Alignment and Partial Structural Analysis of the Cyanogen Bromide Fragments from Yeast Inorganic Pyrophosphatase†

Richard Sterner, Claudia Noyes, and Robert L. Heinrikson*

ABSTRACT: Yeast inorganic pyrophosphatase (EC 3.6.1.1, pyrophosphate phosphohydrolase) was cleaved by cyanogen bromide in 70% formic acid for 48 hr at room temperature, and the reaction mixture was subsequently reduced and S-carboxymethylated. Gel filtration of the cleavage products on a column of Sephadex G-50 developed in 10% formic acid gave three major components, the compositions of which accounted for the total amino acid content of the protein monomer. The alignment of these fragments generated by cleavage at the two methionine residues in the pyrophosphatase subunit was based upon partial sequence analysis at the amino and carboxyl termini of the peptides and the intact enzyme. In addition to these fragments, three minor components were isolated in a highly purified form by the gel filtration step. These were shown to be derived from the internal CNBr fragment, possibly as a result of cleavages during prolonged exposure to 70% formic acid. Subsequent analysis of the six cleavage products by automated Edman degradation and digestion with carboxypeptidase A allowed placement of 108 of the 270 residues in the subunit of yeast inorganic pyrophosphatase.

An earlier publication from this laboratory (Heinrikson *et al.*, 1973) presented (1) evidence that yeast inorganic pyrophosphatase is composed of two identical subunits with an approximate mol wt of 30,000 and (2) limited sequence analysis of the protein monomer by automated Edman degradation and treatment with carboxypeptidase A. These studies were carried out preparatory to the complete covalent structural analysis of yeast inorganic pyrophosphatase, the preliminary results of which are reported here. The present communication describes methods for cleavage of the enzyme with CNBr and for the isolation of pure fragments in high yield. Compositional and partial sequence analyses of the peptides thus obtained have permitted the alignment of CNBr fragments in the native pyrophosphatase monomer, and have provided overlapping sequences in the vicinity of the cleavage sites.

Materials

Inorganic pyrophosphatase was prepared from Red Star baker's yeast by a modification of the Kunitz (1952) method developed recently by Cooperman *et al.* (1973). Acrylamide gel electrophoresis indicated a purity of greater than 95%, and the specific activity of this preparation was 43 Kunitz units per mg. Routine yields of the pure enzyme obtained from 50-lb lots of yeast varied between 450 and 500 mg.

Iodoacetic acid, purchased from the Matheson Co., Inc., East Rutherford, N. J., was recrystallized from a mixture of diethyl ether and petroleum ether prior to use. Cyanogen bromide was a product of Pierce Chemical Co., and guanidine hydrochloride (Ultra Pure grade) was obtained from Schwarz BioResearch. Carboxypeptidase A (code COA DFP-1DA) and trypsin (code TRTPCK-1FA) were purchased from Worthington. All reagents employed in the automated Edman degradation of polypeptides were highly purified products from Beckman Instruments, Palo Alto, Calif. Fluorescamine (Udenfriend *et al.*, 1972) was kindly donated by Roche Diagnostics; the reagent is presently available from commercial sources. All other reagents were of the purest commercially available grade.

Methods

Amino Acid Analysis. Amino acid compositional data were obtained by automated ion exchange chromatography on a single column according to the general procedures of Spackman *et al.* (1958), with a Durrum D-500 analyzer. Analyses of the various peptides were performed by hydrolysis in 6 N HCl for 24 hr at 112° *in vacuo*. Special precautions were observed in the procedure to ensure the highest possible yield of carboxymethylcysteine (Crestfield *et al.*, 1963; Moore and Stein, 1963). As in earlier work (Heinrikson *et al.*, 1973), the quantitation obtained in these analyses served as a reference for the precise determination of polypeptide concentrations in samples subjected to quantitative end group procedures. Tryptophan analyses were performed on samples hydrolyzed with methanesulfonic acid as described by Liu and Chang (1971). Hydrolyses of phenylthiohydantoin (PhNCS[1]) derivatives back to the parent amino acids were performed either in 6 N HCl containing 0.05% 2-mercaptoethanol (Van Orden and Carpenter, 1964), or in HI as described by Smithies *et al.* (1971). The latter procedure was especially useful in affording higher yields of alanine from serine and of carboxymethylcysteine and α-aminobutyrate from threonine. All samples subjected to analysis on the Durrum D-500 analyzer were centrifuged prior to application.

High-Voltage Paper Electrophoresis. This procedure was

† From the Department of Biochemistry and the Franklin McLean Memorial Research Institute, University of Chicago, Chicago, Illinois 60637. *Received August 30, 1973.* The Franklin McLean Memorial Research Institute is operated by the University of Chicago for the U. S. Atomic Energy Commission. This work was supported by Grants GB-29098 and GB-27506 from the National Science Foundation, Grant NP-43 from the American Cancer Society, Training Grant GM-424 from the National Institute of General Medical Sciences (to R. S.), and Research Career Development Award 5K04 GM-11164 (to R. L. H.) from the U. S. Public Health Service.

[1] Abbreviations used are: PhNCS, phenylthiohydantoin; Quadrol, the registered trademark of the Wyandotte Chemical Corp., Wyandotte, Mich., for the compound *N,N,N′,N′*-tetrakis(2-hydroxypropyl)-ethylenediamine.

used to determine peptide purity, to detect peptides during the course of column separation procedures, and to monitor the course of tryptic hydrolysis. Electrophoresis was run on sheets of Whatman No. 1 chromatography paper in 0.61 N formic acid (pH 2.1) for 45 min at 2950 V (52 V/cm). Positions of the peptides were detected by staining both with ninhydrin–$CdCl_2$ reagent (Dreyer and Bynum, 1967) and with fluorescamine (Udenfriend *et al.*, 1972). In the latter procedure, electrophoretograms were sprayed with 0.2 M *N*-ethylmorpholine acetate (pH 8.5) until completely damp, and then immediately sprayed with a solution containing 5 mg of fluorescamine in 30 ml of acetone. The fluorescent spots were detected by inspection of the dried papers with a short-wavelength ultraviolet lamp. The sensitivity of the fluorescamine procedure was found to be five–ten times greater than that utilizing ninhydrin–$CdCl_2$. Peptides containing arginine were detected by a modified Sakaguchi staining procedure (Irreverre, 1965).

Polyacrylamide Gel Electrophoresis. Both the size and purity of various cleavage products derived from yeast pyrophosphatase were assessed by polyacrylamide gel electrophoresis in systems containing sodium dodecyl sulfate and 2-mercaptoethanol, according to the method of Weber and Osborn (1969).

Succinylation of Fragments. Succinylation of amino groups was performed in several cases to enhance the solubility of peptides, and also to effect selective cleavage by trypsin at arginine residues. Peptide (1–10 mg) was dissolved in 2 ml of a solution containing 6 N guanidine hydrochloride, 0.003 M EDTA, and 0.1 M Tris-HCl, pH 8.0, and a tenfold excess by weight of solid succinic anhydride was slowly added. The pH of the reaction mixture was maintained between 7 and 8 throughout the course of succinylation by the dropwise addition of 0.1 N NaOH. Succinylated peptides were desalted by gel filtration on a column of Sephadex G-25 (2 × 50 cm) developed in 0.01 M NH_4OH, and the pooled fractions comprising each were lyophilized.

Isolation of the Cleavage Products Obtained in the Reaction of Inorganic Pyrophosphatase with CNBr. Prior to cleavage with CNBr, 180 mg of native pyrophosphatase was dissolved in 30 ml of a solution containing 5% 2-mercaptoethanol and 0.01 M Tris-HCl (pH 7.0). After standing overnight at 4°, the solution was dialyzed exhaustively against water and lyophilized. The protein treated in this fashion was dissolved in 70% formic acid (15 mg of protein/ml) and a 100-fold molar excess of CNBr relative to methionine content was added. The reaction vessel was sealed tightly under N_2 and the reaction was allowed to proceed for 24 hr in the dark at room temperature. At this time, a second quantity of CNBr equal to the first was added, and the reaction was carried on for an additional 24 hr as before. Prior to separation of the CNBr fragments, the reaction mixture was rotary evaporated and lyophilized to remove solvents and volatile reaction by-products, and the peptides were subjected to reduction and carboxymethylation as described previously (Heinrikson *et al.*, 1973). Upon completion of the reaction, the mixture was brought to 10% formic acid in a total volume of 10 ml, and the solution was added directly to a column (5 × 200 cm) of Sephadex G-50 equilibrated in 10% formic acid. Fractions of 18.6 ml were collected at a flow rate of 60 ml/hr, and the elution of fragments was monitored by absorbance at 280 nm. Based on the elution profile thus obtained, appropriate fractions were pooled and lyophilized.

Sequence Analysis. The determination of carboxyl-terminal sequences in various CNBr fragments was made by the kinetics and stoichiometry of release of residues during digestion with carboxypeptidase A as described by Ambler (1967). Hydrolysis mixtures incubated at 37° contained pre-washed and solubilized carboxypeptidase A (Ambler, 1967) and substrate (50–125 nmol) in a weight ratio of 1:40, dissolved in 1 ml of 0.2 N *N*-ethylmorpholine acetate buffer (pH 8.0). Aliquots of 200 μl were removed at various times during the course of digestion, deproteinated, and analyzed as described earlier (Heinrikson *et al.*, 1973). It should be stressed that removal of protein prior to analysis is an important step to safeguard the capillary column of the Durrum D-500 analyzer.

Procedures for automated Edman degradation of intact pyrophosphatase and CNBr fragments derived therefrom were essentially the same as described earlier (Heinrikson *et al.*, 1973). A Beckman protein–peptide Sequencer (Model 890-B) was employed for these operations with programs supplied by the manufacturer. In addition to the Quadrol program (program D-XI), procedures using the volatile buffer dimethylallylamine were applied in the sequence analysis of large fragments (dimethylallylamine–protein program 050771) and peptides (dimethylallylamine–peptide program 090872). The PhNCS derivatives sequentially liberated by these procedures were identified as such and quantitated by gas chromatography (Pisano and Bronzert, 1969) on a Beckman gas chromatographic unit (GC-45). Identifications were further confirmed by thin-layer chromatography (Jeppsson and Sjöquist, 1967). In some cases, which will be specified for each individual fragment, confirmation and quantitation were afforded by hydrolytic back-conversion of the PhNCS derivative to the parent amino acid.

Results

As a prelude to a preparative scale CNBr cleavage of yeast inorganic pyrophosphatase, a pilot experiment was performed in which 15 mg (500 nmol) of the native protein was treated with CNBr, as described earlier in this paper. Without any attempt to separate the cleavage products, the reaction mixture was lyophilized and the residue was redissolved in 20% acetic acid. An aliquot corresponding to 100 nmol of the original protein was subjected to a single cycle of automated Edman degradation with the Quadrol program. The elution profile obtained by gas chromatography of the PhNCS derivatives from the first degradative stage is presented in Figure 1. In addition to threonine, which is known to be the amino-terminal residue in the enzyme (Eifler *et al.*, 1972; Heinrikson *et al.*, 1973), almost equimolar quantities of alanine and valine were observed as a result of the CNBr cleavage. The yields of alanine and valine were 65 and 70%, respectively, typical of the percentage yields which we normally observe with this approach (Kingdon *et al.*, 1972; Heinrikson *et al.*, 1973). These initial observations suggested the presence of a Met-Val and a Met-Ala sequence in the interior of the pyrophosphatase subunit. Furthermore, they provide strong evidence in support of the contention made earlier (Heinrikson *et al.*, 1973) that the native enzyme is composed of two identical subunits, each of which contains two methionine residues.

Cyanogen bromide cleavage was next performed on a preparative scale, and the fragments were separated by gel filtration. A sample consisting of 180 mg of the native enzyme (6 μmol of protein monomer) was reduced, cleaved with CNBr, and S-carboxymethylated as described earlier in this paper. In Figure 2 is shown the elution profile obtained by gel filtration of the reaction products on a column of Sephadex G-50 developed with 10% formic acid. Ten fractions were pooled, as

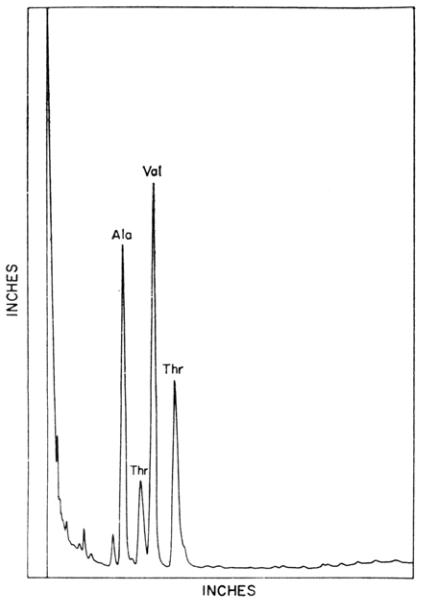

FIGURE 1: Gas chromatographic analysis of the PhNCS derivatives liberated from CNBr-cleaved pyrophosphatase (100 nmol) by a single stage of Edman degradation (*cf.* Results). The graph is an exact reproduction of the experimental data plotted as inches of pen deflection *vs.* inches of peak elution position at a fixed chart speed with reference to the point of injection. The sample contained 8% of the total PhNCS derivatives and the data were obtained at an attenuation of 1600. PhNCS derivatives are designated by the name of the parent amino acid. No additional peaks were observed after silylation of the PhNCS mixture.

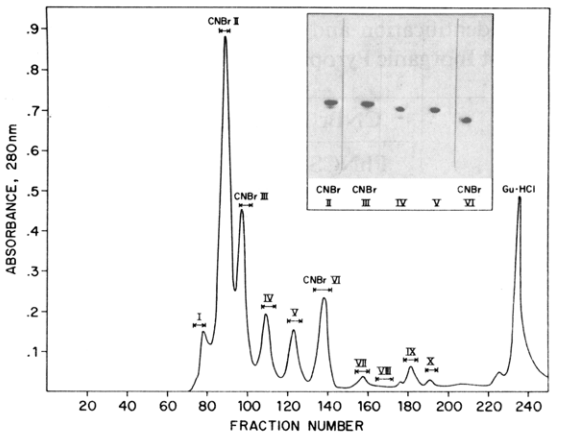

FIGURE 2: Gel filtration of the products formed by CNBr cleavage of yeast inorganic pyrophosphatase and sodium dodecyl sulfate–polyacrylamide gel patterns of five of the resulting peptides. Enzyme (180 mg) was treated with CNBr and, after reducton and S-carboxymethylation, the products were separated on a column of Sephadex G-50. Experimental details for these procedures are given in the Methods section. Various peaks on the chromatogram are designated by Roman numerals based upon their order of elution, and the arrows indicate fractions pooled in each case. Numerals prefaced by CNBr refer to peptides subsequently shown to arise solely by CNBr cleavage. Insert: Photograph of stained sodium dodecyl sulfate–polyacrylamide gels obtained by electrophoresis of 25–30 μg of the indicated fractions, as described by Weber and Osborn (1969).

indicated in Figure 2, on the basis of absorbance at 280 nm The insert in Figure 2 depicts photographs of stained sodium dodecyl sulfate–polyacrylamide gels run on fractions CNBr II, CNBr III, IV, V, and CNBr VI. The ten fractions were labeled consecutively according to their order of elution, and each was characterized in turn to determine its identity and origin. Those in which the Roman numeral is prefaced with CNBr are so designated because subsequent analysis showed them to be derived solely by CNBr cleavage. It was determined that the remaining fractions had arisen from secondary cleavages in addition to those produced by CNBr. In the following paragraphs, each fragment is discussed in regard to purity, compositional analysis, position in the polypeptide chain, and any sequence analysis derived therefrom. Table I summarizes the quantitative aspects of the automated Edman degradation as applied to all the fractions which appeared to be pure on the basis of polyacrylamide gel electrophoresis and end group analysis. Amino acid analyses of the various fractions are given in Table II.

I. Calculation of molecular weights from sodium dodecyl sulfate–polyacrylamide gel data (Weber and Osborn, 1969) for the two species in this fraction, and the small molar yield relative to other fragments, indicated that peak I consisted of incompletely cleaved and intact pyrophosphatase monomer. No further analysis of this fraction was carried out.

CNBr II. Polyacrylamide gel electrophoresis of this fragment indicated a slight degree of contamination by what appeared to be a smaller fragment (*cf.* insert, Figure 2). However, automated Edman degradation revealed alanine as the sole amino-terminal residue. CNBr II is therefore one of the fragments expected on the basis of the small-scale cleavage described earlier (Figure 1). Identification by thin-layer and gas chromatographic methods of the PhNCS derivatives afforded by automated Edman degradation (Quadrol program) yielded the following unambiguous sequence through 22 residues

Ala-Leu-Leu-Asp-Glu-Gly-Glu-Thr-Asp-Trp-Lys-
5 10
Val-Gly-Ala-Ile-Asp-Ile-Asn-Asp-Pro-Leu-Ala-
15 20

The chromatographic methods of identification and quantitation were supplemented in some cases by back-conversion of the PhNCS derivative to the original amino acid by hydrolysis with HI or HCl (*cf.* Methods). Data summarized in Table I show that the initial yield of Ala>PhNCS was 69% and the repetitive yield calculated from the quantities of Ala>PhNCS at steps 1 and 14 was 93%. Residue 8 (Thr) was confirmed by back-conversion in HI which gave α-aminobutyric acid. PhNCS derivatives 15 (Ile), 16 (Asp), 18 (Asn → Asp), 19 (Asp), and 20 (Pro) were hydrolyzed in HCl to give the parent amino acids in high yield. All of the indicated residues were confirmed by thin-layer chromatography. The agreement between these various modes of identification and the good yields obtained provide a sound basis for the sequence reported.

Amino acid compositional data presented in Table II indicate that CNBr II contains 121 amino acids. From this analysis, it was calculated that the yield of CNBr II obtained after gel filtration was 4–5 μmol (80%). To be noted in the analysis is the absolute lack of homoserine or homoserine lactone in this fragment, which indicates that CNBr II is the carboxyl-terminal peptide in the enzyme. Proof for this assertion was obtained by digestion of 125 nmol of CNBr II with carboxypeptidase A. Results identical with those reported earlier for the intact enzyme were obtained (Heinrikson *et al.*, 1973). Equimolar quantities of valine were rapidly removed, followed by serine, glycine, phenylalanine, and isoleucine, in that order.

CNBr III. Polyacrylamide gel electrophoresis in sodium dodecyl sulfate gave a single uniform band for CNBr III (Figure 2), and automated Edman degradation provided

TABLE I: Identification and Quantitation of Phenylthiohydantoins Sequentially Removed by Automated Edman Degradation from Yeast Inorganic Pyrophosphatase and Various Fragments Produced Therefrom.[a]

	CNBr II (450 nmol)		CNBr III (400 nmol)		VII (120 nmol)		Intact Subunit (600 nmol)	
Position	PhNCS Derivative	Yield (nmol)	PhNCS Derivative	Yield (nmol)	PhNCS Derivative	Yield (nmol)	PhNCS Derivative	Yield (nmol)
1	Ala	310	Val	283	Pro	75	Thr	
2	Leu	303	Val	275	Ile	75	Tyr	426
3	Leu	250	Glu	60	Asp		Thr	
4	Asp	149	Ile		Val	72	Thr	
5	Glu	117	Pro	87	Leu	66	Arg	
6	Gly		CmCys[c]	65	Gln[d]	45	Gln	
7	Glu	150	Trp		Ile	73	Ile	
8	Thr[b]	75	Thr		Gly	46	Gly	
9	Asp	96	Asn		Glu		Ala	370
10	Trp		Ala	154	Thr[b]	17	Lys	
11	Lys		Lys[d]	60	Ile	44	Asn	
12	Val	120	Leu	85	Ala	54	Thr	
13	Gly		Glu	15	Tyr	36	Leu	190
14	Ala	127	Ile	63	Thr[b]	16	Glu	121
15	Ile[d]	80	Thr[b]	38	Gly	15	Tyr	164
16	Asp[d]	82	Lys		Gln	18	Lys	
17	Ile		Glu[d]	40	Val	32	Val	190
18	Asn[d]	70	Glu[d]	41	Lys[d]	18	Tyr	155
19	Asp[d]	68	Thr[b]	30	Gln[d]	15	Ile	
20	Pro[d]	62	Leu	42	Val	21	Glu	90
21	Leu	63	Asn				Lys[d]	65
22	Ala	67					Asp[d]	70
23							Gly[d]	60
24							Lys[d]	75
25							Pro	75
26							Val	80
27							Asn	
28							Ala	80
29							Phe	70
30							Asn[d]	60
31							His[d]	50
32							Thr	
33							Ile	61
34							Pro	58
Repetitive yield[e]	93%		93%		94%		93%	

[a] Automated Edman degradation of the indicated polypeptides was performed in a Beckman Model 890-B Protein–Peptide Sequencer, with programs supplied by the manufacturer. CNBr II and intact subunit were analyzed with the Quadrol program (D-XI), CNBr III with the dimethylallylamine–protein program (050771), and VII with the dimethylallylamine–peptide program (090872). Initial amounts of samples subjected to sequence analysis are given in parentheses. The indicated PhNCS derivatives were identified by gas and thin-layer chromatographic procedures, except for Arg, His, and CmCys. [b] Identified and quantitated as α-aminobutyric acid following hydrolysis in HI. [c] Identified and quantitated as alanine following hydrolysis in HI; no serine or alanine was observed by chromatographic methods. [d] Confirmed by hydrolytic back-conversion in 6 N HCl. [e] Calculated as follows: CNBr II, Ala-1 to Ala-14; CNBr III, Val-1 to Ala-10; VII, Ile-2 to Ile-11; intact subunit, Tyr-2 to Tyr-15.

further evidence for the purity of this fragment. The data obtained by automated sequence analysis (dimethylallylamine–protein program) revealed that CNBr III is the second CNBr fragment expected, *i.e.*, that containing valine at the amino terminus (Figure 1). The following sequence of the first 21 residues was elucidated by gas chromatography and hydrolytic back-conversion of the liberated PhNCS derivatives

Val-Val-Glu-Ile-Pro-CmCys-Trp-Thr-Asn-Ala-
5 10
Lys-Leu-Glu-Ile-Thr-Lys-Glu-Glu-Thr-Leu-Asn-
15 20

Carboxymethylcysteine at position 6 was identified as alanine following back-conversion in HI. The fact that neither alanine nor serine was observed at position 6 by chromatographic methods argues in favor of this assignment, but until it is proved by independent means it must remain tentative. Residues 15 (Thr) and 19 (Thr) were also back-converted in HI to α-aminobutyric acid. Residues 11 (Lys), 17 (Glu), and 18 (Glu) were hydrolyzed in HCl to give the expected parent amino acids. The initial yield of Val>PhNCS at position 1 was 71%, and the repetitive yield calculated from the quantities of Val-1>PhNCS and Ala-10>PhNCS was 93% (*cf.*

TABLE II: Amino Acid Composition of Yeast Inorganic Pyrophosphatase Compared to Compositions of a Number of Fragments Derived Therefrom during the Course of CNBr Cleavage.[a]

Amino Acid	RCM-PP_iase[b] Subunit Residues	CNBr II (1.1 nmol)		CNBr III (0.7 nmol)		IV (0.7 nmol)		V (1.2 nmol)		CNBr VI (2.0 nmol)		VII (4.0 nmol)	
		nmol	Residues	nmol	Residues	nmol	Residues	nmol	Residues	nmol	Residues	nmol	Residues
CmCys[c]	1	0	0	0.26	1	0.25	1	0.5	1	0	0	0	0
Asx	39	20.58	19	9.10	13	8.74	12	9.48	8	13.54	7	12.48	3
Thr	17	6.44	6	5.39	8	4.34	6	6.06	5	9.01	5	10.89	3
Ser	11	8.80	8	1.90	3	1.43	2	1.83	2	1.96	1	3.19	1
Hse[d]		0	0	0.50	1	0	0	0	0	1.43	1	2.96	1
Glx	26	9.62	9	8.53	12	6.67	9	9.46	8	8.41	4	22.83	6
Pro	19	9.58	9	4.12	6	2.93	4	3.72	3	4.16	2	11.56	3
Gly	16	6.45	6	5.09	7	3.66	5	4.84	4	4.54	2	15.77	4
Ala	21	12.81	12	4.40	6	3.22	5	3.98	3	5.92	3	12.82	3
Val	13	3.36	3	4.65	7	3.34	5	3.66	3	4.61	2	15.17	4
Met	2												
Ile	28	13.78	13	6.12	9	3.71	5	6.26	5	7.62	4	19.09	5
Leu	17	9.19	8	4.00	6	2.78	4	4.63	4	3.97	2	11.30	3
Tyr	10	4.11	4	1.98	3	1.29	2	2.36	2	7.15	4	7.20	2
Phe	10	5.99	5	2.18	3	2.22	3	3.86	3	3.57	2	1.07	0
His	5	1.05	1	2.66	4	2.67	4	3.69	3	2.67	1	0.95	0
Lys	27	15.30	14	6.81	10	5.40	8	7.47	6	9.21	5	12.40	3
Arg	6	2.05	2	2.03	3	2.12	3	3.58	3	1.62	1	1.08	0
Trp[e]	3		2		1		1		1				
Total	271		121		103		79		64		46		41

[a] Samples were hydrolyzed in 6 N HCl for 24 hr at 110° *in vacuo* and the hydrolysates were analyzed on a Durrum D-500 amino acid analyzer. Figures in parentheses refer to the number of nanomoles of peptide subjected to analysis. [b] RCM-PP_iase is an abbreviation for reduced and carboxymethylated yeast inorganic pyrophosphatase; data from Heinrikson *et al.* (1973). [c] Recoveries of the single residue of CmCys in fragments CNBr III, IV, and V were roughly 0.3 residue, in spite of precautions taken to ensure high yields. This is probably due to a combination of oxidative losses and the low per cent composition of this residue in the peptides. [d] Total of homoserine + homoserine lactone. [e] Determined separately on sample hydrolyzed in 4 N methanesulfonic acid.

Table I). All assignments in the sequence were confirmed by thin-layer chromatography.

Amino acid analysis of this fragment gave a composition of 103 amino acids (*cf.* Table II). Of particular interest were the observations in CNBr III of one residue of homoserine and the presence of the single residue of carboxymethylcysteine in the pyrophosphatase subunit. From the amino acid analysis, a yield of 3–4 μmol (~60%) was calculated. Carboxyl-terminal analysis with carboxypeptidase A of 75 nmol of succinylated CNBr III gave the expected initial release of homoserine (0.6 residue) followed by release of one full residue of isoleucine. As shown in Table III, the release of amino acids as a function of time allowed the assignment of the following tentative sequence in the vicinity of the carboxyl terminus of CNBr III

-(Ala-Leu)-Gly-Ile-Hse-COOH

In view of prior knowledge of the amino-terminal sequence in the native enzyme, these results provided evidence that of the three expected CNBr fragments, CNBr III constitutes the internal segment of the polypeptide.

IV. This fragment was shown by gel electrophoresis to contain a minor component corresponding in mobility to V (*cf.* Figure 2). Automated Edman degradation (dimethylallylamine–peptide program) of IV (100 nmol) through ten cycles gave an amino-terminal sequence identical with that of CNBr III; no additional PhNCS derivatives were observed. Amino acid analysis of IV indicated a composition of 79 amino acids including carboxymethylcysteine, but absolutely no homoserine (Table II). Digestion of 75 nmol of succinylated IV with carboxypeptidase A showed no release of homoserine or of any other amino acids. It was also found that the molar yield of IV

TABLE III: Sequential Liberation of Amino Acids from CNBr III during Digestion with Carboxypeptidase A.[a]

Amino Acid	nmol of Amino Acid Released per nmol of CNBr III after				
	5 min	30 min	60 min	120 min	240 min
Homoserine + lactone	0.22	0.33	0.51	0.68	0.79
Isoleucine	0.19	0.45	0.60	0.76	1.02
Glycine	0.08	0.30	0.53	0.77	1.05
Leucine		0.24	0.46	0.75	1.00
Alanine		0.22	0.46	0.70	0.97

[a] CNBr III (75 nmol) was digested with carboxypeptidase A and aliquots were deproteinated and analyzed as described in Methods.

was only about 15% relative to those of CNBr II and CNBr III. These experimental observations lead to the conclusion that IV is not a unique and genuine CNBr fragment, but is an amino-terminal peptide produced by limited internal cleavage of CNBr III.

V. Polyacrylamide gel electrophoresis indicated that V contained a single polypeptide component (*cf.* Figure 2). Automated Edman degradation (dimethylallylamine–peptide program) of this fraction (100 nmol) through ten cycles revealed an amino-terminal sequence which was exactly the same as determined for CNBr III and IV.

Results of amino acid analysis presented in Table II showed that V was composed of 64 amino acid residues. As had been the case for IV, this fragment contained carboxymethylcysteine, but absolutely no homoserine. No amino acids were released during digestion of 75 nmol of succinylated fraction V by carboxypeptidase A. The yield of V relative to CNBr II or CNBr III was similar to that of IV, namely, about 15%. These findings suggest that V, like IV, is an amino-terminal peptide derived by an internal cleavage of CNBr III, and that this cleavage occurs nearer the amino terminus than that which produced IV. The total molar yield of CNBr III, IV and V is approximately 90%.

CNBr VI. This fraction moved as a single component on sodium dodecyl sulfate–polyacrylamide gel electrophoresis (*cf.* Figure 2). Automated Edman degradation (dimethylallylamine–peptide program) further demonstrated the purity of this fragment, and yielded the same amino acid sequence as had been observed previously from studies on the intact enzyme (Heinrikson *et al.*, 1973), with the exception that position 5, previously left blank, was determined to be arginine. This assignment was made by tryptic digestion of succinylated CNBr VI, followed by isolation of fragments by gel filtration and amino acid analysis. One of the fragments was ninhydrin-negative and Sakaguchi-positive and gave the following composition: Thr_3, Tyr, Arg. These amino acids correspond to the first five in the pyrophosphatase subunit.

Amino acid compositional data presented in Table II indicate that CNBr VI contains 46 residues, including 1 residue of homoserine. The analysis further showed that the total yield of the fragment was 4–5 μmol, in excellent agreement with the yields of CNBr II and CNBr III.

Carboxyl-terminal sequence analysis of CNBr VI was performed by hydrolysis with carboxypeptidase A. Digestion of 100 nmol of peptide gave the expected immediate release of homoserine (0.6 residue), followed by release of a full residue of phenylalanine. The data shown in Table IV allowed the assignment of the following tentative carboxyl-terminal sequence of CNBr VI: -Ile-Phe-Hse-COOH. These experimental observations show that CNBr VI is a unique and genuine CNBr fragment, and that it constitutes the amino-terminal portion of the intact enzyme subunit.

VII. High voltage paper electrophoresis at pH 2.1 was inconclusive regarding the purity of VII, but automated Edman degradation (dimethylallylamine–peptide program) revealed a clearly defined sequence through 20 residues. This partial structure, obtained by the methods described earlier, is

Pro-Ile-Asp-Val-Leu-Gln-Ile-Gly-Glu-Thr-
5 10
Ile-Ala-Tyr-Thr-Gly-Gln-Val-Lys-Gln-Val
15 20

Residues 10 (Thr) and 14 (Thr) were confirmed by back-conversion in HI, and residues 6 (Glu), 16 (Glu), 18 (Lys), and 19 (Glu) were hydrolyzed in HCl to give the expected amino acids. The initial yield of Pro>PhNCS was 63%, and the repetitive yield (*cf.* Table I) based upon Ile-2>PhNCS and Ile-11 >PhNCS was 94%. Amino acid analysis of VII gave a composition of 41 amino acids including one residue of homoserine (Table II). This indicated a total yield of 800 nmol of the fragment, a value very similar to the yields obtained for two of the other minor components, IV and V.

Digestion of 75 nmol of VII with carboxypeptidase A gave the same results as obtained with CNBr III (Table III). Initially, 0.7 residue of homoserine was released, followed by a full residue of isoleucine and the three additional amino acids to give the same tentative sequence of -(Ala-Leu)-Gly-Ile-Hse-COOH.

The results indicate that VII is not a unique CNBr cleavage product, but rather the carboxyl-terminal portion resulting from an internal cleavage of CNBr III. Based upon compositional analyses (Table II) and partial sequence data at the amino and carboxyl termini, it appears very likely that fragments V and VII are produced from CNBr III by cleavage of a single peptide bond between an as yet unidentified residue and proline.

VIII, IX, and X. High-voltage paper electrophoresis of these fractions revealed that each one contained a mixture of small peptides none of which exceeded about 5% in yield. Automated Edman degradation (dimethylallylamine–peptide program) of these fractions through five cycles gave a number of PhNCS derivatives at each step, and it was possible to reconstruct several sequences previously determined with purified fractions. One of these peptide mixtures could contain the carboxyl-terminal peptide produced together with IV by cleavage of CNBr III, but this question has not been pursued further.

Extensive Edman Degradation of Intact Pyrophosphatase. The relatively small size of amino-terminal fragment CNBr VI presented the possibility of performing automated Edman degradation through the first methionine residue and thus establishing an overlap which would provide independent evidence for the proposed ordering of CNBr fragments. Intact, reduced, and S-carboxymethylated pyrophosphatase (600 nmol of subunit) was subjected to 50 stages of the degradative procedure with the Quadrol program. This relatively large amount of protein was used to enhance the chances of elucidating the sequence through the necessary number of steps to establish the overlap. In this way, the sequence of 20 residues originally reported (Heinrikson *et al.*, 1973) was extended by

TABLE IV: Sequential Liberation of Amino Acids from CNBr VI during Digestion with Carboxypeptidase A.[a]

	nmol of Amino Acid Released per nmol of CNBr VI after				
Amino Acid	5 min	30 min	60 min	120 min	240 min
Homoserine + lactone	0.40	0.64	0.63	0.65	0.63
Phenylalanine	0.38	0.80	0.95	0.98	0.98
Isoleucine	0.30	0.73	0.89	0.95	0.97

[a] CNBr VI (100 nmol) was digested with carboxypeptidase A, and aliquots were deproteinated and analyzed as described in Methods.

14 residues to give the following amino-terminal sequence through 34 cycles

H_2N-Thr-Tyr-Thr-Thr-Arg-Gln-Ile-Gly-Ala-Lys-Asn-Thr-Leu-
5 10
Glu-Tyr-Lys-Val-Tyr-Ile-Glu-Lys-Asp-Gly-Lys-
15 20
Pro-Val-Asn-Ala-Phe-Asn-His-Thr-Ile-Pro-
25 30

The arginine at position 5 was definitely established both by the phenanthrenequinone spot test (Yamada and Itano, 1966) and by hydrolytic back-conversion in HCl. Histidine at position 31 was also identified by the latter procedure. The yield of Tyr>PhNCS at position 2 was 71%, and the repetitive yield calculated from Tyr-2>PhNCS to Tyr-15>PhNCS was 93% (*cf.* Table I). The identities of several of the PhNCS derivatives after position 34 were unclear, but residues 46–48 were definitely established to be Met-Val-Val. This shows that CNBr VI should contain 46 residues, a value in accord with the compositional analysis (Table II). Moreover, these findings provide definitive evidence in support of the alignment of the CNBr fragments based upon experiments described above.

Figure 3 summarizes our current knowledge with regard to the covalent structure of the yeast inorganic pyrophosphatase subunit. Of the total of approximately 270 amino acids in the polypeptide, 108 have been positioned largely with the aid of automated Edman degradative procedures. Further analysis of peptides from these large fragments should provide confirmation of the sequences obtained in this way. It should be noted that the numbering of residues following Asn-67 is based upon compositional data only and must be considered tentative until the entire structure is completed.

Discussion

An earlier report from this laboratory (Heinrikson *et al.*, 1973) presented evidence, based upon molecular weight studies, peptide mapping, and limited sequence analysis at the amino and carboxyl termini, that yeast inorganic pyrophosphatase is a dimer composed of identical subunits approximately 30,000 in molecular weight. While this communication was in press, a description of a similar investigation was published by Hansen *et al.* (1972) which indicated the presence of identical subunits of mol wt ~35,000, essentially in accord with our observations. Our identification of threonine and valine as the amino- and carboxyl-terminal residues, respectively, has also been reported by Eifler *et al.* (1972). Moreover, our findings, in agreement with those published by a number of investigators (Negi and Irie, 1971; Eifler *et al.*, 1966), indicated the presence of 2 mol of methionine per 30,000 g of enzyme, or two residues per subunit. The low content of methionine suggested that an initial approach to the eventual complete covalent structural analysis of the enzyme might be through fragmentation with CNBr. The larger fragments which one would expect as a result of this procedure would be few in number (three), and easily aligned in the complete polypeptide chain, especially in view of the sequence information already in hand. Furthermore, such large peptides would be amenable to current methods of automated Edman degradation and should yield considerable additional structural data. Earlier work had provided evidence with regard to terminal sequences in the subunit (Heinrikson *et al.*, 1973), and experiments reported here clearly show that CNBr cleavage produces two new amino-terminal residues identified as valine and alanine (*cf.* Figure 1). Of the total fragments produced by CNBr cleavage of pyrophosphatase under conditions described in this paper, three major fragments designated CNBr II, CNBr III, and CNBr VI (Figure 2) were isolated in high yield and in a high state of purity. These peptides, taken together, constitute the whole of the polypeptide chain of the enzyme. Alignment of these fragments was simple and straightforward. It was shown that CNBr VI is the amino-terminal fragment in the pyrophosphatase monomer. The sequence of amino acids at the amino terminus of this 46-residue peptide is identical with that of the native enzyme, and the fragment terminates in the sequence Ile-Phe-Hse (Table IV, Figure 3). CNBr III is the internal CNBr fragment from the subunit. This was shown by the fact that it terminates in a sequence containing Hse (. . . Gly-Ile-Hse; Table III, Figure 3), and also by the observation of a new amino-terminal sequence beginning with valine (Figures 1 and 3). Automated Edman degradation of intact enzyme through 50 cycles proved that CNBr III arises by virtue of a cleavage at Met-46 and that this fragment is the internal CNBr peptide. The carboxyl-terminal CNBr fragment was proved to be CNBr II, the largest of the three peptides (Table II). CNBr II terminates in a sequence identical with that of native pyrophosphatase, and its amino-terminal residue is alanine; no homoserine was observed in this fragment. CNBr VI is a 46-residue fragment the amino-terminal sequence of which is identical with native pyrophosphatase. As may be seen in Table II, the total amino acids observed in CNBr II, CNBr III, and CNBr VI are in close agreement with the composition of the enzyme reported earlier (Heinrikson *et al.*, 1973).

For several reasons, the choice of CNBr as a first means of protein cleavage turned out to be a fortunate one. Cleavage by the reagent was nearly quantitative, and the major fragments thus produced were isolated in pure form and in high yield simply by gel filtration (Figure 2). Furthermore, by virtue of secondary cleavages, a number of very useful minor fragments were produced; these were also isolated in a pure state along with the bona fide CNBr fragments. Among the minor fragments isolated during the course of cleavage by CNBr, IV, V, and VII proved to be most important. These pure peptides, isolated by gel filtration in yields of about 15%, were shown by means of partial sequence analysis to be the products of internal cleavages in CNBr III. Fractions IV and V represent two different amino-terminal fragments of CNBr III. Although the data do not entirely preclude the excision of a small peptide not recovered in our studies, it seems highly likely that VII is the carboxyl-terminal peptide generated along with V by a single cleavage of CNBr III. Compositional and partial sequence analyses of IV, V, and VII enabled us to position these fragments in the pyrophosphatase subunit and to gain additional structural information which could not have been obtained through analysis of the CNBr fragments alone. Thus far, we have not found the carboxyl-terminal peptide produced simultaneously with IV by a second internal cleavage of CNBr III. The difference in composition between IV and CNBr III indicates that this peptide should contain 20–25 amino acids, and that these should comprise the final 20–25 residues of VII. Although internal sequences observed in VII were also observed during the Edman degradation of VIII and X, the impurity of the latter fractions precludes a definitive statement regarding the possible presence of the missing peptide in these preparations. Fractions VIII, IX, and X were composed of mixtures of peptides none of which exceeded 5% in yield, and which are thus of minor importance to the study at hand.

At present, we do not know the exact mechanism by which these minor cleavages occurred, but the fact that no low-level

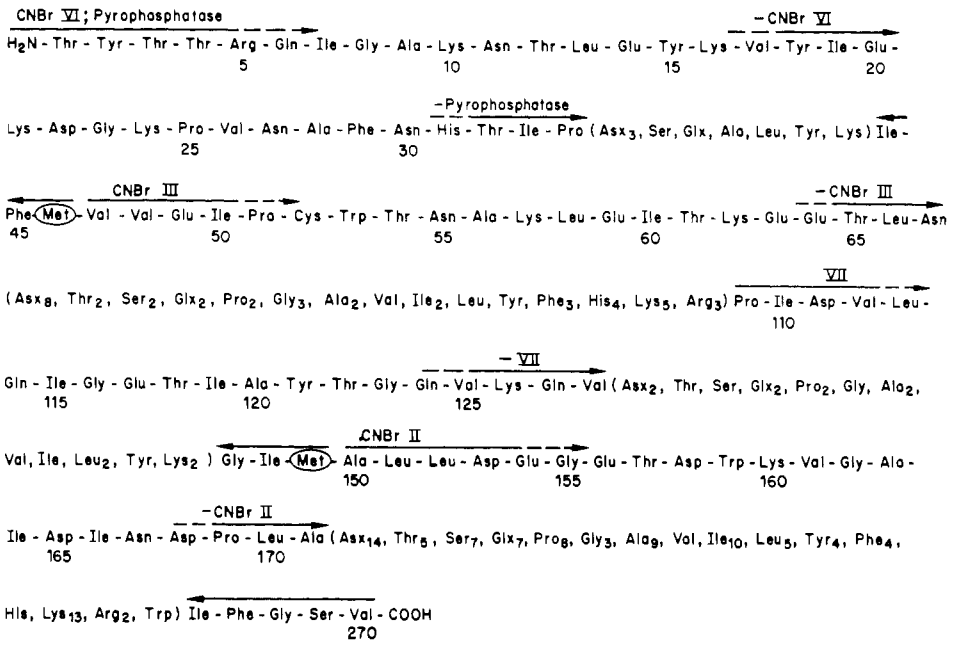

FIGURE 3: Partial amino acid sequence of yeast inorganic pyrophosphatase. Sequences elucidated by automated Edman degradation are defined in each case by two arrows pointing to the right, as shown in the following example: CNBr VI ---→, beginning of sequence analysis of the fraction designated CNBr VI in Figure 2; -CNBr VI ——→, termination of sequence analysis of CNBr VI. Arrows pointing to the left indicate sequences derived by digestion with carboxypeptidase A. Methionine residues are encircled for ease of identification. All residue numbers following Asn-67 are based upon compositional analysis of appropriate fractions, and are therefore tentative until the complete structure is elucidated.

contaminating sequences were observed in our degradative work with intact enzyme (Table I) argues strongly against the existence of nicks which might have been present in the polypeptide backbone prior to cleavage with CNBr. Eifler *et al.* (1972) have discussed the likelihood of proteolysis during the autolysis step of the preparative procedure for yeast pyrophosphatase, but our enzyme, prepared by the method of Cooperman *et al.* (1973), does not appear to have undergone internal cleavages of this kind. All of our sequence data obtained by automated Edman degradation of the intact enzyme point to a single polypeptide chain (*cf.* Table I and Heinrikson *et al.*, 1973), and the carboxyl-terminal sequence derived from intact enzyme is identical with that of CNBr II. It may be that minor cleavages in the polypeptide backbone at sites other than methionine are occurring during treatment with CNBr due to the acidic conditions employed in the cleavage. Peptide bonds involving aspartate residues are known to be acid sensitive (Schultz, 1967). Of particular relevance to the present study, Piszkiewicz *et al.* (1970) demonstrated a minor cleavage of an Asp-Pro bond in glutamate dehydrogenase during treatment with CNBr. Our results indicate that CNBr III undergoes cleavage to yield fractions V and VII, and that VII begins with Pro (Table I, Figure 3). If V terminates with Asp, it would be expected to be resistant to digestion with carboxypeptidase A and this is in fact what we have observed in our attempts to determine carboxyl-terminal sequences in both IV and V. These findings support the contention that the minor fragments arise through acid cleavage of Asx-X linkages.

The advent of automated procedures for Edman degradation of proteins and peptides has not only greatly facilitated sequence analysis, but has also opened up new strategies for attempting structural studies. Whereas the classical manual sequence methods have been limited to the analysis of small peptides, automated degradation can be carried out on much larger fragments and even on intact proteins. Several investigators have had great success with these procedures and have been able to elucidate large, if not entire, segments of the primary structure of particular proteins without recourse to the generation of small fragments. Brandt and Von Holt (1972) have reported the complete amino acid sequence of the 136 residues in histone F_3 from chicken erythrocytes as determined solely by automated procedures and limited specific cleavage. The primary structure of thermolysin obtained by conventional methods has been confirmed in large measure by the application of automated methods to large fragment analysis (Titani *et al.*, 1972). A noteworthy publication relating to the strategy of contemporary sequence analysis and the application of automated methods is that by Hermodson *et al.*, 1972.

The present investigation has relied heavily on automated Edman degradation procedures, and the results obtained therefrom have provided a rationale for alignment of the three large CNBr fragments generated from the pyrophosphatase monomer. In addition, the sequence information obtained has enabled the placement of more than one-third of the total amino acids in the polypeptide. These data, summarized in Figure 3, further substantiate the earlier contention (Heinrikson, *et al.*, 1973) that native yeast pyrophosphatase is composed of two identical subunits with an approximate mol wt of 30,000. The monomers contain methionine residues at position 46 and, tentatively, 149 in the sequence. Evidence presently in hand suggests that the single residue of cysteine is located at position 52. These findings should be of considerable value to attempts currently in progress, both for the pinpointing of residues essential for catalysis (B. S. Cooperman, N. Y. Chiu, and M. B. O'Connor; P. Heitmann[2]) and for determining the three-dimensional structure of yeast inorganic pyrophosphatase (Voet[2]).

[2] Work in progress.

Acknowledgments

The expert technical assistance of Mrs. Elaine Krueger and Mrs. Pamela Keim is gratefully acknowledged.

References

Ambler, R. P. (1967), *Methods Enzymol. 11*, 155.
Brandt, W. F., and Von Holt, C. (1972), *FEBS (Fed. Eur. Biochem. Soc.) Lett. 23*, 357.
Cooperman, B. S., Chiu, N. Y., Bruckmann, R. H., Bunick, G. J., and McKenna, G. P. (1973), *Biochemistry 12*, 1665.
Crestfield, A. M., Moore, S., and Stein, W. H. (1963), *J. Biol. Chem. 238*, 622.
Dreyer, W. J., and Bynum, E. (1967), *Methods Enzymol. 11*, 32.
Eifler, R., Hahn, V., and Hermann, I. (1972), *Acta Biol. Med. Ger. 28*, 707.
Eifler, R., Laesmer, R. G., and Rapoport, S. (1966), *Acta Biol. Med. Ger. 17*, 716.
Hansen, G., Eifler, R., and Heitmann, P. (1972), *Acta Biol. Med. Ger. 28*, 977.
Heinrikson, R. L., Sterner, R., Noyes, C., Cooperman, B. S., and Bruckmann, R. H. (1973), *J. Biol. Chem. 248*, 2521.
Hermodson, M. A., Ericsson, L. H., Titani, K., Neurath, H., and Walsh, K. A. (1972), *Biochemistry 11*, 4493.
Irreverre, F. (1965), *Biochim. Biophys. Acta 111*, 551.
Jeppsson, J.-O., and Sjöquist, J. (1967), *Anal. Biochem. 18*, 264.
Kingdon, H. S., Noyes, C., Lahiri, A., and Heinrikson, R. L. (1972), *J. Biol. Chem. 247*, 7923.
Kunitz, M. (1952), *J. Gen. Physiol. 35*, 423.
Liu, T.-Y., and Chang, Y. H. (1971), *J. Biol. Chem. 246*, 2842.
Moore, S., and Stein, W. H. (1963), *Methods Enzymol. 6*, 819.
Negi, T., and Irie, M. (1971), *J. Biochem. 70*, 165.
Pisano, J. J., and Bronzert, T. J. (1969), *J. Biol. Chem. 244*, 5597.
Piszkiewicz, D., Landon, M., and Smith, E. L. (1970), *Biochem. Biophys. Res. Commun. 40*, 1173.
Schultz, J. (1967), *Methods Enzymol. 11*, 255.
Smithies, O., Gibson, D., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G., and Ballantyne, D. L. (1971), *Biochemistry 10*, 4912.
Spackman, D. H., Stein, W. H., and Moore, S. (1958), *Anal. Chem. 30*, 1190.
Titani, K., Hermodson, M. A., Ericsson, L. H., Walsh, K. A., and Neurath, H. (1972), *Nature (London), New Biol. 238*, 35.
Udenfriend, S., Stein, S., Bohlen, P., and Dairman, W. (1972), *in* The Chemistry and Biology of Peptides, Meienhofer, J., Ed., Ann Arbor, Mich., Ann Arbor Science Publishers, pp 655–663.
Van Orden, H., and Carpenter, F. (1964), *Biochem. Biophys. Res. Commun. 14*, 399.
Weber, K., and Osborn, M. (1969), *J. Biol. Chem. 244*, 4406.
Yamada, S., and Itano, H. A. (1966), *Biochim. Biophys. Acta 130*, 538.

Multiple Forms of Galactosyltransferase from Bovine Milk†

Steve C. Magee, Ramswarup Mawal,‡ and Kurt E. Ebner*,§

ABSTRACT: A galactosyltransferase which transfers galactose from UDP-galactose to form β-(1–4)-glycosides was isolated from bovine milk. Two molecular forms, 42,000–44,000 and 55,000–59,000 molecular weight, were demonstrated by electrophoresis on polyacrylamide gels and chromatography on Bio-Gel P-200. Both forms contained carbohydrate. Both forms had similar catalytic properties based on similar apparent K_m's for UDP-galactose, *N*-acetylglucosamine, and α-lactalbumin. Both forms were retained to the same extent on α-lactalbumin–Sepharose columns and were inhibited to an equal extent by heating at 54° and by sulfhydryl inhibitors.

Galactosyltransferase (UDP-galactose:D-glucose 1-galactosyltransferase, EC 2.4.1.22) catalyzes the transfer of galactose from UDP-galactose forming β-(1–4) linkages with free *N*-acetylglucosamine (Brew *et al.*, 1968) or protein-bound β-glycosides which terminate with *N*-acetylglucosamine (Schanbacher and Ebner, 1970). Galactosyltransferase alone has a low affinity for glucose (K_m > 1 M) as the galactosyl acceptor but α-lactalbumin (*ca.* 0.1 mg/ml) lowers the K_m to the millimolar region (Fitzgerald *et al.*, 1970a) though V_{max} is affected only moderately (Morrison and Ebner, 1971). The general reaction is described by Scheme I, where R_1 is –OH or an oligosaccharide and R_2 is –OH or –$NHCOCH_3$.

Galactosyltransferase occurs in most tissues bound to the Golgi apparatus (Coffey and Reithel, 1968a,b) but is found in soluble form in milk and blood (Wagner and Cynkin, 1971; Ebner and McKenzie, 1972; Bella and Kim, 1972). Molecular weight estimates range from 29,000 to 130,000 (Palmiter, 1969) for the soluble enzymes though the bovine milk enzyme molecular weight is 42,000 and contains approximately 12% carbohydrate, including 2% sialic acid (Trayer and Hill, 1971). Evidence obtained in several laboratories suggests that there may be more than one form of the enzyme in milk (Magee *et al.*, 1972; Trayer and Olsen, 1972; Klee and Klee,

† From the Department of Biochemistry, Oklahoma State University, Stillwater, Oklahoma 74074. *Received August 27, 1973.* Journal Article J-2748 of the Agricultural Experiment Station, Oklahoma State University. Supported by National Institutes of Health Grant AM-10764 and National Science Foundation Grant 24291. A preliminary report has appeared (Magee *et al.*, 1972).

‡ Present address: Ahmednagar College, Ahmednagar, India.

§ Recipient of a Career Development award (GM 42936) from the National Institutes of Health.