Characterization of a R plasmid-associated, trimethoprim-resistant dihydrofolate reductase and determination of the nucleotide sequence of the reductase gene

J. Werner Zolg and Urs J. Hänggi

Institut für Physiologische Chemie, Physikalische Biochemie und Zellbiologie der Universität, Goethestr. 33, 8000 München 2, GFR

Received 17 November 1980

ABSTRACT
The trimethoprim-resistant dihydrofolate reductase associated with the R plasmid R388 was isolated from strains that over-produce the enzyme. It was purified to apparent homogeneity by affinity chromatography and two consecutive gel filtration steps under native and denaturing conditions. The purified enzyme is composed of four identical subunits with molecular weights of 8300. A 1100 bp long DNA segment which confers resistance to trimethoprim was sequenced. The structural gene was identified on the plasmid DNA by comparing the amino acid composition of the deduced proteins with that of the purified enzyme. The gene is 234 bp long and codes for 78 amino acids. No homology can be found between the deduced amino acid sequence of the R388 dihydrofolate reductase and those of other prokaryotic or eukaryotic dihydrofolate reductases. However, it differs in only 17 positions from the enzyme associated with the trimethoprim-resistance plasmid R67.

INTRODUCTION
Trimethoprim prevents the bacterial production of tetrahydrofolate coenzymes by inhibiting the enzyme dihydrofolate reductase (2). The resulting deficiency of tetrahydrofolates leads to impaired protein, RNA, and DNA synthesis and eventually to cell death (3). The production of tetrahydrofolates is not inhibited in bacteria which harbor trimethoprim-resistance plasmids. These cells contain additional dihydrofolate reductases that are insensitive to the drug (4-6). Gene dosage experiments and investigations on the production of the plasmid-associated enzymes in minicells have shown that they are encoded on the plasmid DNA (7-9).

Several trimethoprim-resistance plasmids have been mapped by in vitro recombinant DNA techniques (7-11). These studies have shown that the genes for the resistant dihydrofolate reductases...
of the plasmids R483 and R67 reside on DNA fragments of about 4100 bp and 2400 bp length (9) and that of the plasmid R388 on a DNA segment of less than 1200 bp length (7). In the present study the nucleotide sequence of the latter DNA segment was determined. The resulting amino acid sequences were compared to the amino acid composition of the isolated enzyme. The comparison made it possible to identify the structural gene on the DNA segment, to analyze the organization of the regulatory elements, and to deduce the amino acid sequence of the enzyme.

MATERIALS AND METHODS

Chemicals and enzymes. Acrylamide, 2 times crystallized, was obtained from Serva, Heidelberg. Piperidine, dimethylsulfate, and guanidinium chloride were from E. Merck, Darmstadt, hydrazine from Roth, Karlsruhe, lyophilized alkaline phosphatase (calf intestine), T4 polynucleotide kinase, and NADPH from Boehringer, Mannheim, and 3-(4,5-dimethylthiazolyl-2-)-2,5-diphenyltetrazolium bromide from Sigma, München. Matrex Gel Red A was purchased from Amicon, Witten. All other chemicals used were as described (7). Dihydrofolate was prepared according to (12).

Purification of dihydrofolate reductase. E. coli C carrying pWZ820 or pWZ703 (7) was grown in a fermenter in L-broth to a density of 2-5.10^9 cells/ml. About 60 g of cells were broken up in a French-press in 200-300 ml of 10 mM Tris-HCl, pH 8.0 containing 1 mM mercaptoethanol and 1 mM EDTA (TME-8 buffer). After high speed centrifugation (45 min at 75000 x g, 0°C) solid ammonium sulfate was added first to 2 M, then to 3.6 M. The 2 M precipitate was discarded. The 3.6 M ammonium sulfate precipitate was dissolved in 100-200 ml of TME-8 and applied to a Sephadex G-75 column (10x60 cm) equilibrated with TME-8. 20 ml fractions were collected. The fractions containing the trimethoprim-resistant dihydrofolate reductase were pooled and precipitated by adding ammonium sulfate to 3.6 M, the pellet was dissolved in 50 ml of 10 mM Tris-HCl, pH 9.0 containing 1 mM mercaptoethanol, 1 mM EDTA, and 10 mM NaCl (TMENa-9 buffer), dialyzed against 3 times 1 l of TMENa-9, and applied to a Matrex Gel Red A column (2.5x30 cm) equilibrated with TMENa-9. The column was eluted with the same buffer without using a gradient. The two trimetho-
prim resistant activity peaks (see Results) were pooled and concentrated to 10 ml. 7.6 g of guanidinium chloride and 10 μl of concentrated mercaptoethanol were added, and the mixture was held at 100°C for 5-10 min. After cooling it was applied to a Sephadex G-150 column (3.5x30 cm) which previously was equilibrated with 10 fold concentrated TMENa-9 containing 6 M urea. Elution was with the same buffer mixture. Fractions of 5 ml were collected. The fractions containing dihydrofolate reductase were freed of urea by gel filtration in Sephadex G-25 (3.5x30 cm) equilibrated with 10 mM Na-phosphate, pH 6.8 containing 1 mM mercaptoethanol. The enzyme pool was concentrated to 0.5 ml by flash evaporation and rechromatographed on a small column (1.5x15 cm) of Sephadex G-25 in the phosphate buffer mentioned above. The most active fractions were pooled and stored at -20°C.

Dihydrofolate reductase activity was assayed according to Ref. 13. 2 μM trimethoprim was included in the assay mixture whenever the trimethoprim-resistant activity had to be distinguished from the trimethoprim-sensitive E.coli reductase activity.

Molecular weight determinations. The subunit molecular weight of the purified R388 dihydrofolate reductase was determined by electrophoresis on 7% and 10% polyacrylamide gels containing 0.2% SDS (14). The gels were calibrated with insulin chain B (M_r 3400), aprotinin (M_r 6500), cytochrome C (M_r 12500), and soybean trypsin inhibitor (M_r 21500) (Combithek I, Boehringer). The molecular weight of the purified native enzyme was determined by gel filtration in Sephadex G-200 superfine. The column (1x100 cm) was equilibrated with 10 mM Tris-HCl, pH 8.0, containing 10 mM mercaptoethanol and 1 mM EDTA. It was calibrated with bovine serum albumin (M_r 68000), ovalbumin (M_r 45000), chymotrypsinogen A (M_r 25000), and cytochrome C (M_r 12500) (Combithek II, Boehringer).

Gel electrophoresis. Non-denaturing polyacrylamide gels were prepared and stained for proteins or dihydrofolate reductase activity according to Ref. 15. The conditions used for separating DNA fragments on agarose gels were as described (7). DNA fragments were isolated from agarose by the glass powder adsorption method (16). The sizes of the fragments were determined by coelectrophoresis with pBR322 DNA cleaved with either Sau3A or HpaII.
Restriction enzyme analysis. DNA was isolated and cleaved with restriction enzymes as previously described (7). The restriction nucleases BamHI, EcoRI, Sau3A and HaeIII were kindly donated by T. Igo-Kemenes. AluI, PstI, HhaI, HinfI, Sau96I, and TagI were gifts of R.E. Streeck. HinfI was prepared by chromatography on DEAE-cellulose and heparin-agarose, Sau96I and TagI according to Ref. 17. The latter enzyme was used in a buffer containing 6 mM Tris, pH 7.4, 6 mM MgCl₂, and 0.6 mM dithiothreitol at 50°C. HpaII was prepared by H. Feldmann, and HindIII was obtained from Boehringer, Mannheim.

Terminal labelling and sequence analysis. 20-50 µg of DNA were cleaved with restriction nuclease and the resulting mixture of fragments was treated with phosphatase prior to the separation of the fragments by electrophoresis in polyacrylamide gels. The fragments were recovered from the gels by adding 2 volumes of 1 M NaCl to the gel slices and by passing the gel-NaCl mixture through a syringe. The resulting paste was incubated overnight at 37°C and centrifuged. The supernatant was filtered through Whatman 3MM paper. The isolated fragments were denatured, labelled at their 5' ends with polynucleotide kinase, renatured, and cleaved according to Maxam and Gilbert (18) with minor modifications (19). The reactions specific for C, C+T, A>C, and G were performed according to Ref. 18, and that for A+G according to Ref. 20. Separation of the partial fragments was on 8% or 20% polyacrylamide gels in 8.3 M urea. Autoradiography was at -70°C with and without intensifying screens (SE6, Cawo, Schrobenhausen). Methylated cytosines were detected according to Ref. 21.

RESULTS

Purification and characterization of the R388 dihydrofolate reductase. In a previous study we had found that dihydrofolate reductase is stable under denaturing conditions and that the level of the enzyme is considerably increased in bacteria which carry multiple copies of the reductase gene (7). These two observations were used as part of a novel purification scheme. Starting from cell extracts of bacteria with increased dihydrofolate levels the enzyme was partially purified by ultracentri-
fugation, ammonium sulfate precipitation and gel filtration on Sephadex G-75 as previously described (7). Further purification was achieved by chromatography on triacyl dye agarose (Procion Red HE 3B, Amicon Matrex Gel Red A) which selectively adsorbs NADP$^+$-dependent enzymes (22). Preliminary experiments had shown that the R388 dihydrofolate reductase was retained by the gel. In order to purify the enzyme on a preparative scale, pH and buffer compositions were chosen which allowed the separation of the bulk of the non-dihydrofolate reductase proteins from the enzyme without using a salt gradient (Fig. 1A). 1.5 U of enzyme were retained per ml of affinity gel. Chromosomal dihydrofolate reductase is strongly absorbed and is only eluted at high salt concentrations.

The R388 dihydrofolate reductase displayed a two-peak elution pattern on the Procion Red affinity column (Fig. 1A). However, since no difference between the two fractions could be detected on SDS gels, both were combined. To remove remaining

Figure 1. (A). Affinity chromatography of R388 dihydrofolate reductase. The concentrated and dialyzed Sephadex G-75 pool was chromatographed on Procion Red agarose as described in Materials and Methods. Aliquots of the eluate were assayed for absorbance at 280 nm (Δ—Δ) and for activity (●—●). The bar indicates the fractions which were pooled. (B). Gel filtration under denaturing conditions. The enzyme fractions of the affinity gel were denatured and chromatographed on Sephadex G-150 in the presence of 6 M urea as described in Materials and Methods. Aliquots of 1-10 μl of the eluate were assayed for activity without removing the urea.
non-reductase contaminants, which in their native conformation had copurified with the native R388 dihydrofolate reductase on the first Sephadex G-75 column, the pooled enzyme fractions were denatured by boiling in 6 M guanidinium chloride and chromatographed in Sephadex G-150 in the presence of 6 M urea (Fig. 1B). The R388 dihydrofolate reductase disaggregated under these conditions and was well separated from the larger, non-dihydrofolate reductase proteins. The fractions containing dihydrofolate reductase were freed of urea by gel filtration on Sephadex G-50, concentrated, and rechromatographed on a small column of Sephadex G-50. The resulting enzyme was more than 95% pure as judged by electrophoresis on SDS gels and the overall yield was between 15-20% in the different preparations. The specific activity was about 1.5 U/mg of protein (Table 1).

The molecular weight of the purified, catalytically active enzyme was estimated by gel filtration on Sephadex G-200. R388 dihydrofolate reductase eluted as a single protein and activity peak in a position corresponding to a molecular weight of about 36000. A single protein band was also seen on SDS gels. However, its molecular weight was only about 8400. Since the band showed weak dihydrofolate reductase activity after prolonged staining for the enzyme, it was concluded that the native enzyme is composed of four identical subunits with molecular weights of 8400. A similar subunit structure has been found for the R67 dihydrofolate reductase (23).

Table 1. Purification of R388 dihydrofolate reductase

<table>
<thead>
<tr>
<th>Fraction</th>
<th>Volume</th>
<th>Protein concentration</th>
<th>Activity</th>
<th>Total activity</th>
<th>Specific activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crude extract</td>
<td>300 ml</td>
<td>12.13 mg/ml</td>
<td>127 nM/mg</td>
<td>38 130 nM</td>
<td>10 nM/mg</td>
</tr>
<tr>
<td>(NH₄)₂SO₄ precipitate</td>
<td>170 ml</td>
<td>-</td>
<td>263 nM/mg</td>
<td>44 761 nM</td>
<td>- nM/mg</td>
</tr>
<tr>
<td>Sephadex G-75 pool</td>
<td>260 ml</td>
<td>0.77 mg/ml</td>
<td>122 nM/mg</td>
<td>31 746 nM</td>
<td>157 nM/mg</td>
</tr>
<tr>
<td>Procion Red agarose</td>
<td>100 ml</td>
<td>0.42 mg/ml</td>
<td>181 nM/mg</td>
<td>18 060 nM</td>
<td>430 nM/mg</td>
</tr>
<tr>
<td>Sephadex G-50</td>
<td>5 ml</td>
<td>0.93 mg/ml</td>
<td>1266 nM/mg</td>
<td>6 330 nM</td>
<td>1360 nM/mg</td>
</tr>
</tbody>
</table>
The amino acid composition of the purified enzyme is shown in Table 2. The calculated molecular weight (8020) is in close agreement with the value determined from SDS gels.

Identification of the coding sequence of the dihydrofolate reductase gene. Our previous analysis of the trimethoprim-resistance gene of R388 has shown that the plasmid-induced dihydrofolate reductase is encoded on a DNA segment of less than 1200 bp length (7). Further cloning experiments (not shown) suggested that the entire segment might be essential for proper expression of the gene. Therefore the DNA was cleaved with various restriction nucleases and sequenced by the Maxam and Gilbert technique. The restriction map and the sequence strategy

Table 2. Amino acid composition of R388 dihydrofolate reductase. About 0.6 nmol of purified enzyme were hydrolyzed for 30 min at 160°C in 5.7 N HCl as described by Wachter and Werhahn (24).

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>nmoles found/nmol of enzyme</th>
<th>suggested theoretical content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cys c)</td>
<td>0.98</td>
<td>1</td>
</tr>
<tr>
<td>Asx</td>
<td>4.06</td>
<td>4</td>
</tr>
<tr>
<td>Asp</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Asn</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>Met d)</td>
<td>0.32</td>
<td>1</td>
</tr>
<tr>
<td>Thr</td>
<td>2.98</td>
<td>3</td>
</tr>
<tr>
<td>Ser</td>
<td>6.55</td>
<td>7</td>
</tr>
<tr>
<td>Glx</td>
<td>10.02</td>
<td>10</td>
</tr>
<tr>
<td>Glu</td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>Gln</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Pro</td>
<td>4.06</td>
<td>4</td>
</tr>
<tr>
<td>Gly</td>
<td>8.90</td>
<td>9</td>
</tr>
<tr>
<td>Ala</td>
<td>10.68</td>
<td>11</td>
</tr>
<tr>
<td>Val</td>
<td>6.21</td>
<td>8</td>
</tr>
<tr>
<td>Ile</td>
<td>1.19</td>
<td>1</td>
</tr>
<tr>
<td>Leu</td>
<td>5.07</td>
<td>5</td>
</tr>
<tr>
<td>Tyr</td>
<td>2.38</td>
<td>3</td>
</tr>
<tr>
<td>Phe</td>
<td>1.93</td>
<td>2</td>
</tr>
<tr>
<td>His</td>
<td>1.00</td>
<td>1</td>
</tr>
<tr>
<td>Lys</td>
<td>3.23</td>
<td>3</td>
</tr>
<tr>
<td>Arg</td>
<td>2.56</td>
<td>3</td>
</tr>
<tr>
<td>Trp e)</td>
<td>1.58</td>
<td>2</td>
</tr>
</tbody>
</table>

M_r: 8020 M_L: 8266

a) mean of two determinations; b) deduced from the nucleotide sequence shown in Fig. 3 and 4; c) Cystein was determined as cysteic acid; d) Methionine was determined as methionine sulfone; e) Tryptophan was determined according to Beaven and Holiday (25)
are shown in Fig. 2. The sequence, 1125 nucleotides long, was determined, about 80% of which was confirmed by sequencing both DNA strands. The whole nucleotide sequence is shown in Fig. 3. The GC-content is 60% between bases 1 and 400, 50% between bases 400 and 500, and 63% in the remaining part of the DNA. The overall GC-content is 56%. Three methylated cytosines were found at positions 234, 333, and 959.

Eight open reading frames could be detected by locating the positions of the start and stop codons in the nucleotide sequence. Three have no stop codons within the sequenced part. The lengths of the proteins which are encoded in the open reading frames vary between 43 and 177 amino acids or more. In order to find the coding sequence of the dihydrofolate reductase gene, the amino acid compositions of all proteins that could be derived from the nucleotide sequence were compared to the composition of the isolated enzyme. Only a single protein had the composition, the length, and the characteristic His/Ile/Met/Cys (1:1:1:1), Leu/Ile (5:1), and Tyr/Trp (3:2) ratios of the isolated enzyme (cf. Table 2). Therefore it was concluded that this protein is the dihydrofolate reductase and that the corresponding nucleo-

Figure 2. Restriction map of the DNA segment carrying the R388 dihydrofolate reductase gene and strategy used for determining the nucleotide sequence. The numbers refer to the cleavage sites on the DNA sequence shown in Fig. 3. The directions in which the DNA fragments were sequenced and the distances are indicated by arrows. The reductase structural gene is delineated by the bar.
Figure 3. Nucleotide sequence of the R388 DNA segment carrying the dihydrofolate reductase gene. The coding sequence for the R388 dihydrofolate reductase is enclosed.

tide sequence is the structural gene for the enzyme. This sequence is enclosed in Fig. 3.

Comparison of the amino acid sequence of the R388 and R67 dihydrofolate reductases. The complete amino acid sequence deduced from the nucleotide sequence of the R388 dihydrofolate
The dihydrofolate reductase gene is shown in Fig. 4. Extensive homologies can be found to the amino acid sequence of the dihydrofolate reductase isolated from the trimethoprim-resistance plasmid R67 (26). Both enzymes have exactly the same number of amino acid residues and 61 of the 78 amino acids are identical. The remaining 17 amino acids, for the most part, are conservative replacements, eight of which can be explained by transitions of single bases (amino acid residues 3, 6, 8, 9, 17, 18, and 21; Fig. 4) and two by transversions of single bases (residues 49 and 61). At least two substitutions are needed to explain the remaining 7 exchanges (residues 2, 10, 15, 20, 26, 77, and 78).

### Organization of the dihydrofolate reductase gene

The dihydrofolate reductase gene (cf. Fig. 3) is only 234 bp long. Its promoter contains the sequence CATAAT (pos. 744-749) and GTTGCATT (pos. 726-734) which differ only minimally from the -10 and -35 consensus sequences described by Rosenberg and Court (27). In analogy to other mRNAs by which the dihydrofolate reductase mRNA is expected to start at the A at position 756, seven nucleotides downstream of the ubiquitous T of the -10 hexamer (pos. 749). Between the start of the mRNA and the AUG initiation codon (Fig. 4), the sequence is compared to that of the R67 enzyme (26). Identical amino acids are enclosed. The number above the nucleotide triplets refer to the amino acid residues.

![Amino acid sequence of the R388 dihydrofolate reductase](image)

**Figure 4.** Amino acid sequence of the R388 dihydrofolate reductase. The sequence is compared to that of the R67 enzyme (26). Identical amino acids are enclosed. The number above the nucleotide triplets refer to the amino acid residues.
tion codon there is an untranslated region of 16 nucleotides which contains the sequence AAGGA (pos. 761-765) that could serve as a ribosome binding site (28). The mRNA can form a hairpin structure in this region by pairing GGAUUC (763-768) to GAAGCC (pos. 790-795). The initiation codon thus would be exposed in a non-paired loop.

Two GC-rich sequences with dyad symmetries are found at the 3' side of the gene: the sequence GCCGCTA (pos. 1054-1059) and TGACGC (pos. 1065-1070) and the sequence AGTCG (pos. 1094-1099) and CGACT (pos. 1108-1112). Both inverted repeat structures resemble the terminator signals (27) and thus might be involved in termination of mRNA transcription. There are two stop codons in phase between the end of the coding sequence and those structures (TAA 1006-1008 and TAG 1051-1053).

DISCUSSION

The amino acid sequence of the R388 dihydrofolate reductase shown in Fig. 4 bears no apparent homology to the sequences of other prokaryotic or eukaryotic dihydrofolate reductases (for a recent compilation see Ref. 29). This lack of homology raises the intriguing question of whether the R388 and the similarly organized R67 dihydrofolate reductase represent a novel type of dihydrofolate reductases or if these enzymes are non-reductase proteins with fortuitous reductase activities. The latter possibility is suggested by the low turnover number which can be deduced from the specific activity of the purified R388 enzyme. Its value (50-100 moles/min/mole of enzyme) is one to two orders of magnitude lower than those of other dihydrofolate reductases (3). However, Smith and Burchall (30), who have asked the same question for the R67 enzyme, could detect no other activity besides that of the dihydrofolate reducing one.

A comparison of the amino acid sequences of the R388 and the R67 dihydrofolate reductase gives instructive insights into the organization of the two enzymes. Of the 17 amino acids which are different, all but four map at the amino- and carboxy termini of the enzymes (Fig. 4). The sequences between amino acid residues 22 and 76 are virtually identical. This suggests that only the central part of the molecule is essential for the acti-
vity. This is supported by the conspicuous accumulation of functional groups therein (e.g. Arg-Val-Arg-Lys-Lys-Ser (29-34), Trp-Tyr-Cys-Thr (45-48), Thr-Pro-Glu-Gly-Tyr (51-55), Glu-Ser-Glu (58-60), or Ile-Tyr-Pro-Val (68-71)). In this respect it would be interesting to know if there are kinetic differences between the R67 and the R388 dihydrofolate reductase since glutamine residue 49 in the R67 dihydrofolate reductase is replaced by a lysine residue in the R388 enzyme.

The amount of R388 dihydrofolate reductase activity is rather low in E.coli (4,7). This might be due to a low level of gene expression. However, the organization of the promoter region is as in other prokaryotic genes. The nucleotide sequences of the "Pribnow box" (31), of the RNA polymerase recognition site (32,33), and of the ribosome binding site (28) have no more deviations from the prototype sequences than other genes (27). This suggests that the low amount of enzyme in E.coli is mostly, if not exclusively due to the low specific activity of the enzyme. However, it should be recalled that we never have obtained a trimethoprim-resistant recombinant plasmid that is smaller than about 1200 bp. This might well indicate that far more sequence information is needed to express the gene than the information stored in the putative promoter region and that the auxiliary sequences modulate the rate of gene expression.

The nucleotide sequence at the end of the gene contains two termination codons in phase. It is conceivable that the first stop codon is translated in suppressor positive strains. This would add the sequence X-Gln-Leu-Ala-Gln-Glu-Gly-Ser-Pro-Pro-Ala-Glu-Leu-Cys-Arg to the carboxy end of the protein and increase the molecular weight by about 1600. The recent finding of Fling and Elwell that the R388 dihydrofolate reductase is somewhat larger than the R67 enzyme in minicell producing strains (9) might be explained by this mechanism.

The R388 dihydrofolate reductase gene is unusually small in size. With a length of less than 240 bp it should be possible to insert the gene into viral genomes without grossly affecting the size of the viral DNA. Since the R388 dihydrofolate reductase gene also confers resistance to methotrexate, which is toxic for fungi (34,35), plants (36), and animal cells (2),
recombinants between the dihydrofolate reductase gene and the viruses could possibly be used as selectable DNA vectors to transfer foreign genes into the cells of fungi, plants, and animals.

ACKNOWLEDGEMENTS

We would like to thank Dr. H.G. Zachau for helpful suggestions and discussions during this work, Dr. E. Wachter for performing the amino acid analysis, and Ms. Ch. Graf, Ms. I. Wagner, and Mr. Neumaier for their excellent technical assistance. This work was supported in parts by the Deutsche Forschungsgemeinschaft (Forschergruppe Genomorganisation) and the Fonds der Chemischen Industrie. J.W. Zolg was supported by Cusanuswerk, Bischöfliche Studienförderung.

REFERENCES

1. Present address: Institute of Animal Genetics, University of Edinburgh, West Mains Road, GB Edinburgh EH9 3JN