Padlan, E. A., & Love, W. E. (1974) *J. Biol. Chem. 249*, 4067-4078.
Parkhurst, L. J., Sima, P., & Gross, D. J. (1980) *Biochemistry 19*, 2688-2692.
Seamonds, B. (1971) in *Probes of Structure and Function of Macromolecules and Membranes* (Chance, R., Yonetani, T., & Mildvan, A. S., Eds.) Vol. II, pp 317-320, Academic, New York.
Seamonds, B., & Forster, R. E. (1972) *Am. J. Physiol. 223*, 734-738.
Seamonds, B., Forster, R. E., & George, P. (1971) *J. Biol. Chem. 246*, 5391-5397.
Seamonds, B., McCray, J. A., Parkhurst, L. J., & Smith, P. D. (1976) *J. Biol. Chem. 254*, 2579-2583.
Smith, R. M., & Martell, A. E. (1976) *Critical Stability Constants*, Vol. 4, pp 26, Plenum, New York.
Vega-Catalan, F. J., Odeyemi, O. J., & Okonjo, K. O. (1986) *J. Biol. Chem. 261*, 10576-10581.
Ver Ploeg, D. A., Cordes, H., & Gurd, F. R. N. (1971) *J. Biol. Chem. 246*, 2725-2733.

A Histidine-Rich Protein from the Vitellaria of the Liver Fluke *Fasciola hepatica*†

J. Herbert Waite*,‡ and Allison C. Rice-Ficht§

Marine Biology/Biochemistry Program, College of Marine Studies, University of Delaware, Lewes, Delaware 19958, and Department of Medical Biochemistry and Genetics, Texas A&M University, College Station, Texas 77843

Received January 26, 1989; Revised Manuscript Received March 28, 1989

ABSTRACT: The vitellaria are an extensive network of glandular cells and ducts distributed throughout the peripheral tissues of the liver fluke *Fasciola hepatica*. Eggshell precursor proteins are produced and stockpiled in the vitelline cells of mature flukes. Vitelline protein C has an extraordinary composition: the amino acid 3,4-dihydroxyphenyl-L-alanine (DOPA) and histidine each comprise about 20% of the residues, while glycine represents 41-42% in all variants of what appears to be a microheterogeneous protein family. Protein C has an apparent molecular weight of 16 000-17 000 by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Although the protein appears homogeneous following polyacrylamide gel electrophoresis in Tris-glycine with SDS and in acetic acid-urea, electrophoresis in borate, however, suggests that the vitelline protein consists of four or more closely related proteins weighing from 16 000 to 18 500. Isoelectric focusing of the protein family in the presence of 8 M urea resolves only two species having p*I* values of 6.89 and 6.99. A single N-terminus having the sequence H-H-W-D-G-DOPA-G-DOPA-G was detected. The primary structure of vitelline protein C is characterized by a repeated motif consisting of $(G\text{-}X)_n$, where X is Ser, DOPA, or His. Most of the His occurs as G-H repeats in a pepsin-resistant fragment of the protein. Previously, a 31-kDa protein, representing up to 6% of the total protein in the fluke, was reported [Waite, J. H., & Rice-Ficht, A. (1987) *Biochemistry 26*, 7819-7825] to contain significant levels of DOPA. In other respects, however, it is distinct from vitelline protein C. Present studies suggest the existence of at least one other distinct DOPA-containing protein in the vitellaria.

The formation of eggshells in helminths has long captivated the attention of parasitologists [see reviews by Smyth and Clegg (1959) and Cordingley (1987)]. This is due in part (i) to the large proportion (20-30%) of the total energy budget of helminths devoted to eggshell production (Wharton, 1983), (ii) to pathological complications caused by the entrapment of helminth eggs in host tissues (Malek, 1980), and (iii) to the extraordinary chemical and physical stability of the eggshells, which are not unlike well-tanned leather bags (Stephenson, 1947). Despite all the attention, however, it is ironic that very little is actually known about the biochemistry of helminth eggshells. The protein precursors for trematode eggshells are generally thought to be synthesized in the vitelline cells that, when mature, migrate from the vitellaria through specialized conducting tubules to Mehli's gland and the proximal uterus. There, following some as yet unidentified biochemical cue, the cells release globules of eggshell precursors that coalesce to form the nascent eggshell around the fertilized ovum (Irwin & Threadgold, 1970). In recent years, researchers have speculated that the expression of gender-specific genes in blood flukes (*Schistosoma mansoni*) and gonad-specific genes in the liver fluke (*Fasciola hepatica*) is linked to putative eggshell precursor proteins (Johnson et al., 1987; Köster et al., 1988; Kunz et al., 1987; Bobek et al., 1986, 1988; Zurita et al., 1987). The gene sequences code for proteins especially rich in glycine, tyrosine, aspartic acid, histidine, and lysine. Recently, a protein representing perhaps 6% of the total protein in mature *F. hepatica* was purified from the vitellaria. This protein (M_r 31 000) was distinctive in containing up to 10 residues/100 of 3,4-dihydroxyphenyl-L-alanine (DOPA)[1] and having a composition generally reminiscent of that of *F. hepatica* eggshells, i.e., high levels of Asp/Asn, Gly, and Lys (Waite & Rice-Ficht, 1987). In the present paper, we describe

† This research was supported by grants from the National Institute of Allergenic and Infectious Diseases and the Office of Naval Research.
‡ University of Delaware.
§ Texas A&M University.

[1] Abbreviations: Bis, *N*,*N*'-methylenebis(acrylamide); DOPA, 3,4-dihydroxyphenyl-L-alanine; EDTA, ethylenediaminetetraacetic acid; HPLC, high-performance liquid chromatography; PAGE, polyacrylamide gel electrophoresis; SDS, sodium dodecyl sulfate; Tris, tris(hydroxymethyl)aminomethane; vpC, vitelline protein C.

0006-2960/89/0428-6104$01.50/0

another DOPA-containing protein from *F. hepatica* vitellaria having an unusually high level of His and Gly. While the function of this protein in eggshell formation is not known, it does ostensibly resemble another putative eggshell protein rich in histidine and aromatic amino acids recently described in *S. mansoni* (Johnson et al., 1987).

Materials and Methods

Extraction and Purification of Eggshell Precursor Proteins. Liver flukes were removed from condemned fresh bovine livers at the abattoir in Sealy, TX. The flukes were briefly rinsed in buffered saline (0.9% NaCl) and frozen in lots of 30 in liquid nitrogen for storage. Whole flukes were stained for 1,2-benzenediols by soaking first for 20 min in 2 mL of 0.5 N HCl, then for 20 min in an equal volume of 1.45 M sodium nitrite with 0.4 M sodium molybdate, and, finally, for another 20 min in 2 mL of 1 N NaOH (Arnow, 1937). This procedure produces a brilliant red (λ_{max} = 500 nm) in the presence of monosubstituted 1,2-benzenediols for which it is very specific (Waite & Tanzer, 1981). Prior to extraction, the deep-frozen material was thawed on ice, and flukes were spread out flat on a glass plate, frozen at −20 °C, and carefully dissected to separate the vitellaria. In bulk preparations, vitellaria were homogenized in 50-mL ground glass tissue grinders with neutral salt buffer [0.15 M NaCl, 0.05 M tris(hydroxymethyl)aminomethane hydrochloride buffer (Tris-HCl) at pH 7.5, 1 mM phenylmethanesulfonyl fluoride, 10 mM *N*-ethylmaleimide, 25 mM ethylenediaminetetraacetic acid (EDTA), and 1 mM potassium cyanide]. The homogenate was centrifuged at 5000*g* for 10 min, and the pellet was redispersed in more neutral salt buffer and centrifuged as before. The second pellet was rehomogenized in cold 5% (v/v) acetic acid with 4 M urea, 1 mM iodoacetamide, and 10 μM leupeptin (Boehringer-Mannheim, Indianapolis, IN) and centrifuged for 20 min at 35000*g*. Supernatants were frozen at −20 °C until further use. When supernatant extracts corresponding to 15–20 g of liver flukes (wet weight) had been accumulated, these were thawed and pooled. The pooled volumes were brought to 10% (w/v) ammonium sulfate at 4 °C for 30 min. The resulting precipitate was removed by centrifugation for 20 min at 5000*g*, and the supernatant was brought to 20% (w/v) ammonium sulfate at 10 °C for 60 min and centrifuged as before to remove the precipitate. The supernatant was dialyzed against 100–200 volumes of cold 2.5% acetic acid by using tubing with a molecular weight cutoff of 1000 (Spectrum Industries, Los Angeles, CA). For unknown reasons, the protein of interest precipitated out of solution during dialysis and could be easily harvested by light centrifugation at 10000*g* for 30 min. This precipitate was redissolved in 8 M urea in 5% acetic acid.

For protein extractions from the vitellaria of individual flukes, an abbreviated procedure was used: Vitellaria were dissected as described earlier and homogenized in 1.5 mL of 5% acetic acid containing 4 M urea, 1 mM iodoacetamide, and 10 μM leupeptin by using a 2-mL hand-held glass tissue grinder (Kontes, Vineland, NJ). The homogenate was spun for 10 min in a microcentrifuge (Eppendorf, Brinkman Instruments, NY) at 15000*g*, and the resulting supernatant (1.5 mL) was applied directly to the C_8 reversed-phase HPLC described under Purification and Characterization.

Purification and Characterization. The vitelline protein C was purified from the dialysis-precipitated material described above by C_8 reversed-phase (250 mm × 4.6 mm, Brownlee RP-300) HPLC using a gradient of acetonitrile (0–30%) in double-distilled water with 13 mM trifluoroacetic acid (flow rate, 1 mL/min) and lyophilized at −80 °C for 12 h. Protein hydrolysis was carried out in 7 M HCl with 10% trifluoroacetic acid and 1% phenol in vacuo at 155 °C for 22, 45, and 66 min (Tsugita et al., 1987). Amino acid composition of hydrolyzed protein was determined following flash evaporation of the acid using a single-column Beckman 6300 autoanalyzer. Corrections were made for the loss of Ser, Thr, and DOPA during hydrolysis. Tryptophan content was separately determined following hydrolysis in 4 N methanesulfonic acid (Simpson et al., 1972) and by acid–borate difference spectrophotometry of the protein (Waite, 1984). The latter was done by estimating the concentration of DOPA in a 0.01% (w/v) solution of protein C from difference absorbance spectra in 0.1 M sodium borate (pH 8.2) and 0.1 M HCl, respectively ($\Delta\epsilon_{292}$ = 3200 M^{-1} cm^{-1}). Trp was assumed to be roughly equivalent to

$$(A - [\text{DOPA}]\epsilon_1)/\epsilon_2$$

where A is taken as the absorbance of protein in 0.1 M HCl at 282 nm, [DOPA] is the concentration of DOPA, ϵ_1 is the molar extinction coefficient of DOPA in acid at 282 nm (2700 cm^{-1}), and ϵ_2 is the molar extinction coefficient of Trp in acid at 282 nm (5550 cm^{-1}). N-Terminal amino acid sequence of peptides was determined by automated Edman degradation with the use of an Applied Biosystems Model 470A (Foster City, CA) gas-phase sequenator (Ozols, 1986).

Electrophoresis. Routine electrophoresis was done on polyacrylamide gels [7% acrylamide and 0.2% *N,N'*-methylenebis(acrylamide)] containing 5% acetic acid with 2 M urea (Waite & Benedict, 1984). This system was convenient here because it lends itself readily to staining either for protein with Serva Blue R (Serva Fine Chemicals, Westbury, NY) or for DOPA with the Arnow reagents (Waite & Benedict, 1984). Molecular weights were determined by polyacrylamide gel electrophoresis (PAGE) in the presence of sodium dodecyl sulfate (SDS) using either the discontinuous Tris–glycine (Hoefer Scientific Catalogue, San Francisco, p 131), or continuous Tris–borate buffer system (Grierson, 1983). For the latter, the polyacrylamide gels contained 10% acrylamide, 2.5% bis, and 4 M urea, whereas the running buffer (gel and reservoir) consisted of 90 mM Tris, 90 mM boric acid, 2.5 mM EDTA (disodium salt), and 0.1% SDS at a pH of 8.2. Isoelectric focusing was performed in the presence of 8 M urea with ampholytes (Servalytes, Serva Fine Chemicals) in the pH range 6–9 according to a procedure suggested by the instruction manual for protean II slab gels (Bio-Rad Chemical Division, Richmond, CA, Bulletin 87-0133).

Protein Blotting. Electrophoretic transfer of proteins from polyacrylamide gels (acid–urea or sodium dodecyl sulfate) onto nitrocellulose (Sigma N-0639) was accomplished by using techniques outlined by Towbin et al. (1979). The detection of bound primary antibody was by alkaline phosphatase coupled goat anti-rabbit antiserum at a dilution of 1:3000 (Boehringer-Mannheim). Antisera to the pure 31-kDa *F. hepatica* eggshell precursor and preimmune serum were contractually prepared in three female New Zealand rabbits by Cambridge Research Biochemicals (Cambridge, England) according to standard immunization schedules and harvest bleeding at 12 weeks and diluted 1:500 before use. Optimal dilution range was experimentally determined by diluting antigen serially on nitrocellulose dot blots.

Preparation of Peptides. Digestion of protein was carried out at a protein-to-enzyme weight ratio of 50:1 in 0.2 M sodium borate (pH 8.2) with 3.5 M urea and 0.1 mM $CaCl_2$ under 40 psi of nitrogen and constant stirring at 27 °C. Enzymes used were trypsin, pepsin (both Boehringer-Mannheim), and α-chymotrypsin (U.S. Biochemical Corp., Cleveland, OH).

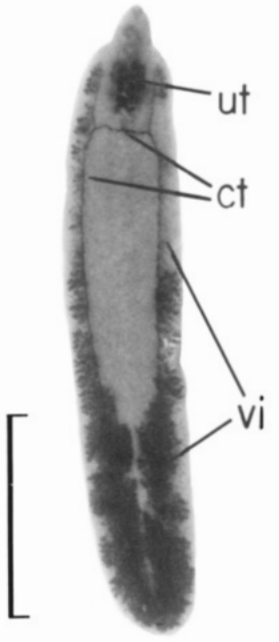

FIGURE 1: Dorsal view of a mature liver fluke (*Fasciola* sp.) stained for DOPA with nitrite–molybdate. The extensive vitellaria (vi), which are colorless in unstained flukes, stain bright red, as do the conducting tubules (ct) and uterus (ut). The latter contains thousands of eggs each encapsulated by a leathery DOPA-rich eggshell. Bar is 1 cm.

In the case of pepsin, the digestion was done in 5% acetic acid and 3 M urea. The peptides were purified by HPLC on a C_8 reversed-phase column (Brownlee RP-300) HPLC using a gradient of acetonitrile (0–30%) in double-distilled water with 13 mM trifluoroacetic acid.

RESULTS

Vitelline DOPA–Proteins. The vitellaria of mature *F. hepatica* are distributed over the peripheral two-thirds of the fluke and stain intensely with nitrite–molybdate (Figure 1). Nitrite–molybdate (also known as Arnow's reagent) is specific for the determination of 1,2-benzenediols, which in *F. hepatica* appear to occur primarily as protein-bound DOPA residues (Waite & Rice-Ficht, 1987). The DOPA-containing proteins are preferentially extractable at acid pH; however, this useful property is compromised by highly active acid proteases that are coextracted with the DOPA–proteins. Proteins extracted with and without added leupeptin were stained in parallel for protein and protein-bound DOPA following separation by acid–urea polyacrylamide gel electrophoresis (Figure 2A). Note, in particular, the paucity of higher molecular weight proteins in the samples prepared without leupeptin. In contrast, when the inhibitor is present, there are at least three distinct proteins (A–C) that stain for DOPA as well as protein. Of these, proteins A and B are largely degraded within 30 min of extraction in the absence of leupeptin. The fate of protein C is less clear because it becomes progressively obscured by degradation products of similar mobility (Figure 2A). Other studies based on HPLC of acid-extracted vitellaria, however, reflect that protein C is also degraded (data not shown). The most prominent of the DOPA-containing vitelline proteins, protein B, was described previously (Waite & Rice-Ficht, 1987). Protein A has yet to be characterized.

Purification of Vitelline Protein C. We isolated vitelline protein C, which corresponds to the fastest moving of the three on acid–urea gels. This protein is readily purified from acid extracts containing leupeptin by the steps outlined under Materials and Methods. Curiously, protein C coprecipitates

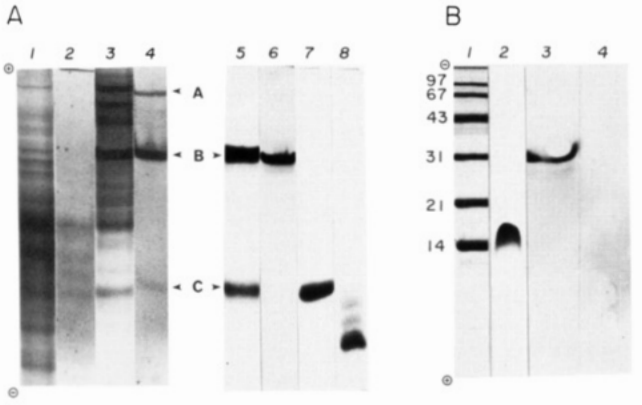

FIGURE 2: Acid–urea and SDS–PAGE of the vitelline proteins A–C of *F. hepatica*. (A) Acid–urea PAGE: (lane 1) extract of vitellaria without leupeptin, 30 μg of protein stained with Serva Blue R; (lane 2) same but stained for DOPA, 60 μg; (lane 3) with leupeptin, 30 μg; (lane 4) same as lane 3 but stained for DOPA, 63 μg; (lane 5) dialysis precipitate, 10 μg; (lane 6) HPLC-purified vitelline protein B (31 000), 6 μg; (lane 7) HPLC-purified vitelline protein C, 5 μg; (lane 8) pepsin-digested protein C, 4 μg. The gels in lanes 5–7 were all stained for protein with Serva Blue R. (B) Sodium dodecyl sulfate–PAGE: (lane 1) standard protein mixture containing lysozyme, soybean trypsin inhibitor, carbonic anhydrase, ovalbumin, bovine serum albumin, and phosphorylase *b*, 40 μg; (lane 2) protein C, 10 μg (protein stain is Serva Blue R); (lane 3) immunoblot of protein B from SDS–PAGE using antiserum against vitelline protein B; (lane 4) blot of protein C and anti-B antiserum. Preimmune serum did not significantly bind to either protein.

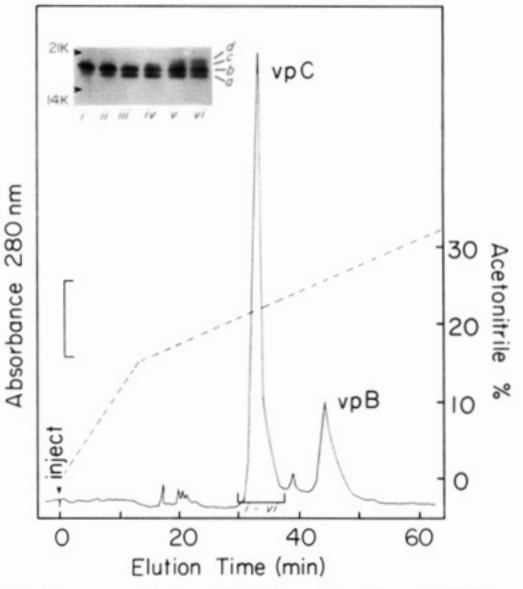

FIGURE 3: C_8 reversed-phase HPLC purification of vitelline protein C from the dialysis precipitate. Conditions of chromatography are given under Methods and Materials. Bar represents 0.1 adsorbance unit at 280 nm. Inset shows SDS–PAGE in Tris–borate of the fractions labeled i–vi under protein C. The gels were stained with molybdate–nitrite for DOPA; molecular weight markers are indicated in the margin.

almost exclusively with vitelline protein B during dialysis against 2.5% acetic acid of the fraction soluble in 20% ammonium sulfate (Figure 2A). From this it is easily separated by reversed-phase HPLC on C_8 silica (Figure 3). The purity of the protein following HPLC is suggested by a sharp band on acid–urea gels (Figure 2) and an unambiguous amino-terminal sequence (Table III). Discontinuous SDS–PAGE with Tris–glycine seems less than ideal for determining the molecular weight of the protein (Figure 2B) because the protein consistently smears in the shape of a "scowl" over an apparent molecular weight range of 14 000–18 000 with the

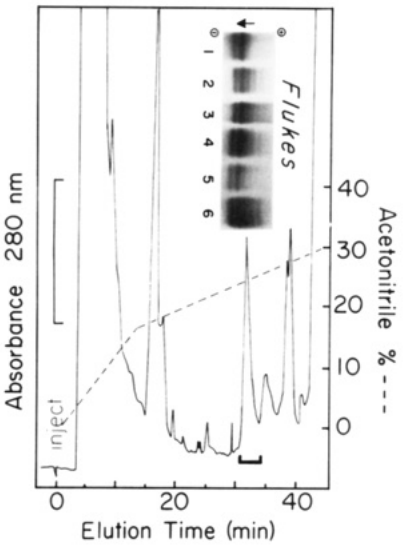

FIGURE 4: Acid-extracted vitellaria of single liver flukes subjected to C_8 reversed-phase HPLC. The bar at left represents 0.05 absorbance unit. The fractions containing the peak at 32 min were pooled (bracket) and run on SDS–PAGE with Tris–borate following lyophilization. Inset shows the microheterogeneity of vitelline protein C derived from six individual flukes. Protein load/lane is approximately 30–40 μg, and the stain is the molybdate–nitrite method for DOPA.

bulk concentrated at 17 000. SDS–PAGE in the presence of Tris–borate, however, provides a crisp resolution of at least four distinct bands staining for protein and DOPA in fractions of the HPLC-purified vitelline protein C (Figure 3, inset). These have apparent molecular weights ranging from 16 000 to 18 500 with the most prominent weighing 17 500. The difference in the ability of these two electrophoretic methods to resolve vitelline protein C may be due to borate, which at pH 8–9 strongly complexes DOPA groups (Waite, 1984), thus perhaps preventing them from interacting with the polyacrylamide. A further caveat with regard to DOPA–borate complexes seems appropriate: The microheterogeneity observed in protein C in Tris–borate may not necessarily reflect differences in molecular weight since the protein standards and vitelline protein C do not interact with the buffer in the same way. Notwithstanding this, the microheterogeneity apparent in bulk preparations of vitelline protein C following SDS gel electrophoresis with Tris–borate is also evident in protein C prepared from six individual flukes (Figure 4). Most of these show the four bands in proportions similar to those seen in the bulk preparations. Only two bands could be detected following isoelectric focusing in the presence of 8 M urea of any of the fractions in Figures 3 and 4: the observed p*I*'s were 6.89 and 6.99 (95% CL ± 0.07) (data not shown) with the latter predominating, usually at a ratio of about 1:2. Protein C is not likely to be derived from protein B by proteolytic degradation since it does not cross-react with polyclonal rabbit anti-vitelline protein B antiserum (Figure 2B) and has a distinctive composition.

Characterization of Vitelline Protein C. The amino acid composition of the protein in six fractions under peak in Figure 3 is given in Table I and is remarkably consistent, excepting the slightly higher levels of Asx, Glx, and Tyr in the latter fractions. DOPA, His, and Gly predominate throughout. Trp is also evident; acid–borate difference spectrophotometry of intact protein C suggests a ratio of one residue of Trp to every eight to nine residues of DOPA. This is in agreement with the amino acid composition following hydrolysis in methanesulfonic acid. Incidentally, the specific extinction coefficients ($E^{1\%}$) of vitelline protein C in 0.1 M HCl at 280 nm and 0.1 M borate pH 8.5 at 287.5 nm were found to be 46 and 67 cm^{-1}, respectively. Borate is the buffer of choice at the higher pH since the protein is not significantly soluble in the pH range 6–9 in other buffers. The composition of vitelline protein C resembles that of protein B only in having a high content of DOPA and Gly (Table I). However, it is different in having Trp, very high His, no Lys, and lower Asp. Gly, His, and DOPA account for over 80% of the residues in protein C, whereas in protein B, these same residues account for only 32% of the residues. Minimum molecular weights can be estimated from amino acid compositions (Segel, 1976). On the assumption that one residue of tyrosine is present per molecule of protein in fractions v and vi (Table I), a minimum M_r of 17 000 is obtained.

The prevailing motif in the primary structure of vitelline protein C would appear to be $(Gly\text{-}X)_n$, where X is Ser, His, or DOPA (Table II). At the amino terminus, there is a His-His-Trp- sequence, followed by an Asp, two Gly-DOPA repeats, two Gly-Ser-Gly-DOPA repeats, two repeats of Gly-Ser, and possibly Gly-His-His. Pepsin digestion of protein C results in a quantitative conversion to a large fragment (P-5), consisting principally of Gly, His, and DOPA, and four smaller ones, three of which are dominated by Gly, Ser, Asp, DOPA, and His (P-2, P-3, and P-4); the fourth (P-1) seems dominated by Gly and DOPA (Figure 5; Table III). The fragment P-5 was purified by reversed-phase HPLC (Figure 5) and found

Table I: Amino Acid Composition of the Fractions i–vi of HPLC-Purified Vitelline Protein C and Vitelline Protein B (Residues/1000) (See Figure 4)[a]

	protein C							
amino acid	i	ii	iii	iv	v	vi	res/M_r[b]	protein B[c]
Asx	35	34	30	32	37	41	4–5	140
Thr	1	1	1	1	2	2	0	18
Ser	77	76	82	78	79	79	11	52
Glx	4	3	3	3	6	6	0–1	83
Pro	21	20	20	19	24	23	3	16
Gly	414	412	425	415	416	405	57	165
Ala	3	2	3	3	3	4	0	69
$^1/_2$-Cys	0	0	0	0	0	0	0	0
Val	1	0	1	1	4	3	0	9
Met	1	1	2	2	0	0	0	23
Ile	1	0	0	1	0	0	0	5
Leu	0	0	0	0	0	0	0	38
DOPA	198	204	198	208	183	182	26–27	106
Tyr	2	2	2	4	7	11	0–1	21
Phe	0	1	0	1	1	2	0	38
His	211	212	210	208	209	206	28	45
Lys	0	0	0	0	1	0	0	120
Trp	17	16	12	14	15	18	3	0
Arg	14	15	12	13	14	16	2	60

[a] Values represent the mean of three runs with a standard deviation of ±5%. [b] Estimated number of residues per molecule weighing 17 000. [c] Waite and Rice-Ficht (1987).

Table II: Sequencer Analysis of the N-Terminus of 250 pmol of Vitelline Protein C of *F. hepatica*

cycle	amino acid	yield (pmol)	cycle	amino acid	yield (pmol)
1	His	198	13	Gly	169
2	His	220	14	Ser	48
3	Trp	132	15	Gly	143
4	Asp	233	16	DOPA	41
5	Gly	166	17	Gly	141
6	DOPA	140	18	Ser	51
7	Gly	216	19	Gly	99
8	DOPA	100	20	Ser	47
9	Gly	195	21	Gly	84
10	Ser	53	22	Ser, Gly, His	41, 89, 14
11	Gly	186	23	Gly, Ser, His	80, 24, 19
12	DOPA	60			

Table III: Amino Acid Composition of Peptide Fragments Prepared by Pepsin Digestion of Vitelline Protein C (Residues/1000)[a]

amino acid	P-1	P-2	P-3	P-4	P-5
Asx	61	100 (1)	70 (1)	65 (1)	33 (2)
Thr	2	17	5	2	4
Ser	62	116 (1)	136 (2)	200 (4)	27 (2)
Glx	57	24	7	5	6
Pro	4	9	3	8	33 (2)
Gly	392	341 (4)	377 (6)	354 (6)	405 (27)
Ala	12	19	8	15	6
Val	3	5	3	0	1
Met	1	0	3	0	0
Ile	7	0	2	3	1
Leu	0	0	0	0	0
DOPA	325	140 (2)	200 (3)	190 (4)	195 (13)
Tyr	10	6	11	3	12 (1)
Phe	10	7	1	1	1
His	50	139 (2)	118 (2)	92 (2)	252 (18)
Lys	1	3	0	2	0
Trp	0	70 (1)	58 (1)	58 (1)	0
Arg	3	4	2	4	22 (2)
total residues	(?)	(11)	(15)	(18)	(65)

[a] Each value represents the mean of three runs with a standard deviation of ±5%. Parenthetical numbers represent residues per peptide.

Table IV: Sequencer Analysis of the N-Terminus of 600 pmol of Pepsin-Resistant Fragment P-5 of the Vitelline Protein C of *F. hepatica*

cycle	amino acid	yield	cycle	amino acid	yield
1	DOPA	460	10	Gly	164
2	Gly	503	11	His	97
3	His	121	12	Gly	151
4	Gly	520	13	His	83
5	DOPA	157	14	Gly	146
6	Gly	294	15	His	71
7	His	108	16	Gly	123
8	Gly	202	17	His	56
9	His	110	18	(Arg, His, Gly)	(17, 45, 112)

to consist of two tandem repeats of DOPA-G-H-G followed by at least three repeats of H-G (Table IV). Although the other peptides have yet to be sequenced, the results obtained so far suggest that most or all of the Asp, Trp, and Ser is confined to a terminal end of the protein. Moreover, the amino acid composition of peptic fractions P-2–4 is so similar to the composition of the N-terminus that perhaps the 20 or so amino acids of the N-terminus are repeated three times. The greater part of the protein consists of an inner domain that is His rich and pepsin resistant. Protein C was not detectably digested following a 12-h exposure to trypsin, suggesting that Arg may be protected by Pro or is at the C-terminus. Chymotrypsin did digest the protein somewhat, although the fragments produced (all large) were not sufficiently well separated by HPLC to afford characterization.

Discussion

F. hepatica produces up to 2500 eggs daily (Björkman & Thorsell, 1963). Each of these eggs is normally encapsulated within a leathery protein eggshell that is derived from droplets or globules that are released by vitelline cells in the uterus of the trematode. Due to the insolubility and low antigenicity of trematode eggshells washed clean of soluble surface antigens (von Lichtenberg & Raslavicius, 1967), it has not yet been technically possible to determine how many and which proteins contribute to eggshell structure. Previously, we argued an eggshell precursor role for vitelline protein B on the basis of a rather circumstantial "guilt by association" due to the presence of DOPA in both protein B and immature eggshells (Waite & Rice-Ficht, 1987). Results obtained here indicate

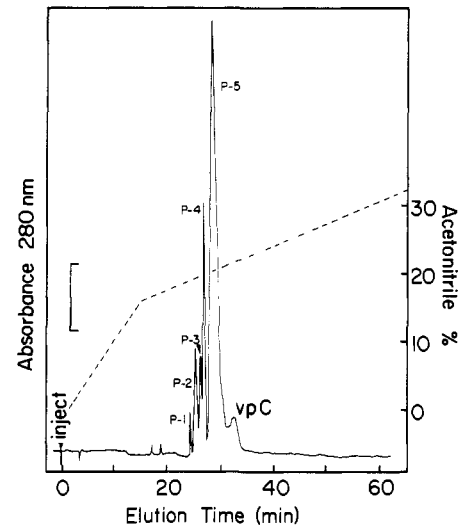

FIGURE 5: Pepsin-resistant fragments of protein C. Conditions of chromatography are given under Materials and Methods. Bar represents 0.1 adsorbance unit at 280 nm.

that there are at least three vitelline proteins containing DOPA. Thus, protein B has no monopoly on DOPA, although it clearly is the most abundant of the three. All three may be eggshell precursors. The isolation of these putative eggshell precursor proteins is not trivial. While they are only soluble at acid pH, they are particularly susceptible to hydrolysis at this pH by acid proteases from the gut of *F. hepatica* (Simpkin et al., 1980). Since the digestive diverticulum interdigitates somewhat with the vitellaria, even careful dissection of the vitellaria prior to extraction does not entirely relieve proteolysis; addition of micromolar levels of leupeptin to the acidic extraction buffer, however, does. Here we characterized a putative precursor named vitelline protein C. SDS–PAGE in the presence of borate at pH 8.2 resolved the protein into at least four components. These could represent (1) limited degradation of a single protein or (2) subtle protein polymorphism due to allelic variants or gene families. The first possibility seems unlikely in view of the precautions taken with protease inhibitors and the clean N-terminal sequence, although a C-terminal exopeptidase-catalyzed degradation cannot yet be categorically discounted. The existence of protein polymorphism in silkworm eggshells is widely known (Regier et al., 1978) and appears to be emerging as a feature in trematode eggshells as well. In this regard, Bobek et al. (1988) recently reported finding three homologous putative eggshell precursor protein genes in *S. mansoni* that differed from one another by only three to four nucleotides.

Like vitelline protein B, protein C has an amino acid composition rich in glycine and DOPA. The actual proportion of DOPA in protein C is about twice that of vitelline protein B, but there is also an additional qualitative difference in that the conversion of tyrosine to DOPA is almost 100% complete in protein C. In contrast, vitelline protein B is always isolated containing an apparent continuum of molecules with tyrosyl-to-DOPA conversions ranging from no conversion to about 85% (Waite & Rice-Ficht, 1987). In addition, protein C has high levels of histidine (21 residues/100), whereas protein B is distinguished by high aspartate/asparagine (14 residues/100) and lysine (12 residues/100) (Waite & Rice-Ficht, 1987). Both precursors are characterized by sequences rich in DOPA and glycine (Figure 6). In fact, DOPA in both proteins occurs with Gly-rich flanking sequences. The DNA sequences of putative eggshell precursor genes in *F. hepatica* (Zurita et al., 1987) and *S. mansoni* (Bobek et al., 1986, 1988; Johnson et

A VPC P-5 residues #1-18

Dopa-G-H-G-Dopa-G-H-G-H-G-H-G-H-G-H-G-H-(R)

B BHMWK residues #459-483

G-H-G-H-Q-K-Q-H-G-L-G-H-G-H-K-H-G-H-G-H-G-H-G-K

FIGURE 6: Comparison of the primary sequences of the P-5 peptide of vitelline protein C–VPCP-5 (A) and bovine high molecular weight kininogen–BHMWK (B) taken from Kitamura et al. (1983).

al., 1987; Simpson et al., 1987; Kunz et al., 1987) also reflect Gly flanking sequences on one or both sides of most of the Tyr. This suggests that the Tyr-to-DOPA conversion may be carried out by the same or related tyrosyl hydroxylases that may depend on the proximity of Gly for substrate recognition.

With the exception of the N-terminal and penultimate His, much of the histidine in protein C seems clustered in the pepsin-resistant fragment (P-5) where it occurs in G-H runs. This is distinct from the poly[L-histidine] domain detected by Johnson et al. (1987) near the carboxy terminus in the gene sequence of a putative eggshell precursor (M_r 48 000) in *S. mansoni*; however, the two proteins may yet have a similar function. His-rich proteins have been discovered in blood serum (Morgan, 1985; Takagaki et al., 1985; Koide et al., 1986), polychaete fangs (Voss-Foucart et al., 1973), the knob proteins and cytoplasmic granules of the human malarial parasite *Plasmodium* (Kemp et al., 1987; Kilejian et al., 1986), and others. In many of these, the histidine has been implicated in the chelation of metals, particularly Cu and Zn. The Gly- and His-rich proteins of the fangs of marine polychaetes, for example, apparently mineralize with zinc containing up to 3% of the metal by dry weight (Bryan & Gibbs, 1979). Moreover, His-Trp at the amino terminus of proteins has strong Ni(II) and Co(II) binding properties (Smith et al., 1987). Unfortunately, nothing is known about the metal content of trematode eggshells. The closest reported sequence homology to vitelline protein C would appear to be the 10 or so repeats of (Gly)-His-Gly in the His-rich domain of the high molecular weight I kininogen shown in Figure 6 (Kitamura et al., 1983), the function of which is still unclear. Sugo et al. (1980) have recently determined that the His-rich domain of kininogen is involved in the molecule's binding to kaolin and other negatively charged surfaces in vitro. This they suggest may be important in the kininogen-accelerated surface-induced activation of blood factor XII, thus triggering coagulation. Eggshell precursors also coalesce and coagulate on a negatively charged lipoprotein from Mehli's gland in *F. hepatica* (Clegg, 1965), but the role (if any) of vitelline protein C in this process is completely unknown.

The presence of DOPA in the vitelline proteins is thought to be related to the stabilization of the eggshells by quinone tanning. Catechol oxidase, an enzyme capable of converting DOPA residues to *o*-quinonyl groups, reportedly occurs in the vitelline globules and in the eggshell (Smyth & Clegg, 1959). It is believed that cross-links arise from the nucleophilic addition of histidinyl and lysyl groups to the *o*-quinones (Schaefer et al., 1987).

ACKNOWLEDGMENTS

We thank George Korza of the University of Connecticut Protein Sequencing Facility for the gas-phase sequencing of the proteins and Kathleen Little and Will Eppes for their help in dissecting vitellaria.

REFERENCES

Arnow, L. E. (1937) *J. Biol. Chem. 118*, 531–537.

Björkman, N., & Thorsell, W. (1963) *Exp. Cell Res. 32*, 153–156.

Bobek, L., Rekosh, D. M., van Keulen, H., & LoVerde, P. T. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5544–5548.

Bobek, L., Rekosh, D. M., & LoVerde, P. T. (1988) *Mol. Cell. Biol. 8*, 3008–3016.

Bryan, G. W., & Gibbs, P. E. (1979) *J. Mar. Biol. Assoc. U.K. 59*, 969–973.

Clegg, J. A. (1965) *Ann. N.Y. Acad. Sci. 118*, 969–986.

Cordingley, J. S. (1987) *Parasitol. Today 3*, 341–344.

Grierson, D. (1983) in *Gel Electrophoresis of Nucleic Acids* (Rickwood, D., & Hames, D. B., Eds.) p 15, IRL Press, Oxford, U.K.

Irwin, S. W. B., & Threadgold, L. T. (1970) *Exp. Parasitol. 31*, 321–331.

Johnson, K. S., Taylor, D. W., & Cordingley, J. S. (1987) *Mol. Biochem. Parasitol. 22*, 89–100.

Kemp, D. J., Coppel, R. L., & Anders, R. F. (1987) *Annu. Rev. Microbiol. 41*, 181–208.

Kilejian, A., Sharma, Y. D., Karoui, H., & Naslund, L. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7938–7941.

Kitamura, N., Takagi, Y., Furuto, S., Tanaka, T., Nawa, H., & Nakanishi, S. (1983) *Nature (London) 305*, 545–549.

Koide, T., Foster, D., Yoshitake, S., & Davie, E. W. (1986) *Biochemistry 25*, 2220–2225.

Köster, B., Dargatz, H., Schröder, J., Hirzmann, J., Haarrmann, C., Symmons, P., & Kunz, W. (1988) *Mol. Biochem. Parasitol. 31*, 183–198.

Kunz, W., Opatz, K., Finken, M., & Symmons, P. (1987) *Nucleic Acids Res. 15*, 5894.

Malek, E. A. (1980) in *Snail-Transmitted Parasitic Diseases*, Vol. 1, pp 261–288, CRC, Boca Raton, FL.

Morgan, W. T. (1985) *Biochemistry 24*, 1496–1501.

Ozols, J. (1986) *J. Biol. Chem. 261*, 3965–3979.

Schaefer, J., Kramer, K. J., Garbow, J. R., Jacob, G. S., Stejskal, E. O., Hopkins, T. L., & Speirs, R. D. (1987) *Science (Washington, D.C.) 235*, 1200–1204.

Segel, I. (1976) *Biochemical Calculations*, 2nd ed., pp 111–112, Wiley, New York.

Simpkin, K. G. Chapman, C. R., & Coles, G. C. (1980) *Exp. Parasitol. 49*, 281–287.

Simpson, A. J. G., Chaudry, M., Knight, M., Kelly, C., Rumjanek, F., Martin, S., & Smithers, S. R. (1987) *Mol. Biochem. Parasitol. 22*, 169–176.

Simpson, R. J., Neuberger, M. R., & Liu, T.-Y. (1976) *J. Biol. Chem. 251*, 1936.

Smith, M. C., Furman, T. C., & Pidgeon, C. (1987) *Inorg. Chem. 26*, 1965–1969.

Smyth, J. D. (1954) *Q. J. Microsc. Sci. 95*, 139–154.

Smyth, J. D., & Clegg, J. A. (1959) *Exp. Parasitol. 8*, 286–323.

Stephenson, W. (1947) *Parasitology 38*, 128–139.

Sugo, T., Ikari, N., Kato, H., Iwanaga, S., & Fujii, S. (1980) *Biochemistry 19*, 3215–3220.

Takagaki, Y., Kitamura, N., & Nakanishi, S. (1985) *J. Biol. Chem. 260*, 8601–8609.

Towbin, H., Staehelin, T., & Gordon, J. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4350–4354.

Tsugita, A., Uchida, T., Mewes, H. W., & Ataka, T. (1987) *J. Biochem. 102*, 1593–1597.

von Lichtenberg, F., & Raslavicius, P. (1967) *Lab. Invest. 16*, 892–904.

Voss-Foucart, M. F., Fonze-Vignaux, M. T., & Jeuniaux, C. (1973) *Biochem. Syst. 1*, 119–122.

Waite, J. H. (1984) *Ann. Chem. 56*, 1935-1939.
Waite, J. H., & Tanzer, M. L. (1981) *Anal. Biochem. 111*, 131-136.
Waite, J. H., & Benedict, C. V. (1984) *Methods Enzymol. 107*, 391-413.
Waite, J. H., & Rice-Ficht, A. C. (1987) *Biochemistry 26*, 7819-7825.
Wharton, D. A. (1983) *Parasitology 86*, 85-97.
Zurita, M., Bieber, D., Ringold, G., & Mansour, T. E. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 2340-2344.

Amino Acid Sequence of Bovine Angiogenin†

Michael D. Bond and Daniel J. Strydom*

Center for Biochemical and Biophysical Sciences and Medicine and Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115

Received December 29, 1988; Revised Manuscript Received March 27, 1989

ABSTRACT: The amino acid sequence and disulfide bridges of bovine plasma derived angiogenin were determined by sequencer analysis of the intact protein and fragments derived by enzymatic and chemical digestion. Bovine angiogenin is a single-chain protein of 125 amino acids; it contains six cysteines and has a calculated molecular weight of 14 595. In contrast to the human protein its amino terminus is unblocked. It has the following sequence: H_2N-Ala[1]-Gln-Asp-Asp-Tyr-Arg-Tyr-Ile-His-Phe[10]-Leu-Thr-Gln-His-Tyr-Asp-Ala-Lys-Pro-Lys[20]-Gly-Arg-Asn-Asp-Glu-Tyr-Cys-Phe-Asn-Met[30]-Met-Lys-Asn-Arg-Arg-Leu-Thr-Arg-Pro-Cys[40]-Lys-Asp-Arg-Asn-Thr-Phe-Ile-His-Gly-Asn[50]-Lys-Asn-Asp-Ile-Lys-Ala-Ile-Cys-Glu-Asp[60]-Arg-Asn-Gly-Gln-Pro-Tyr-Arg-Gly-Asp-Leu[70]-Arg-Ile-Ser-Lys-Ser-Glu-Phe-Gln-Ile-Thr[80]-Ile-Cys-Lys-His-Lys-Gly-Gly-Ser-Ser-Arg[90]-Pro-Pro-Cys-Arg-Tyr-Gly-Ala-Thr-Glu-Asp[100]-Ser-Arg-Val-Ile-Val-Val-Gly-Cys-Glu-Asn[110]-Gly-Leu-Pro-Val-His-Phe-Asp-Glu-Ser-Phe[120]-Ile-Thr-Pro-Arg-His-OH. Disulfide bonds link Cys(27)-Cys(82), Cys(40)-Cys(93), and Cys(58)-Cys(108). Bovine angiogenin is 64% identical with human angiogenin; like the human protein, it is homologous to the pancreatic ribonucleases, with conservation of active site residues. Two regions, 6-22 and 65-75, are highly conserved between the angiogenins but are significantly different from those of the ribonucleases, suggesting a possible role in the molecules' biological activity.

Angiogenin is a nonmitogenic protein that stimulates angiogenesis. It was first isolated from human tumor cell conditioned media (Fett et al., 1985) and later from normal human serum (Shapiro et al., 1987). Its amino acid sequence is homologous to that of RNase A[1] (Strydom et al., 1985), and it contains many of the important structural and active site residues critical to RNase activity, including the catalytically essential His-12, Lys-41, and His-119 (RNase A numbering). Indeed, angiogenin displays ribonucleolytic activity (Shapiro et al., 1986) and also binds tightly to a protein RNase inhibitor from placenta (Shapiro & Vallee, 1987).

Despite the similarities between the two molecules, only angiogenin is able to induce blood vessel growth. Therefore, this activity derives at least in part from some structural feature(s) unique to angiogenin. One means of identifying such features is to examine the expected sequence variation in angiogenins from different species. Individual residues and/or regions involved in angiogenesis should be conserved and should vary, perhaps dramatically, from the corresponding residues in RNase A.

To investigate this possibility, we have isolated (Bond & Vallee, 1988) and sequenced angiogenin from bovine plasma. The isolation was made possible by the recent development of a convenient assay system (Bond, 1988). This paper describes the first sequence determination of an angiogenin other than the human protein, namely, the bovine one, examines correspondences and differences between the two angiogenins, and points to specific regions of the molecule that might be important for biological activity.

Materials and Methods

Angiogenin was isolated from bovine plasma as described (Bond & Vallee, 1988). The methodology of sequence determination is given in the supplementary material.

Nomenclature

Unless otherwise indicated, the numbering of amino acid residues is based on the sequence of bovine angiogenin. Tryptic peptides are numbered according to their location in the sequence, from T1, the amino-terminal peptide, to T18, the carboxyl-terminal peptide.

Results

Overlaps and Complete Sequence. Figure 1 summarizes the sequence of bovine angiogenin and the major pieces of

† This work was supported by funds from Hoechst, A.G., under an agreement with Harvard University.

* Address correspondence to this author at the Center for Biochemical and Biophysical Sciences and Medicine, Harvard Medical School, 250 Longwood Ave., Boston, MA 02115.

[1] Abbreviations: RNase(s), ribonuclease(s); RNase A, pancreatic ribonuclease A; RNase S-protein and S-peptide, subtilisin-generated fragments of RNase A containing residues 21-124 and 1-20, respectively; C18, octadecylsilane; HPLC, high-performance liquid chromatography; Tris, tris(hydroxymethyl)aminomethane; TFA, trifluoroacetic acid; DTT, dithiothreitol; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; PTH, phenylthiohydantoin.

0006-2960/89/0428-6110$01.50/0