THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 244, No. 8, Issue of April 25, pp. 2132-2136, 1969
Printed in U.S.A.

The Enzymes of the Galactose Operon in *Escherichia coli*

II. THE SUBUNITS OF URIDINE DIPHOSPHOGALACTOSE 4-EPIMERASE*

(Received for publication, September 11, 1968)

DAVID B. WILSON‡ AND DAVID S. HOGNESS

From the Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305

SUMMARY

Uridine diphosphogalactose 4-epimerase of *Escherichia coli* had previously been shown to have a molecular weight of 7.9 × 10^4 and to contain 1 molecule of tightly bound DPN (WILSON, D. B., AND HOGNESS, D. S., *J. Biol. Chem.*, *239*, 2469 (1964)). This enzyme is further characterized here with regard to (*a*) amino acid composition, (*b*) number of tryptic peptides, (*c*) amino-terminal residues, and (*d*) molecular weight in guanidine hydrochloride. The number of tryptic peptides indicates that an amino acid sequence of 360 residues is represented twice within the native enzyme. The finding of only one pair of amino-terminal residues (aspartic acid) per enzyme molecule and the reduction of the molecular weight by a factor of 2 in guanidine hydrochloride are consistent with the simplest conclusion that the sequence of 360 residues is contained within a single polypeptide. Uridine diphosphogalactose 4-epimerase therefore appears to be a complex consisting of two identical polypeptides and one molecule of DPN held together by noncovalent bonds. Some implications regarding the size of the structural gene for this enzyme and the nature of the binding sites for DPN are discussed.

The first three reactions of galactose metabolism are catalyzed by galactokinase, galactose-1-P uridyl transferase, and uridine diphosphogalactose 4-epimerase (1). In *Escherichia coli*, the syntheses of these enzymes are coordinately induced (2), and their structural genes lie adjacent to one another to form the genetic configuration known as the galactose operon (3–5).

In the first article of this series on the enzymes of the galactose operon (6), we described a procedure for the purification of UDP-galactose 4-epimerase, and showed that the active enzyme has a molecular weight of 7.9 (±0.8) × 10^4 and contains one mole of tightly bound DPN per mole of enzyme. In this and the succeeding article (7) we seek a definition of the structures of UDP-galactose 4-epimerase and of galactokinase sufficient to determine the sizes of their structural genes, thereby allowing the study on the frequency of gene translation presented in the accompanying, fourth article of this series (8). Here, we have extended the characterization of UDP-galactose 4-epimerase to include determinations of (*a*) the amino acid composition, (*b*) the number of tryptic peptides, (*c*) the amino-terminal residues, and (*d*) the molecular weight in guanidine hydrochloride.

The results of these determinations consistently indicate that the enzyme is a complex consisting of two equivalent polypeptide subunits and the single DPN molecule, held together by noncovalent bonds. As each of the two subunits contains 360 amino acid residues which are not divisible into multiple, identical sequences of significant size (*i.e.* large enough to contain multiple tryptic peptides), the size of the structural gene for this enzyme is shown to be 3 times 360, or 1080 base pairs.

EXPERIMENTAL PROCEDURE

Materials

UDP-galactose 4-epimerase was purified through Step VII of our procedure (6). It was then rechromatographed on DEAE-Sephadex and concentrated as described in the preparation of Fraction IX (6) to yield the product used here. The specific activity (14 × 10^3 units per mg) and the homogeneity, as tested by electrophoresis in polyacrylamide gels, were equal to those of the most highly purified preparations previously reported (6). Trypsin, obtained from Worthington, had been crystallized five times. Urea was deionized by the procedure of Benesch, Lardy, and Benesch (9) prior to use; all other reagents were commercial preparations.

Methods

Amino Acid Analyses—Except as noted below, the amino acid determinations (Table I, below) were performed according to the procedures of Hirs, Moore, and Stein (10). Protein samples of 0.9 mg were hydrolyzed in 6 N HCl at 110° for 18, 49, or 72 hours. Analyses were performed with the Beckman-Spinco amino acid analyzer.

Cysteine and cystine were determined as cysteic acid after performic acid oxidation (11).

Two methods were used to assay the free sulfhydryl groups in UDP-galactose 4-epimerase. In Method A (Table I), cysteine

* This work was supported by grants from the National Institutes of Health.

‡ Present address, Department of Biochemistry, Cornell University, Ithaca, New York 14850.

residues were determined as *S*-(1,2-dicarboxyethyl)-L-cysteine after modification of the enzyme by reaction with *N*-ethylmaleimide and subsequent hydrolysis (72 hours). The procedure of Smyth, Blumenfeld, and Konigsberg (12) was used, except that the protein was not reduced prior to reaction with *N*-ethylmaleimide and this reaction was carried out in 3.5 M guanidine hydrochloride, pH 7.0. In Method B (Table I), free sulfhydryl groups were determined by reaction with Ellman's reagent (5,5′-dithiobis(2-nitrobenzoic acid)) according to his procedure (13), modified by the addition of 3.5 M guanidine hydrochloride 22 min after initiation of the reaction. In the absence of guanidine hydrochloride the number of —SH groups which have reacted with the reagent increases rapidly to 1.7 per enzyme molecule (mol wt 7.9×10^4) by 9 min, rising slowly thereafter to 2.1 by 22 min. Addition of guanidine hydrochloride at this time caused a rapid further increase to the value of 3.7 residues per enzyme molecule given in Table I, no further increase being observed in the succeeding 30 min.

Tryptophan was determined by the spectrophotometric method of Goodwin and Morton (14).

Reduction and Alkylation Prior to Tryptic Hydrolysis—UDP-galactose 4-epimerase (5 mg) was dissolved in 2.0 ml of 0.1 M Tris-HCl buffer, pH 8.8, 5.0 M guanidine hydrochloride, and 1.0 M mercaptoethanol. The solution was flushed with N_2, the sealed vessel was incubated at 37° for 6 hours and then cooled to 25°, and the solution was brought to pH 7.0 by the addition of 0.12 ml of 0.1 M HCl. *N*-Ethylmaleimide (0.35 g) was added and the solution was kept at 25° for 15 min, after which the reaction was stopped by the addition of 0.5 ml of glacial acetic acid. The solution was then dialyzed overnight against water at 4°. The protein precipitated during the dialysis and was recovered by centrifugation.

TABLE I

Amino acid composition of UDP-galactose 4-epimerase

Amino acid	Residues per molecule of enzyme		Residues per subunit (closest integer)
	No. of residues[a]	Standard deviation	
Aspartic acid	91.1	0.8	46
Threonine	33.0[b]		16
Serine	39.9[b]		20
Glutamic acid	64.6		32
Proline	45.3	0.6	23
Glycine	65.7	0.4	33
Alanine	57.6	0.4	29
Valine	62.3[c]	0.4	31
Methionine	17.9[d]	0.6	9
Isoleucine	38.7[c]	0.2	19
Leucine	60.6	0.6	30
Tyrosine	27.7[c]	0.2	14
Phenylalanine	20.5[c]	0.2	10
Lysine	30.6[e]		15
Histidine	19.7[e]		10
Arginine	25.6[e]		13
Tryptophan	12.6[f]		6
Half-cystine	8.1[f]		4
Total			360
—SH groups			2
Method A	4.6[f]		
Method B	3.8[f]		

[a] The determinations are described under "Methods." Except where indicated, the average of five analyses is given.

[b] Obtained by extrapolation to zero time of values from 18-, 49-, and 72-hour hydrolysates.

[c] Average of four analyses.

[d] As methionine determined before and after performic oxidation was not significantly different, all values are included in this average.

[e] Average of two analyses.

[f] Single determinations.

Tryptic Hydrolysis and Separation of Peptides—Both the reduced, alkylated UDP-galactose 4-epimerase and the unmodified enzyme, lyophilized after dialysis against water, were hydrolyzed by trypsin; the resulting peptides separated on filter paper sheets essentially as described by Helinski and Yanofsky (15). Several milligrams of protein (*e.g.* 5 mg) were dissolved in 0.8 ml of a freshly prepared solution containing 6.0 M deionized urea and 0.3 M ammonium carbonate. A freshly prepared solution of trypsin (60 μl of 2.5 mg of trypsin per ml) was added, followed immediately by the addition of 1.2 ml of water. Hydrolysis was allowed to proceed at 25° for 2 hours; it was then stopped by rapid freezing. The peptides were freed of salt and separated on filter paper as described by Helinski and Yanofsky (15). Increasing the time of hydrolysis did not significantly alter the pattern of peptides on the filter paper.

Analysis of Amino-terminal Residues—Amino-terminal residues were determined by the method of Stark and Smyth (16) with the following specifications. Carbamylation of 5.5 mg of UDP-galactose 4-epimerase was carried out for 6 hours at 50° in 4.0 M guanidine hydrochloride-50% *N*-ethylmorpholine-acetate buffer, pH 8.0. In calculating the moles of amino-terminal residues per mole of enzyme, the amount of enzyme analyzed was calculated from an amino acid analysis of a 48-hour hydrolysate of Fraction C in the procedure (16). The yields of residues given in Table II (below) were corrected by subtraction of blank values obtained by subjecting an uncarbamylated sample of protein to the procedure; in no case did the blank values exceed 0.01 residue per enzyme molecule of molecular weight 7.9×10^4.

RESULTS

Amino Acid Composition—The frequencies of amino acid residues in UDP-galactose 4-epimerase are given in Table I as the number of residues of each amino acid per molecule of active enzyme (mol wt 7.9×10^4). The number of residues per subunit is given in the last column. These values depend upon the assumption that the enzyme consists of two equivalent subunits, an assumption based upon the results presented below.

Tryptic Peptides—The sum of the lysine and arginine residues in UDP-galactose 4-epimerase is 56 (Table I). The maximum number of different peptide bonds per enzyme molecule which should be sensitive to trypsin hydrolysis is therefore 56. If this molecule contains N identical polypeptides, then the maximum number of different tryptic peptides is $(56/N) + 1$, and for $N = 2$ this number is 29.

Hydrolysis of the protein by trypsin and separation of the resulting peptides on filter paper sheets (peptide maps) were performed as described in "Methods." Four such two dimensional peptide maps were made; two were obtained from digests of UDP-galactose 4-epimerase which had been reduced and alkylated with *N*-ethylmaleimide, and two from digests of unmodified enzyme, 1 to 3 mg of peptides being used in each case.

TABLE II

Amino-terminal residues in UDP-galactose 4-epimerase

The determinations are described under "Methods." The values have not been corrected for recoveries observed by Stark and Smyth (16) when amino acids were carried through the procedure (see the text). The reported recoveries were aspartic acid, 99%; glycine, 98%; lysine, 84%; glutamic acid, 60%; and serine, 20%.

Amino acid	Residues per molecule of molecular weight	
	79,000	39,500
Aspartic acid	2.32	1.16
Glutamic acid	0.30	0.15
Glycine	0.21	0.10
Lysine	0.17	0.08
Serine	0.13	0.06

One member of each pair was stained with *tert*-butyl hypochlorite reagent (17, 18). Ninhydrin was used to stain the other member of the pair derived from the reduced, alkylated protein, whereas the other member of the pair derived from the unmodified protein was stained with an arginine-specific reagent (15).

The number of distinct peptide spots seen on the three maps stained either with *tert*-butyl hypochlorite or ninhydrin ranged from 27 to 29. The peptide map derived from the unmodified enzyme exhibited a streak near the origin which was absent from both peptide maps of the reduced, alkylated protein. Evidently reduction and alkylation prevent formation of an undigested core. These results are compatible with $N = 2$ and are incompatible with N equal to any other integer. Consistent with this conclusion is the fact that the arginine-specific reagent stained only 13 of the tryptic peptides, whereas there are 26 arginine residues per epimerase molecule.

Amino-terminal Residues—The results of an analysis for amino-terminal residues in UDP-galactose 4-epimerase according to the carbamylation method of Stark and Smyth (16; see "Methods") are given in Table II. The value found for aspartic acid is consistent with the model of two identical subunits. The method does not distinguish between aspartic acid and asparagine, and consequently the two subunits could differ in this regard.

The values in Table II have not been corrected for the yields reported by Stark and Smyth (16) when individual amino acids were carried through the entire procedure. Such a correction would be insignificant for aspartic acid (99%) yield, and is of unknown applicability in the case of the four amino acids detected in minor amounts because of their unknown origin. However, when such corrections are applied to the minor amino acids (see the legend to Table II), in no case does the resulting value approximate the 1 residue per subunit defined by the tryptic peptides. The largest value is 0.3 serine residue per subunit, and in this case the correction factor is very large, owing to the small (20%) yield. Hence, the detection of these minor residues neither indicates nor can be explained by structures containing more than one polypeptide per subunit.

Some of the minor residues may arise from proteins contaminating the enzyme preparation. However, the data on the homogeneity of these preparations indicates less than 10% contamination, whether derived from chromatography on DEAE-Sephadex, sedimentation, or starch and polyacrylamide gel electrophoresis (6; "Materials"); further, when such contamination is indicated, it arises from material moving somewhat more slowly during gel electrophoresis or sedimenting slightly faster than the majority component. Hence, it is difficult to ascribe all four minor residues to this source. There remains the possibility that one or more of these residues results from infrequent cleavage of the polypeptide chain during the carbamylation. We have, however, no evidence regarding the probability of such events in UDP-galactose 4-epimerase.

Molecular Weight in Guanidine Hydrochloride—The molecular weight of UDP-galactose 4-epimerase in solutions of guanidine hydrochloride was determined by sedimentation to equilibrium using the method described for Fig. 6*b* of the first article of this series (6), in which $(1/r)$ (dc/dr) is plotted against the protein concentration, c, at equilibrium. Specifically, 2 mg of protein per ml of 0.01 M potassium phosphate buffer (pH 7.0) containing guanidine hydrochloride at 4.65 M or 6.14 M was centrifuged at 25,980 rpm for 21 hours as a 3-mm liquid column in a 12-mm, 2.5°, double sector cell at 22°. The coordinates of points on the above equilibrium plot were obtained from the schlieren patterns as previously described (6). No significant curvature was observed, and the points conform to the straight line determined by the method of least squares to the same extent as exhibited in Fig. 6*b* of Reference 4 for the case of the native enzyme in the absence of guanidine hydrochloride.

Taking the slope of this line as $M\ (1 - \bar{v}\rho)\ (w^2/RT)$, the value for $M\ (1 - \bar{v}\rho)$ was calculated to be 7.70×10^3 in 4.65 M guanidine hydrochloride and 6.64×10^3 when the concentration was raised to 6.14 M. If one assumes no preferential interactions in such a multicomponent system, the partial specific volume ($\bar{v}$) of UDP-galactose 4-epimerase can be taken as 0.73 ml per g. This is the value which was previously assumed in the calculation of the molecular weight of the active enzyme (6) and which can now be calculated from the amino acid composition given in Table I (19). With the use of this value and the densities (ρ) of the respective solutions, the calculated molecular weights (M) are 4.1×10^4 and 3.9×10^4 in 4.65 and 6.14 M guanidine hydrochloride, respectively.[1] These values are very close to one-half the value of

[1] The method of Schachman and Edelstein (pp. 2699–2701 of Reference 20) allows for preferential interaction between either water or guanidine hydrochloride and polypeptides, and can be applied to our data. It includes the peculiar assumption that the interaction is invariant with respect to change in concentration of guanidine hydrochloride at these high concentrations and hence that the molecular weight of the complex (M_c) and its partial specific volume ($\bar{v}_c$) are constant. One then obtains $\bar{v}_c = (R - 1)/(R\rho_2 - \rho_1) = 0.762$ ml per g, where $R = [M_c(1 - \bar{v}_c\rho_1)/M_c(1 - \bar{v}_c\rho_2)]$, ρ_1 and ρ_2 are the solution densities for the two concentrations of guanidine hydrochloride used, and the values for the $M_c(1 - \bar{v}_c\rho)$ in R are those given in the text. This allows the computation of M_c, and, on the assumption that $\bar{v}$ for the free polypeptide is 0.73 ml per g, the preferential interaction, χ, which we find is 0.13 g of water per g of polypeptide. The further assumption that $M_c = M(1 + \chi)$ yields $M = 4.5 \times 10^4$, the molecular weight of the polypeptide. Although this value is also in reasonable agreement with the model of two equivalent subunits (the ratio of molecular weights in the absence and in the presence of guanidine hydrochloride becomes 1.8), we have not emphasized this treatment because of the numerous assumptions for which we have no substantiation, or good intuition, and because our data are limited to only two conditions of concentration.

7.9×10^4 obtained in the absence of guanidine hydrochloride (6), and hence are consistent with the model of two equivalent subunits. Furthermore, the linearity of the plots indicates a homogeneity of the polypeptide population demanded by equivalence of the subunits. We have not, however, examined the equilibrium behavior in higher centrifugal fields, which would be useful for the detection of considerably smaller polypeptide components.

DISCUSSION

The results of three sets of experiments consistently indicate that the UDP-galactose 4-epimerase of *E. coli* is divisible into two equivalent subunits. The first set, in which the number of tryptic peptides is compared to the number of arginine and lysine residues, offers the strongest, self sufficient evidence for this conclusion. This comparison indicates that an amino acid sequence of 360 residues is represented twice within the native enzyme and, further, that this sequence is not divisible into multiple, identical subsequences of significant size, *i.e.* subsequences which include multiple tryptic peptides. It will be convenient to refer to this sequence of 360 residues as the half-unit.

Translating the conclusion as to the existence of half-units into genetic terms, one obtains the corollary that the sequence of base pairs in the structural gene (or genes) for UDP-galactose 4-epimerase contains 3 times 360, or 1080 base pairs. As the first set of experiments does not determine the number of different polypeptides contributing to the half-unit, it also does not specify the number of different structural genes included in the 1080 base pairs. The second and third sets of experiments, in which the amino-terminal residues and the molecular weight in guanidine hydrochloride were examined, represent attempts to determine the number of different polypeptides in the half-unit, and hence the number of different structural genes specifying the primary structure of UDP-galactose 4-epimerase.

One pair of amino-terminal residues is expected per enzyme molecule for each polypeptide in the half-unit, providing that no amino-terminal residues are blocked and that the two half-units are not themselves connected by a peptide bond. Asking the question of whether one, two, or more pairs of amino-terminal residues can be detected, we find (Table II) that only one pair (aspartic acid) is detectable, and therefore reach the simplest conclusion that each half-unit contains but one polypeptide.

An expectation of this conclusion is that the value for the ratio of the molecular weight of the native enzyme (7.9×10^4 (6)) to that determined in guanidine hydrochloride will be 1.0 or 2.0, depending upon whether the two-half units are covalently connected or not. The observed value for this ratio is 1.9 in 4.65 M guanidine hydrochloride and 2.0 in 6.14 M guanidine hydrochloride, consistent with the conclusion of one polypeptide per half-unit, and indicating that the half-units are not covalently connected. More specifically, it indicates that the half-units are not connected by the two disulfide bonds allowed from a comparison of the number of half-cystine residues to the number of free sulfhydryl groups (Table I).

It is of interest to note that the eight half-cystine residues of the enzyme divide into three classes, each of which contains an even number of residues, namely: (*a*) two residues with —SH groups capable of reacting with Ellman's reagent in the native state of the enzyme (see "*Methods*"), (*b*) two residues with —SH groups reacting readily with this reagent only after denaturation in 3.5 M guanidine hydrochloride, and (*c*) the remaining four residues with no —SH groups capable of reacting with either Ellman's reagent or *N*-ethylmaleimide under denaturing conditions. Clearly, the even number of residues in each class is another characteristic of the enzyme consistent with equivalent half-units, each half-unit containing one cysteine residue from each of the first two classes and two half-cystine residues from the third class, which are presumably linked by a disulfide bond.

The structure we have proposed is the simplest one indicated by the data. Clearly, alternate structures are allowed if one assumes the existence of blocked amino-terminal residues in the enzyme, *e.g.* two representatives of each of two different types of polypeptides (α and β) where each α is connected to a β by a disulfide bond and two of the four amino termini are blocked. However, to our knowledge blocked amino termini have not been found in enzymes purified from *E. coli* and, in any case, such less likely possibilities do not affect the conclusion that the genetic sequence necessary to specify the primary structure of the enzyme contains about 1080 base pairs[2] which, in its RNA form, must be translated twice to form one enzyme molecule. We emphasize this conclusion as it is the one used in the kinetic analysis of enzyme synthesis given in an accompanying article (Paper IV of this series (8)).

The proposed structure is peculiar in that two equivalent polypeptides are associated with only one molecule of DPN, a condition which induces the question of how a single binding site can be created from two equivalent subunits. A single site can be created in many ways by asymmetric association of the two polypeptides. On the other hand, if the association of the two polypeptides creates an axis of two-fold symmetry, a single binding site either can exist as a symmetrical entity centered on this axis, or can be created after DPN has been bound to one polypeptide if that binding forces a change in the conformation of the other polypeptide.

An asymmetric association of two equivalent subunits would be expected to yield a structure capable of further polymerization (21), and since the sedimentation data described previously (6) indicate that this does not occur, we prefer the structure with the axis of two-fold symmetry (21). At present there appears no reason to prefer either of the above two modes for deriving a single site from such a symmetrical structure.

REFERENCES

1. Kalckar, H. M., *Advan. Enzymol.*, **20**, 111 (1958).
2. Buttin, G., *J. Mol. Biol.*, **7**, 164 (1963).
3. Lederberg, E. M., in W. Hayes and R. C. Clowes (Editors), *Microbial genetics*, Cambridge University Press, New York, 1960, p. 115.
4. Morse, M. L., *Proc. Nat. Acad. Sci. U. S. A.*, **48**, 1314 (1962).
5. Adler, J., and Kaiser, A. D., *Virology*, **19**, 117 (1963).
6. Wilson, D. B., and Hogness, D. S., *J. Biol. Chem.*, **239**, 2469 (1964).
7. Wilson, D. B., and Hogness, D. S., *J. Biol. Chem.*, **244**, 2137 (1969).

[2] It should be noted that the size of this genetic sequence is not calculated from the value for the molecular weight of the subunits determined in guanidine hydrochloride, but rather is derived from the more accurate value for the molecular weight of the native enzyme (6). Consequently, it is not affected by the possibility that the values determined in guanidine hydrochloride are first approximations resulting from (*a*) lack of knowledge about preferential interactions of water or of guanidine hydrochloride, or (*b*) the absence of data necessary for extrapolation of the observed molecular weight to infinite dilution of the subunits.

8. Wilson, D. B., and Hogness, D. S., *J. Biol. Chem.*, **244,** 2143 (1969).
9. Benesch, R. E., Lardy, H. A., and Benesch, R., *J. Biol. Chem.*, **216,** 673 (1955).
10. Hirs, C. H. W., Moore, S., and Stein, W. H., *J. Biol. Chem.*, **219,** 623 (1956).
11. Moore, S., *J. Biol. Chem.*, **238,** 235 (1963).
12. Smyth, D. G., Blumenfeld, O. O., and Konigsberg, W., *Biochem. J.*, **91,** 589 (1964).
13. Ellman, G. L., *Arch. Biochem. Biophys.*, **82,** 70 (1959).
14. Goodwin, T. W., and Morton, R. A., *Biochem. J.*, **40,** 628 (1946).
15. Helinski, D. R., and Yanofsky, C., *Biochim. Biophys. Acta*, **63,** 10 (1962).
16. Stark, G. R., and Smyth, D. G., *J. Biol. Chem.*, **238,** 214 (1963).
17. Schwartz, D. P., and Pallansch, M. J., *Anal. Chem.*, **30,** 219 (1958).
18. Mazur, R. H., Ellis, B. W., and Cammarata, P. S., *J. Biol. Chem.*, **237,** 1619 (1962).
19. Cohn, E. J., and Edsall, J. T., in *Proteins, amino acids and peptides*, Reinhold Publishing Corporation, New York, 1943, p. 355.
20. Schachman, H. K., and Edelstein, S. J., *Biochemistry*, **5,** 2681 (1966).
21. Monod, J., Wyman, J., and Changeux, J.-P., *J. Mol. Biol.*, **12,** 88 (1965).