ser Arg Gly Asn	51	181	33	
His Arg	52	182	34	
Phe Leu Ile	53	183	35	
Gly Glu	54	184	36	
Lys Glu Thr Gln	55	185	37	
Cys	56	186	38	

[a] Three numbering systems are presented: turkey third domain, entire chicken ovomucoid (see preceding paper), and Schechter and Berger (1967). The horizontal line divides the sequence at the reactive site peptide bond. The residues that are in contact with the enzyme in the turkey ovomucoid third domain—*Streptomyces griseus* proteinase B (Read et al., 1983)—are starred.

Table VII: Frequency with Which the Various Number of Alternatives Occur in Table VI[a]

no. of alternatives	no. of positions	positions involved
8	2	P_1 (*18*), P'_2 (*20*)
7	1	P'_{14} (*32*)
6	2	P_4 (*15*), P'_3 (*21*)
5	2	P_2 (*17*), P'_{18} (*36*)
4	2	P_6 (*13*), P'_{25} (43)
3	6	P'_5 (23), P'_{11} (29), P'_{24} (42), P'_{31} (49), P'_{33} (51), P'_{37} (55)
2	7	P_9 (10), P'_1 (*19*), P'_{12} (30), P'_{15} (33), P'_{23} (41), P'_{32} (50), P'_{35} (53)
1	15	P_{13} (6), P_{12} (7), P_8 (11), P_5 (*14*), P'_9 (27), P'_{10} (28), P'_{16} (34), P'_{21} (39), P'_{26} (44), P'_{27} (45), P'_{28} (46), P'_{29} (47), P'_{30} (48), P'_{34} (52), P'_{36} (54)
0	14	P_{11} (8), P_{10} (9), P_7 (12), P_3 (*16*), P'_4 (22), P'_6 (24), P'_7 (25), P'_8 (26), P'_{13} (31), P'_{17} (35), P'_{19} (37), P'_{20} (38), P'_{22} (40), P'_{38} (56)

[a] The residues are given both in the reactive site (Schechter & Berger, 1967) notation and (in parentheses) in ovomucoid third domain notation. The positions that are in contact with the enzyme are italicized.

pression is overwhelming and yet totally contrary to what we have been taught to expect. If we grant that the biological role of at least some[3] of the ovomucoid third domains is to inhibit serine proteinases, then we have shown that in at least one set of orthologous proteins functional residues are not conserved but instead are hypervariable. The proof that

[3] Normally in thinking about a set of orthologous and closely related proteins in which the functional residues are conserved, we think of a common function for all members of the set. The findings reported here suggest that some of the members of the set may function as serine proteinase inhibitors while others may not. An alternative view is that the target enzyme varies from species to species, again mitigating against the hypothesis of a common function.

Table VIII: Strictly Unvaried and Almost[a] Unvaried Residues

		residue	reason for conservation[b]	exceptions
Val	P_{13}	6	part of the hydrophobic core	Val^6 → Ile^6 in OMCTI3 and OMHPE3
Asp	P_{12}	7	hydrogen bonds to Ser^9 N and Ser^9 OG	Asp^7 → Asn^7 in OMRDR3
Cys	P_{11}	8	disulfide bridge to Cys^{38}	none
Ser	P_{10}	9	hydrogen bonding to Asp^7	none
Pro	P_7	12	*cis*-proline	none
Pro	P_5	14	rigidity for the reactive site loop	Pro^{14} → His^{14} in OMMLF3
Cys	P_3	16	disulfide bridge to Cys^{35}	none
Pro	P'_4	22	buried, contacts with 7 different residues	none
Cys	P'_6	24	disulfide bridge to Cys^{56}	none
Gly	P'_7	25	type 1 reverse turn	none
Ser	P'_8	26	reverse turn, hydrogen bonds to Thr^{49} N and Thr^{49} O	none
Asp	P'_9	27	hydrogen bond to Tyr^{31} OE in center of molecule	Asp^{27} → Asn^{27} in OMGPP3
Tyr	P'_{13}	31	hydrogen bond to Asp^{27} OD1 in center of molecule	none
Asn	P'_{15}	33	spacer [hydrogen bonds to residues P_2 (Thr^{17} O) and P'_1 (Glu^{19} O)]	Asn^{33} → Ser^{33} in OMSCQ3A and Asn^{33} → Asp^{33} in OMBUQ3
Lys	P'_{16}	34	hydrogen bond to Asp^7 O	Lys^{34} → Arg^{34} in OMWFD3
Cys	P'_{17}	35	disulfide bridge to Cys^{16}	none
Phe	P'_{19}	37	totally buried, hydrophobic interactions	none
Cys	P'_{20}	38	disulfide bridge to Cys^8	none
Asn	P'_{21}	39	hydrogen bonds to Lys^{13} N and Lys^{13} O	Asn^{39} → Ser^{39} in OMASC3
Ala	P'_{22}	40	interacts with conserved Tyr^{31} and Phe^{37}, larger residues would interfere	none
Ser[c]	P'_{26}	44	hydrogen bonds to Leu^{48} N and to Val^{41} O	Ser^{44} → Lys^{44} in OMBTI3 and OMCTI3
Asn	P'_{27}	45	glycosylation site	Asn^{45} → Ser^{45} in OMOST3
Gly[c]	P'_{28}	46	type 1′ reverse turn	Gly^{46} → Val^{46} in OMBTI3 and OMCTI3
Thr	P'_{29}	47	hydrogen bond to Ser^{44} O, glycosylation signal	Thr^{47} is deleted in OMBTI3 and OMCTI3
Leu[c]	P'_{30}	48	buried, contacts with other conserved residues	Leu^{48} is deleted in OMBTI3 and OMCTI3
Phe[c]	P'_{35}	53	hydrophobic interactions with other buried residues	Phe^{53} → Leu^{53} in OMBTI3 and Phe^{53} → Ile^{53} in OMCTI3
Gly	P'_{36}	54	main chain angles favorable only for Gly	Gly^{54} → Glu^{54} in OMSPL3 and OMBGL3 (these two sequences are identical)
Cys	P'_{38}	56	disulfide bridge to Cys^{24}	none

[a] All of the positions listed in Table VII as unvaried are listed here. An arbitrary choice was made of positions showing one or two alternative in Table VII on the basis of our opinion that the position was little varied in evolution. Residues such as P_8 (residue 11) that show only a single alternative in Table VII (Tyr or His) but that obviously change a great deal during evolution (see Table II) are excluded. [b] In this table the letters O, N, and C designate atoms in amino acid residues. This is a capsule subjective interpretation made at Purdue on the basis of three-dimensional structural data of Weber et al. (1981), Papamokos et al. (1982), Fuginaga et al. (1982), Read et al. (1983), and Bode et al. (1985). [c] All of the residues are unvaried in all of the sequences in Table II except for OMBTI3 and OMCTI3. Since all these nonconservations occur at or very close to the site of the deletion in the tinamou system, they are most probably all a consequence of this deletion. We can therefore think of these residues as unvaried in all of the full length sequences.

ovomucoid third domains are in fact physiological serine proteinase inhibitors is lacking, and as the discussion below attempts to show, it may be especially difficult to obtain. We continue the argument without proof and with the hope that some of our biological colleagues may choose to provide it.

The proper proof of hypervariability should involve a count of fixations, not of alternatives, at each position. In collaboration with Professor Fitch, we have constructed phylogenetic trees on the basis of these data and on some first domain sequences. The results do not alter the conclusions; the number of fixations in contact positions is anomalously high compared to other positions.

At first it may appear that all that is needed to explain the conflict between conventional beliefs and the results given here is to postulate that none of the ovomucoid third domains function as proteinase inhibitors. In our opinion, this is not enough. The correlation between highly variable positions and contact positions is so strong that an additional explanation would be needed as to why other surface residues do not evolve so rapidly. The side of the molecule opposite from the reactive site accepts mutations slowly.

Objections could also be raised that contacts with *Streptomyces griseus* proteinase B are taken as representative of contacts with all serine proteinases and with a target enzyme(s) in particular. In a recent paper we showed that eight different serine proteinases are inhibited by OMTKY3 at the same reactive site (Ardelt & Laskowski, 1985). The contacts between OMTKY3 and chymotrypsin (Read et al., 1984, M. N. G. James, personal communication) are essentially the same as those with *Streptomyces griseus* proteinase B. The exceptions are that P_6 (13) Lys does not make contact and neither does P'_{14} Gly. In the latter case, however, it is clear that any larger residue than Gly would make contact. A glancing contact is made between P'_{15} (33) Asn and chymotrypsin, which is not seen with *Streptomyces griseus* proteinase B. In the earlier modeling work of Papamokos et al. (1982), we have made models of complexes of trypsin, chymotrypsin, and pancreatic elastase with the known structure of OMJPQ3. The contact positions appeared to be essentially the same as those listed here [the complex between OMJPQ3 and pancreatic elastase could not be built until P_1 (18) Lys was replaced by a smaller residue]. Bolognesi et al. (1982) find that porcine pancreatic secretory trypsin inhibitor (Kazal) makes roughly the same contacts with bovine trypsinogen as those described here.

On the other hand, Papamokos et al. (1982) failed to construct a credible complex of OMJPQ3 with subtilisin. Hirono et al. (1984) find that binding of *Streptomyces* subtilisin inhibitor (SSI) to subtilisin involves contacts not easily anticipated from the structures of the enzyme and inhibitor alone,

and therefore, some variation of contact residues in subtilisin complexes might be expected. Even then most of the expected contacts are made, as well as some unexpected ones.

Another line of support for the contact positions being essentially the same for all enzymes are our results on the determination of the association equilibrium constant between most of the unique sequence domains reported here and chymotrypsin, elastase, subtilisin, and *Streptomyces griseus* proteinases A and B (Laskowski et al., 1981, 1983; Empie & Laskowski, 1982; Tashiro & Laskowski, 1983; Park, 1985). We find that changes in contact residues almost always cause changes in K_a values and many of these changes are very large. On the other hand, noncontact residue changes seldom affect K_a values.

One could argue that the residues in contact are hypervariable but that the changes that occur in the avian ovomucoid third domain set have little effect on enzyme–inhibitor contact with the target enzyme. If so, this enzyme would be a truly unusual serine proteinase. We have measured the enzyme–inhibitor association constants for 55 of the 65 unique sequence third domains reported in this paper with the five serine proteinases listed before. Of these, three are bacterial and two are vertebrate; they represent both the chymotrypsin and the subtilisin families. The *least* sensitive of them is *Streptomyces griseus* proteinase B, whose K_a with members of our set ranges from 1.9×10^5 M^{-1} with OMGUA3 to 8.5×10^{10} M^{-1} with OMPWD3D and OMPWD3E mixture. This is a 6 order of magnitude range. The other enzymes are far more sensitive. In the above example the lower bound of the range is clearly understated. The 10 ovomucoid third domains reported here whose K_a's we did not measure were not measured because they proved to be ineffective as inhibitors in preliminary tests and their amounts were not sufficient for detailed study. (If an inhibitor is *very* weak, very large amounts are needed for measurement).

If ovomucoids are active as inhibitors, then it appears that the evolution of ovomucoid third domains yields greatly different inhibitory activities against a common target enzyme or comparable inhibitory activities against very different (and unknown) target enzymes in closely related species. This is epitomized in OMMNQ3L and OMMNQ3S where the polymorphism is at P_1 (18) (the primary specificity residue of the reactive site) and the effect of the Leu to Ser change makes the K_a's of the Ser form weaker by 25 times against subtilisin and by 5×10^3 times against chymotrypsin with the other enzymes falling between these two extremes.

The reasons why most of us believe that ovomucoid third domains are likely to be physiological proteinase inhibitors are as follows.

(1) Among the ovomucoid third domains reported here are some of the strongest protein inhibitors for bovine chymotrypsin, porcine pancreatic elastase I, *Streptomyces griseus* proteinases A and B, and subtilisin Carlsberg thus far described in the world literature (Empie & Laskowski, 1982; Laskowski et al., 1983; Park, 1985). These are the enzymes we study most intensely. Many other serine proteinases are also very strongly inhibited by several ovomucoid third domains (Ardelt & Laskowski, 1985).

(2) Ovomucoid third domains are homologues of proteins for which inhibition of serine proteinases is an agreed-on function. The most obvious set of such proteins is the set of pancreatic secretory trypsin inhibitors (Kazal) that are present in the zymogen granules and pancreatic secretions of vertebrates, including birds (Laskowski et al., 1980), and whose accepted role is to prevent the premature activation of trypsinogen(s) and, in turn, of most other pancreatic zymogens (Greene et al., 1976). The other and possibly more relevant example is the 7 Kazal domain ovoinhibitor that is present both in the eggs and the blood of birds (Barrett, 1974), where its function as a proteinase inhibitor is not questioned. (This inhibitor is discussed again below.)

(3) Avian eggs are very rich in actual and putative inhibitors (listed below) of many proteinases. These inhibitors have very different structures; what they have in common is being inhibitors. Therefore, it is suggestive that they should function as inhibitors.

If it is true that ovomucoid third domains differ from many other proteins by not conserving their functional residues, then how and why do they do it? Our knowledge of this is limited. Our group is divided in thinking either that their inhibitory activity is bacteriocidal and, thus, the rapid evolution is in response to the differences in bacterial floras most likely to infect egg whites of a given species or that their inhibitory activity is directed against the bird's own proteinases, possibly those involved in some of the last events in embryonic tissue differentiation. In either case, if our premises are correct, ovomucoid third domains are species-specific proteins as opposed to species-universal proteins such as cytochrome c, hemoglobin, etc. They are likely to be much less important to the organism than the species-universal proteins, but this relative lack of importance allows them to evolve in an unusual way. This unusual evolution is greatly aided by the presence of numerous putative or actual proteinase inhibitors in egg whites. We hasten to add that to the best of our knowledge none of these inhibitors have a clearly assigned physiological function in egg whites. However, the presence of these inhibitors would allow for two ways in which the unusual evolution of ovomucoid third domains could be supported: (1) as the third domains lose their inhibitory activity toward their target enzymes, other inhibitors continue inhibiting those enzymes and thus carry the load; (2) various inhibitory reactive sites could switch target enzymes (coevolution) with the result that the egg white as a whole would still inhibit the same ones. The inhibitors present in avian egg whites are described below.

(a) Ovomucoid is clearly the most studied proteinase inhibitor in egg whites. It is present in egg white at a concentration of 10 mg/mL or approximately 3×10^{-4} M, and its amount in various egg whites is surprisingly constant [Feeney and Allison (1969) and Table I of this paper]. It consists of three tandem, homologous domains, all of which have an actual or putative reactive site (see the preceding paper). This should allow the domains either to carry the load for one another or to coevolve with one domain taking over the function lost by the other. However, this does not appear to happen very frequently as first and second domains often have Lys or Arg P_1 and, thus, inhibit trypsin-like enzymes while in third domains P_1 has never been found to be Arg and it is Lys only in three sequences.

(b) The most likely candidate for coevolution with ovomucoid is *ovoinhibitor* (Matsushima, 1958). This protein consists of seven tandem Kazal-type domains (Laskowski et al., 1980). In chicken the first four from the NH_2 terminus have Arg at their reactive site P_1 positions, the fifth has Phe, and the last two have Met. Chicken ovoinhibitor strongly inhibits many of the enzymes typically inhibited by ovomucoid third domains such as OMTKY3. On the other hand, chicken third domain is an ineffective inhibitor of any enzyme tested thus far. Thus, ovoinhibitor is also a good candidate simply for carrying the load when a third domain becomes weak. Another important aspect of ovoinhibitor is that the same protein is expressed in

egg whites and in avian blood, where it presumably functions as an important proteinase inhibitor (Barrett, 1974; Kato & Kohr, 1978; Laskowski, et al., 1980) and therefore may be subject to evolutionary pressures extrinsic to egg whites.

(c) It was only recently recognized (Ikai et al., 1983; Nagase et al., 1983) that ovomacroglobulin, which was described in many avian egg whites (Feeney & Allison, 1969), is a proteinase inhibitor of very low specificity. Like its distant homologue blood serum α_2-macroglobulin, it "inhibits" proteinases from all four mechanistic classes. Since its specificity is so low, it is not a good candidate for coevolution with avian ovomucoid third domains, but it could take a great deal of pressure off the third domain evolution.

(d) Ovocystatin is a newly named and sequenced (Turk et al., 1983; Schwabe et al., 1984) papain inhibitor of Fossum and Whittaker (1968). While it inhibits only cysteine proteinases and ovomucoid inhibits only serine proteinases, it is mentioned here to point out that egg whites contain many proteinase inhibitors with very different stabilities, molecular weights, and structural characteristics.

(e) Ovalbumin. This major protein of egg whites is a member of the serpin family, which includes mammalian α_1-proteinase inhibitor, α_1-antichymotrypsin, and antithrombin III (Hunt & Dayhoff, 1980), all of which appear to have agreed-on physiological functions (Travis & Salvesen, 1983). Chicken ovalbumin does not inhibit any of the common serine proteinases tested against it thus far. On the other hand, both subtilisin (Linderstrom-Lang & Ottesen, 1949) and porcine elastase (Wright, 1984) hydrolyze it rapidly and specifically at the Ala^{352}–Ser^{353} bond, which is (by alignment to the known reactive sites of other members of the family) the putative reactive site. Estell and Laskowski (1980) have shown that there exist enzymes, homologous and closely related to enzymes that are strongly inhibited by protein inhibitors, that instead of being inhibited rapidly hydrolyze the reactive sites of many protein inhibitors. Thus, ovalbumin may well be an inhibitor for an as yet undiscovered or simply untested serine proteinase. It would be interesting to compare its putative reactive site sequence from several avian species.

(*3*) *How General Is Contact Area Hypervariability*? It is highly likely that hypervariability is not limited to ovomucoid third domains but that it encompasses avian ovomucoid first and second domains. This conclusion is based on inspection of a large ($\simeq$100) number of unpublished first domain sequences and of a smaller ($\simeq$15) number of second domain sequences available in our laboratory. However, it should be noted that three-dimensional structures of complexes of enzymes with either first or second domains are not available; therefore, the conclusion is based on the assumption that residues equivalent to those in third domains (P_6–P'_3, P'_{14}, P'_{18}) are in contact. With this assumption, the contact area hypervariability of first and second domains seems quantitatively similar to that of third domains, except that the P_1 position is clearly less variable in the first and second domains than it is in third domains.

The notion that P_1 hypervariability is a characteristic of many and possibly all protein inhibitors of serine proteinases is quite old [see Laskowski and Kato (1980) for a review]. Alignments of all of the known sequences in each inhibitor family show both great P_1 variation and considerable variation of the surrounding residues. Thus, it appears that hypervariability of contact residues is a general property of this group. It seems even to include the serpins (Carrell & Travis, 1985) in spite of the considerable difference in inhibitory behavior between this inhibitor family and the other families whose members obey the "standard mechanism" quite strictly.

Another indication of hypervariability among protein inhibitors of serine proteinases is the recent finding of Kingston and Anderson (1986), who sequenced the paralogous genes for bovine pancreatic trypsin inhibitor (Kunitz) and bovine spleen inhibitor II. There are nine nucleotide differences in the coding region. Of these, only two are silent changes. Among the seven amino acid replacements, three are of enzyme–inhibitor contact residues. These results (relatively small number of silent changes and high percentage of changes of residues in contact) indicate evolutionary selection.

Thus, hypervariability seems to be a characteristic of most or possibly all families of protein inhibitors of serine proteinases. The other group where there are sufficient number of sequences to ask the question are protein inhibitors of cysteine proteinases. In an alignment of their sequences, Mueller-Esterl et al. (1985) noticed that the sequence ...Gln-Val-Val-Ala-Gly... is strongly conserved (but not unvaried) in all the inhibitors. On this basis they postulate that this sequence is likely to encompass the reactive site. If they are correct, then protein inhibitors of cysteine proteinases evolve the way most other proteins do and strikingly differently from protein inhibitors of serine proteinases.

(*4*) *Analogy to Immunoglobulins*. Many biochemists when first exposed to the ovomucoid system wish to draw a direct analogy to the immunoglobulins. In both systems there are multiple, highly defined domains, and in both systems the combining region (enzyme–inhibitor contact residues in ovomucoids, the antigen combining site in immunoglobulins) is hypervariable. The analogy is somewhat flawed. In multidomain immunoglobulin chains only the NH_2-terminal domain has the antigen combining site; in ovomucoids and in multidomain Kazal inhibitors [see Laskowski and Kato (1980) and Laskowski et al. (1980)], all domains have either actual or at least potential reactive sites, and cases are known when all of them simultaneously bind enzymes. The second comparison involves hypervariability, first discussed for immunoglobulins by Wu and Kabat (1970). In both cases we have large sets of natural proteins in which the structurally important residues are conserved while ones in the combining sites vary widely. Thus, both systems present physical protein chemists with rather similar challenges. On the other hand, the systems are completely different from the point of view of an evolutionist. In immunoglobulins the comparison involves many homologous proteins from the same species. In contrast, there appears to be only one ovomucoid gene in each avian species (Stein et al., 1980; Bogard et al., 1980), and the hypervariability involves the comparison of orthologous single-copy proteins from various species. Thus when the inhibitory properties of an ovomucoid third domain change, there is both a gain of a new specificity and a loss of an old one (not the case with immunoglobulins where so many different variants coexist). We have been saying, however, that the presence of the two other inhibitory domains in ovomucoid, of seven domains in ovoinhibitor, and of a broad spectrum ovomacroglobulin may in fact make the two situations more parallel than they seemed to us at first.

Ovomucoid Third Domains as Variants for Sequence to Property Algorithms. Many scientists recently have realized that a powerful approach to protein chemistry is now possible. One generates an appropriate set of variants, which are likely to have nearly identical and well determined conformations, each differing from its nearest neighbor by one or at most a few amino acid substitutions. One then measures the property of interest for all members of the set in order to infer how this

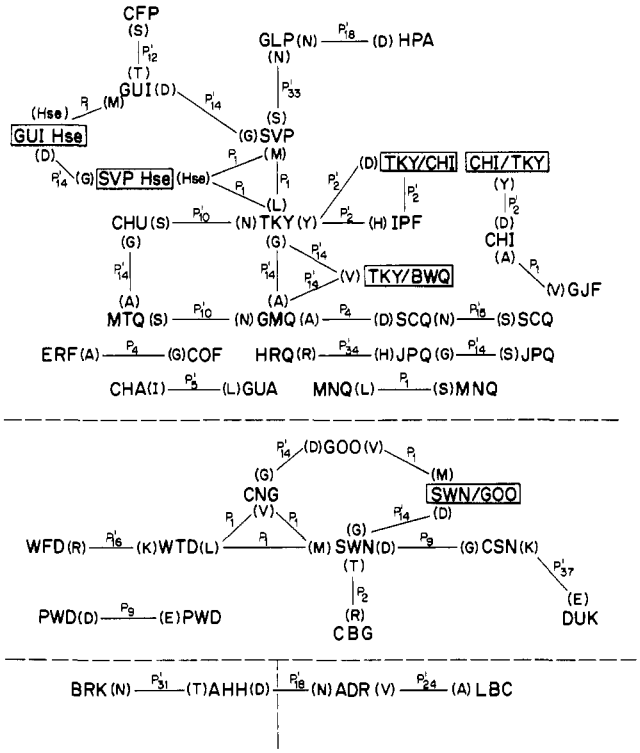


FIGURE 5: Diagram showing the interrelationship of the ovomucoid third domains *proper* (residues 6–56), which differ from their nearest neighbor(s) by a single amino acid substitution, i.e., correspond to ones in the difference matrix (Table III). In order to avoid cluttering this diagram even further, the letters OM and the final 3 are omitted so that we write TKY for OMTKY3. Note that in a few cases, e.g., GMQ and TKY, the connecting peptides also differ, and thus, there are two differences between the connecting peptide extended third domains of these species. The differences between the sequences and the positions where they occur are indicated. The figure is augmented by six enzymatic semisynthetic derivatives (boxed). Four of them are covalent hybrids of the natural species; i.e., SWN/GOO corresponds to covalently joined residues 1–18 of OMSWN3 and residues 19–56 of OMGOO3 [Wieczorek & Laskowski, 1983; Wieczorek et al. (1983) and unpublished results]. The derivatives SVP-Hse and GUI-Hse were obtained by converting the single Met^{18} (P_1) in OMSVP3 and OMGUI3 respectively to Hse followed by enzymatic resynthesis of the Hse^{18}–Glu^{19} peptide bond (Wieczorek and Laskowski, unpublished results). The figure is divided into four sections to indicate the phylogenetic classification of the birds that form these closely related clusters. All the birds in the top clusters belong to Phasianoidea, in the middle clusters to Anseriformes, on the bottom on the left-hand side to the family of Acciptitridae of the order Falconiformes, and on the right-hand side to the family Phalacrocaracidae of the order Pelecaniformes. It is remarkable that the interordinal distance AHH to ADR is only one replacement. In this figure, which is a working figure for the algorithm work, the species chosen from the identity clusters of Table IV is not always the first listed there but rather the first sequenced in our laboratory or the most accessible. Otherwise, CPV would replace AHH in this figure.

property is affected by the sequence of the protein (Ackers & Smith, 1985).

The set of ovomucoid third domains described in this paper, augmented by six semisynthetic third domains described in Figure 5, is such a protein set. As this research developed, we decided quite early (in the late 1970s) that the hypervariability of the enzyme–inhibitor contact residues offered an exceptional opportunity in sequence to reactivity studies. This is so because most wild-type proteins conserve their functional residues and thus do not differ in functional properties until they have accepted many mutations. Ovomucoid third domains often show important changes after fixation of only a single substitution. Thus, as this work continued to generate a larger and larger data set for proving hypervariability, its principal goal switched to the acquisition of the best possible variant set for sequence to reactivity studies.

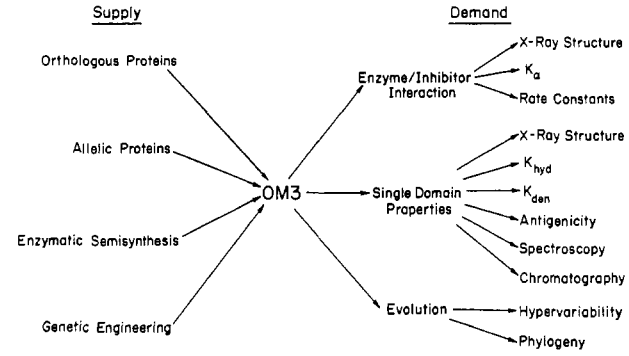


FIGURE 6: Flow diagram showing the supply of and demand for ovomucoid third domains.

Thus far we have made a case only for the use of the ovomucoid third domain set in the studies of enzyme–inhibitor interaction. However, as the set grew and became known to others, both they and we have suggested many additional uses unrelated to the contact residue hypervariability.

In sequence to property algorithms it is easiest to draw correct conclusions if the variant pairs compared differ by a single replacement. Figure 5 is a list of such sequences presented in diagramatic form. It contains 37 natural entries plus 6 semisynthetic variants. Earlier in this paper we called Table III our "failure" list. Now we can call Figure 5 our "success" list. To the best of our ability, the choice of species was made to make this set as large as possible. It should be noted that many of the remaining 28 (65 – 37 = 28) variants have turned out to be highly valuable. As the set grows, they may also be promoted to Figure 5 because intermediate sequences may be found or prepared.

Figure 6 is a flow diagram showing how new variants of ovomucoid third domain are generated (the supply side) and how they are used in our group and in other groups (the demand side). The main body of this paper shows the generation of 60 orthologous proteins and of 5 additional allelic variants. The additional six entries in Figure 5 show what has been accomplished by enzymatic protein semisynthesis. This procedure is now being further extended to coupling synthetic residues 1–18 to natural 19–56 peptides, thus allowing for choice of placement of residues in the important contact positions P_6, P_4, P_2, and P_1 (unnatural amino acids can also be included this way, e.g., Hse). An impressive characteristic of a really large variant set of very small proteins is that both the covalent hybrid possibilities and the possibilities of useful chemical modifications rise greatly as one can choose the optimal variant to use. The obvious way to augment the set with desirable variants is genetic engineering. N. Warne of our laboratory is doing so in collaboration with J. E. Dixon's laboratory at Purdue.

On the demand side, we roughly, but possibly somewhat inaccurately, divide the properties we measure into three sets. The properties of enzyme–inhibitor interaction are the "functional" characteristics we wish to study, but they depend also upon the enzyme partner of the enzyme–inhibitor pair. Since many of the third domains interact with many interesting serine proteinases, the work offers huge numbers of pairs to study. In order to reduce the number of pairs to a manageable number, we focus only on those enzymes whose three-dimensional structures are known at high resolution, but even that leaves us with a lot to do.

The determination of the three-dimensional structure of complexes is carried out primarily in the laboratory of Prof.

N. M. G. James at Edmonton, Alberta, Canada. They have already determined the structures of complexes of OMTKY3 with *Streptomyces griseus* proteinase B (Fujinaga et al., 1982; Read et al., 1983) and with chymotrypsin [Read et al. (1984) and unpublished results].

The main thrust of our own work on the demand side is to determine enzyme–inhibitor equilibrium constants for essentially all of the available variants and to analyze the data in terms of our sequence to reactivity algorithm (Laskowski et al., 1981, 1983, 1985; Empie & Laskowski, 1982; Park, 1985). We currently have measured 62 (variants) × 5 (enzymes) = 310 highly accurate equilibrium constants.

As this work continues to (almost) keep up with the rate of new variant production, we have expanded it to the systematic study of rate constants for association and dissociation (Empie & Laskowski, 1982; Ardelt & Laskowski, 1985; Beatty and Laskowski, unpublished results). The salient conclusion is that association rate constants are very roughly the same for almost all enzyme–inhibitor pairs, while dissociation rate constants are responsible for most of the variation in equilibrium constants.

For the properties of the single domains, the central aspect is X-ray crystallography done by the Huber group at Martinsried. They have already determined the structure of OMJPQ3S and OMJPQ3G mixture (Weber et al., 1981; Papamokos et al., 1982) and more recently OMSVP3 (Bode et al., 1985). More are on the way. In our own laboratory we also study the values of K_{hyd}, the equilibrium constant for reactive site (the bond broken by the solid line in Table II) (Ardelt & Laskowski, 1983, 1984). The surprise here was that K_{hyd} values have a broad range (currently 0.4 to ~30) among the set of variants. Similarly (Otlewski & Laskowski, 1985), we have started domain stability studies. In spectroscopy we have had extensive collaboration on NMR (Markley et al., 1984, 1985). We have done some ultraviolet spectroscopy of the tyrosyls (March, 1980), and we have supplied samples to R. G. Spiro of Princeton for laser Raman studies. We supplied samples for ^{252}Cf plasma desorption mass spectrometry to P. Cotter of Johns Hopkins University. We have an extensive collaboration with S. Smith-Gill (National Institutes of Health) on the use of this variant set in better defining the interaction of protein antigens with monoclonal antibodies. Finally, we use our variants to develop both ion-exchange (Ardelt & Laskowski, 1982) and reverse-phase HPLC (Lin et al., 1985) separations. R. Lewontin (Harvard University) is using a part of our set to calibrate disc gel electrophoresis.

Finally, the sequences are good in themselves. They were used to define the hypervariability problem, but they can also be used (although this was not done yet) to construct phylogenies. We collaborate here with W. Fitch (Wisconsin) and C. Sibley (Yale).

In examining this set of proteins solely as an object for sequence to protein property studies (ignoring for the moment the major conclusion of this paper about unusual evolution), one can ask whether for a comparably high amount of effort one could not do as well or better by starting with a single small protein with an exceptionally well-determined three-dimensional structure, e.g., bovine pancreatic trypsin inhibitor (Kunitz), and then produce variants either by some fully synthetic or semisynthetic procedure or by site-specific mutagenesis. The clear advantage of such an approach would be that we would get the variants we specify. In all probability this would often reduce to making all the possible replacements at some significant position [e.g., Estell et al. (1985)]. This would allow for a very systematic evaluation of the resultant data. The present approach, while far less tidy, has two advantages. First, it provides us with a guarantee that each product protein folds well and possesses a modicum of stability. Ovomucoid is present in egg whites at high concentration, and both our data and simple thought lead to the conclusion that a minimal requirement for its acceptability to a species is proper folding. Thus, we avoid making domains that fail to fold properly. The second advantage of natural variants is the element of surprise. Were we to control variant choice we would undoubtedly concentrate on a narrower range of positions to substitute than we were forced to examine by using this natural set. Several very interesting observations—the special interaction of Arg^{21} (P'_3) with chymotrypsin or the effect of residues other than Gly at position P'_{14} on the interaction with *Streptomyces griseus* proteinase B (Park, 1985)—probably would not have been discovered in an investigator-initiated variant set. We are thus forced to admit that occasionally we would have been less shrewd than nature in choosing the best variant. On the other hand, occasionally we would have done better as we hope to do in the future by controlling the type of the change. As the choice does not have to be made, we opt for the best of both worlds: a large set of natural variants, which provides information on what is allowed to vary and suggests some unanticipated questions augmented by a large set of artificial variants obtained by design. Ultimately, it is clear that very many variants are needed to unravel the complexities of protein chemistry.

ACKNOWLEDGMENTS

We thank the following for their generous gifts of eggs; without them there would be no paper: Bronx Zoo, Brookfield Zoo, Cheyenne Mountain Zoological Park, Colorado Division of Wildlife, Columbian Park Zoo (Lafayette, IN), Dallas Zoo, Denver Zoological Gardens, Fukuoka Municipal Zoo, Indianapolis Zoological Society, International Crane Foundation, Jasper-Pulaski Game Preserve, Lincoln Park Zoological Gardens, Mergus Farms, Marilyn Millberger, Mesker Park Zoo, Osaka Municipal Tennoji Zoo, Phoenix Zoo, San Antonio Zoological Gardens and Aquarium, Zoological Gardens of San Diego, San Diego Sea World, Smithsonian Institution National Zoological Park, St. Louis Zoological Park, Charles Swenson, Tokyo Ueno Zoo, Dr. Jozef Witkowski and Ludwik Tomialojc, Woodland Park Zoological Gardens, Mosquito Creek Game Bird Farm, Delbert Gue, Mickey Ollson, Edith Pierson, Udo Schultze, William A. Sumner, Traina Brothers, and Marvin Trice. We are also grateful to many colleagues both in biochemistry and in ornithology for their interest, discussions, and help in facilitating egg acquisitions: Wolfram Bode (Martinsried), Siwo DeKloet (Tallahassee), Walter Fitch (Wisconsin), Robert Huber (Martinsried), Atsushi Ikai (Tokyo), Tokuji Ikenaka (Osaka), Sadaaki Iwanaga (Fukeoka), Michael James (Edmonton), Paul Johnsgard (Nebraska), Randy Read (Edmonton), Charles Sibley (Yale), Allan Wilson (Berkeley), and Tadeusz Wilusz (Wroclaw). Finally, we thank members of our own group who over the years made major contributions to this research. They are Chauce Bogard, Cheng-wu Chi, William Finkenstadt, Clyde Kelly, Tiao-Yin Lin, Carl March, Jacek Otlewski, and Nicholas Warne.

APPENDIX

s1 Northwest Farm, Inc., Box 3003, Portland, OR 97208

s2 St. Louis Zoo, Forest Park, MO 63110; Stephen R. Willie, Curator

s3 Mickey Ollson, Route 1, Box 152, Glendale, AZ 83301

s4 San Diego Zoo, Zoological Society of San Diego, San Diego, CA 92112; Arthur C. Risser, Jr., Curator

s5 Charles G. Sibley, Peabody Museum, P.O. Box 6666, New Haven, CT 06511-8161
s6 Allan C. Wilson, Department of Biochemistry, University of California, Berkeley, CA 94720
s7 Tony Alexander, Rt. 5, Box 385A, Theodore, AL 36582
s8 Mesker Park Zoo, Bement Avenue, Evansville, IN 47712; Alice Kuhn, Curator
s9 Phoenix Zoo, P.O. Box 5155, Phoenix, AZ 85010; Kristin Buhl, Senior Keeper
s10 Osaka Municipal Tennoji Zoo, Osaka, Japan (facilitated by T. Ikenaka, Department of Chemistry, Faculty of Science, Osaka University, 1-1 Machikaneyama, Toyonaka (T560), Japan)
s11 Brookfield Zoo, Chicago, IL 60513; Dennis Decourcey, Curator
s12 M. Millberger, 37 Elmwood Place, Oberlin, OH 44074
s13 Siwo R. DeKloet, 4355 Cripple Creek Drive, Tallahassee, FL 32308
s14 Vance Grannis, Jr., 9249 Barnes Avenue, Inver Grove Heights, MN 55075
s15 Mary Popple, Mosquito Creek Game Farm, Rt. 2, Box 149, Weyauwega, WI 54983
s16 Denver Zoo, East 23rd Avenue & Steele, Denver, CO 80216; Ed Schmidt, Curator
s17 Columbian Park Zoo, Lafayette, IN 47904; Debbie Burrows, Kirk Burnet
s18 J. Witkowski and L. Tomialojc, Institute of Zoology, and T. Wilusz, Institute of Biochemistry, University of Wroclaw, Wroclaw, Poland
s19 Mel Quale, 144 Cedar Point, Redwood Falls, MN 56283
s20 Rollin Packer, Rt. 2, Box 40, Wood River, NE 68883
s21 Charles Nugent, Rt. 2, 72474 Mica Road, Kimbolton, OH 43749
s22 San Diego Sea World, 1720 South Shore Road, San Diego, CA 92109; Scot Drieschman, Curator
s23 Smithsonian Institution/National Zoological Park, 3001 Connecticut Avenue, N.W., Washington, DC 20008
s24 Udo Schulze, P.O. Box 46, State Hill, NY 10973
s25 William A. Sumner, 6015 High Point Road, Greensboro, NC 27407
s26 Woodland Park Zoo, 5500 Phinney Avenue North, Seattle, WA 98103; Wally English, Curator
s27 Tokyo Ueno Zoo, Tokyo, Japan (facilitated by A. Ikai, Department of Biophysics and Biochemistry, Facult of Science, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan)
s28 Ed Benhardt, Twin Spruce Aviary, P.O. Box 311, Reardan, WA 99029
s29 Irvin Hickman, Lebo Rt., Box 87A, West Plains, MO 65775
s30 Clait E. Braun, Colorado Division of Wildlife, Wildlife Research Center, 317 W. Prospect Road, Ft. Collins, CO 80526
s31 Manesa Company, Inc., 764 Ora Avo Drive, Vista, CA 92083
s32 Charles Mitchell, 37154 Camarillo Avenue, Barstow, CA 92311
s33 Robert J. White, Sr., Box 213 Dodswirth Lane, Cold Spring, KY 41076
s34 L. M. Schumpert Game Farm, P.O. Box 36, Port Tales, MN 88130
s35 Gallant Bevy, P.O. Box 1465, Amarillo, TX 79105
s36 Glen Graves, Rt. 1, Box 144, Dexter, NM 88230
s37 Kraft's Game Bird Farm, Route 1, Box 258, Princess Anne, MD 21853
s38 Twin Cedar Farm, Liberty, KY 42539
s39 Meadowbrook Game Birds, Richfield, PA 17086
s40 J & M Quail Farm, 1790 Cardinal Circle, Fayetteville, NC 28301
s41 Warren L. Boyd, Missouri Quail Farm, Rt. 1, Box 265, Stover, MO 65078
s42 Indianapolis Zoo, 3120 East 30th Street, Indianapolis, IN 46218; Julian Duval, Curator
s43 Robert Tybie, P.O. Box 32, Doyle, CA 96109
s44 Delbert Gue, 2701 W. Robin Lane, Phoenix, AZ 85027
s45 Oak Ridge Game Farm, R.R. 2, Gravette, AZ 72736
s46 Marvin Trice, Box 131, Orestes, IN 46063
s47 Traina Brothers, 107 Cranford Boulevard, Mastic, NY 11950
s48 Marsh Farms, 14232 Brookhurst Street, Garden Grove, CA 92643
s49 Canaan Farms, P.O. Box 879, Tombstone, AZ 85638
s50 Denton Bird Farm, 1566 Wetmore Road, Livermore, CA 94550
s51 Charles Sivelle, 41 Westcliff, Dix Hills, NY 11746
s52 Miller's Exotic Game Bird and Poultry Farm, Ionia, MO 65335
s53 Fukuoka Municipal Zoo, Fukuoka, Japan (facilitated by S. Iwanaga, Department of Biology, Faculty of Science, Kyushu University 33, Fukuoka 812, Japan)
s54 Jasper-Pulaski Game Preserve, Medaryville, IN 47957
s55 Rev. Eritt T. Darnell, Sr., Rt. 1, Box 132, Ft. Blackmore, VA 24250
s56 Frank Zeitler, 2225 West Elder, Santa Ana, CA 92704
s57 Frank Newhall, Acampo, CA 95220
s58 Martin Granica, RFD 1, Glenwood City, WI 54013
s59 Friends Bird Farm, Fredonia, KS 66736
s60 Shenandoah Farms, R.R. 1, Patoka, IN 47666
s61 Janson's Farms, P.O. Box 136, Zeeland, MI 49464
s62 Dougal House, 269 Orange Avenue, Goleta, CA 93017
s63 Fordham Quail Farm, 10520 S.W. 47 Street, Miami, FL 33165

References

Ackers, G. K., & Smith, F. R. (1985) *Annu. Rev. Biochem. 54*, 597–629.

Ardelt, W., & Laskowski, M., Jr. (1982) *Anal. Biochem. 120*, 198–203.

Ardelt, W., & Laskowski, M., Jr. (1983) *Acta Biochim. Pol. 30*, 115–126.

Ardelt, W., & Laskowski, M., Jr. (1984) *Biochemistry 23*, 3359 (Abstr).

Ardelt, W., & Laskowski, M., Jr. (1985) *Biochemistry 24*, 5313–5320.

Barrett, A. J. (1974) *Biochim. Biophys. Acta 371*, 52–62.

Bode, W., Epp, E., Huber, R., Laskowski, M., Jr., & Ardelt, W. (1985) *Eur. J. Biochem. 147*, 387–395.

Bogard, W. C., Jr., Kato, I., & Laskowski, M., Jr. (1980) *J. Biol. Chem. 255*, 6569–6574.

Bolognesi, M., Gatti, G., Menegatti, E., Guarneri, M., Marquart, M., Papamokos, E., & Huber, R. (1982) *J. Mol. Biol. 162*, 839–868.

Carrell, R. W., & Travis, J. (1985) *Trends Biochem. Sci. (Pers. Ed.) 10*, 20–21.

Drapeau, G. (1976) *Methods Enzymol. 45*, 469–475.
Empie, M. W., & Laskowski, M., Jr. (1982) *Biochemistry 21*, 2274–2284.
Estell, D. A., & Laskowski, M., Jr. (1980) *Biochemistry 19*, 124–131.
Estell, D. A., Grayear, T. P., & Wells, J. A. (1985) *J. Biol. Chem. 260*, 6518–6521.
Feeney, R. E. (1970) in *Proceedings of the International Research Conference on Proteinase Inhibitors* (Fritz, H., & Tschesche, H., Eds.) pp 162–168, de Gruyter, Berlin and New York.
Feeney, R. E., & Allison, R. G. (1969) *Evolutionary Biochemistry of Proteins*, Wiley-Interscience, New York.
Fossum, K., & Whitaker, J. R. (1968) *Arch. Biochem. Biophys. 125*, 367–375.
Fujinaga, M., Read, R. J., Sielecki, A., Ardelt, W., Laskowski, M., Jr., & James, M. N. G. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4868–4872.
Greene, L. J., Pubois, M. H., & Bartelt, D. C. (1976) *Methods Enzymol. 45*, 813–825.
Gruson, E. S. (1976) *Checklist of the World's Birds*, Quadrangle, The New York Times Co., New York.
Hirono, S., Akagawa, H., Iitaka, Y., & Mitsui, Y. (1984) *J. Mol. Biol. 178*, 389–443.
Hoy, T. G., Ferdinand, W., & Harrison, P. M. (1974) *Int. J. Pept. Protein Res. 6*, 121–140.
Hunkapiller, M. W., & Hood, L. E. (1978) *Biochemistry 17*, 2124–2133.
Hunt, L. T., & Dayhoff, M. O. (1980) *Biochem. Biophys. Res. Commun. 95*, 864–871.
Ikai, A., Kitamoto, T., & Nishigai, M. (1983) *J. Biochem. (Tokyo) 93*, 121–127.
Kato, I., & Kohr, W. J. (1978) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 37*, 1337 (Abstr.).
Kato, I., Schrode, J., Wilson, K. A., & Laskowski, M., Jr. (1976) *Protides Biol. Fluids 23*, 235–243.
Kato, I., Kohr, W. J., & Laskowski, M., Jr. (1978) *Proc. FEBS Meet. 47*, 197–206.
Kato, I., Schrode, J., Kohr, W. J., & Laskowski, M., Jr. (1986) *Biochemistry* (preceding paper in this issue).
Kingston, I. B., & Anderson, S. (1986) *Biochem. J. 233*, 443–450.
Klapper, D. G., Wilde, C. E., III, & Capra, J. D. (1978) *Anal. Biochem. 85*, 126–131.
Laskowski, M., Jr., & Kato, I. (1980) *Annu. Rev. Biochem. 49*, 593–626.
Laskowski, M., Jr., Kato, I., Kohr, W. J., March, C. J., & Bogard, W. C. (1980) *Protides Biol. Fluids 28*, 123–128.
Laskowski, M., Jr., Empie, M. W., Kato, I., Kohr, W. J., Ardelt, W., Bogard, W. C., Jr., Weber, E., Papamokos, E., Bode, W., & 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., & Wieczorek, M. (1983) in *Proteinase Inhibitors: Medical and Biological Aspects* (Katanuma, N., Umezawa, H., & Holtzer, H., Eds.) pp 55–68, Japan Scientific Societies, Tokyo, and Springer Verlag, Berlin.
Lin, T.-Y., Cook, J. A., & Laskowski, M., Jr. (1985) *Biochemistry 24*, 3385 (Abstr.).
Linderstrom-Lang, K., & Ottesen, M. (1949) *C. R. Trav. Lab. Carlsberg Ser. Chem. 26*, 403–442.
Lineweaver, H., & Murray, C. W. (1947) *J. Biol. Chem. 171*, 565–581.
Lottspeich, F. (1980) *Hoppe-Seyler's Z. Physiol. Chem. 361*, 1829–1834.
March, C. J. (1980) Ph.D. Thesis, Purdue University.
Markley, J. L., Westler, M. W., Chan, T.-M., Kojiro, C. L., & Ulrich, E. L. (1984) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 43*, 2648–2656.
Markley, J. L., Croll, D. H., Krishnamoorthi, R., Ortiz-Polo, G., Westler, W. M., & Bogard, W. C., Jr., & Laskowski, M., Jr. (1986) *J. Cell. Biochem. 30*, 291–309.
Matsushima, K. (1958) *Science (Washington, D.C.) 127*, 1178–1179.
Mueller-Esterl, W., Fritz, H., Kellermann, J., Lottspeich, F., Machleidt, W., & Turk, V. (1985) *FEBS Lett. 191*, 221–226.
Nagase, H., Harris, E. D., Jr., Woessner, J. F., & Brew, K. (1983) *J. Biol. Chem. 258*, 7481–7489.
Otlewski, J., & Laskowski, M., Jr. (1985) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 44*, 1807.
Papamokos, E., Weber, E., Bode, W., Huber, R., Empie, M. W., Kato, I., & Laskowski, M., Jr. (1982) *J. Mol. Biol. 158*, 515–537.
Park, S. J. (1985) Ph.D. Thesis, Purdue University.
Prager, E. M., & Wilson, A. C. (1976) *J. Mol. Evol. 9*, 45–57.
Read, R. J., Fujinaga, M., Sielecki, A. R., & James, M. N. G. (1983) *Biochemistry 22*, 4420–4433.
Read, R., Fujinaga, M., Sielecki, A., Ardelt, W., Laskowski, M., Jr., & James, M. (1984) *Acta Crystallogr., Sect. A.: Found. Crystallogr. 40* (Suppl.), 50–51.
Rhodes, M. B., Bennett, N., & Feeney, R. E. (1960) *J. Biol. Chem. 235*, 1686–1693.
Satoch, S., Kurecki, T., Kress, L. F., & Laskowski, M., Sr. (1979) *Biochem. Biophys. Res. Commun. 86*, 130–137.
Schechter, I., & Berger, M. (1967) *Biochem. Biophys. Res. Commun. 27*, 157–162.
Schwabe, C., Anastasi, A., Crow, H., McDonald, J. K., & Barrett, A. J. (1984) *Biochem. J. 217*, 813–817.
Sibley, C. G., & Ahlquist, J. E. (1972) *A Comparative Study of the Egg White Proteins of Non-Passerine Birds*, Peabody Museum of Natural History, Yale University, New Haven.
Stein, J. P., Catterall, J. F., Kristo, P., Means, A. R., & O'Malley, B. W. (1980) *Cell (Cambridge, Mass.) 21*, 681–687.
Tarr, G. E., Beecher, J. F., Bell, M., & McKean, D. (1978) *Anal. Biochem. 84*, 622–627.
Tashiro, M., & Laskowski, M., Jr. (1983) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 42*, 1953.
Travis, J., & Salvesen, G. S. (1983) *Annu. Rev. Biochem. 52*, 655–709.
Turk, V., Brzin, J., Longer, M., Ritonja, A., Eroplin, M., Borchardt, V., & Machleid, W. (1983) *Hoppe-Seyler's Z. Physiol. Chem. 364*, 1487–1496.
Weber, E., Papamokos, E., Bode, W., Huber, R., Kato, I., & Laskowski, M., Jr. (1981) *J. Mol. Biol. 158*, 109–123.
Wieczorek, M., & Laskowski, M., Jr. (1983) *Biochemistry 22*, 2630–2636.
Wieczorek, M., Tashiro, M., & Laskowski, M., Jr. (1983) *Biochemistry 22*, 21A.
Wright, H. T. (1984) *J. Biol. Chem. 259*, 14335–14336.
Wu, T. T., & Kabat, Z. A. (1970) *J. Exp. Med. 132*, 211–250.