Notides, A. C., Hamilton, D. E., & Rudolph, J. H. (1973) *Endocrinology 93*, 210-216.
Okey, A. B., Bondy, G. P., Mason, M. E., Kahl, G. F., Eisen, H. J., Guenthner, T. M., & Nebert, D. W. (1979) *J. Biol. Chem. 254*, 11636-11648.
Okey, A. B., Bondy, G. P., Mason, M. E., Nebert, D. W., Forster-Gibson, C. J., Muncan, J., & Dufresne, M. J. (1980) *J. Biol. Chem. 255*, 11415-11422.
Perdew, G. H. (1988) *J. Biol. Chem. 263*, 13802-13805.
Poellinger, L., Lund, J., Gillner, M., Hansson, L.-A., & Gustafsson, J.-A. (1983) *J. Biol. Chem. 258*, 13535-13542.
Poland, A., & Knutson, J. C. (1982) *Annu. Rev. Pharmacol. Toxicol. 2*, 517-554.
Poland, A., & Glover, E. (1987) *Biochem. Biophys. Res. Commun. 146*, 1439-1449.
Poland, A., & Glover, E. (1988) *Arch. Biochem. Biophys. 261*, 103-111.
Poland, A., Glover, E., & Kende, A. S. (1976) *J. Biol. Chem. 251*, 4936-4946.
Poland, A., Glover, E., Ebetino, F. H., & Kende, A. S. (1986) *J. Biol. Chem. 261*, 6352-6365.
Pratt, W. B. (1987) *J. Cell. Biochem. 35*, 51-68.
Prokipcak, R. D., & Okey, A. B. (1988) *Arch. Biochem. Biophys. 267*, 811-828.
Puca, G. A., Nola, E., Sica, V., & Bresciani, F. (1977) *J. Biol. Chem. 252*, 1358-1366.
Puca, G. A., Abbondanza, C., Nigro, V., Armetta, I., Medici, N., & Molinari, A. M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5367-5371.
Radanyi, C., Renoir, J.-M., Sabbah, M., & Baulieu, E. E. (1989) *J. Biol. Chem. 264*, 2568-2573.
Reker, C. E., Kovacic-Milivojevic, B., Eastman-Reks, S. B., & Vedekis, W. V. (1985) *Biochemistry 24*, 196-204.
Rowley, D. R., Premont, R. T., Johnson, M. P., Young, C. Y. F., & Tindall, D. J. (1986) *Biochemistry 25*, 6988-6995.
Rucci, G., & Gasiewicz, T. A. (1988) *Arch. Biochem. Biophys. 265*, 197-207.
Schmidt, T. J., & Litwack, G. (1982) *Physiol. Rev. 62*, 1131-1192.
Schmidt, T. J., Miller-Diener, A., Webb, M. L., & Litwack, G. (1985) *J. Biol. Chem. 260*, 16255-16262.
Schmidt, T. J., Diehl, E. E., Davidson, C. J., & Puk, M. J. (1986) *Biochemistry 25*, 5955-5961.
Tiengrungroj, W., Pratt, S. E., Grippo, J. F., Holmgren, A., & Pratt, W. B. (1987) *J. Steroid Biochem. 28*, 449-457.
Tukey, R. H., Hannah, R. R., Negishi, M., Nebert, D. W., & Eisen, H. J. (1982) *Cell 31*, 275-284.
Whitlock, J. P., Jr. (1987) *Pharmacol. Rev. 39*, 147-161.
Wilhelmsson, A., Wikstrom, A.-C., & Poellinger, L. (1986) *J. Biol. Chem. 261*, 13456-13463.
Willmann, T., & Beato, M. (1986) *Nature 324*, 688-691.

A Developmentally Regulated Gene of Trypanosomes Encodes a Homologue of Rat Protein-Disulfide Isomerase and Phosphoinositol-Phospholipase C†,‡

MiMi P. Hsu,§ Michael L. Muhich,∥ and John C. Boothroyd*
Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, California 94305-5402
Received December 12, 1988; Revised Manuscript Received April 17, 1989

ABSTRACT: We have isolated and characterized a developmentally regulated gene in *Trypanosoma brucei*, arbitrarily termed BS2. BS2 mRNA is substantially more abundant in bloodstream-form trypanosomes than in procyclic culture forms. Its nucleotide sequence reveals a single contiguous open-reading frame of 497 codons and is predicted to encode a protein of ~55.5 kilodaltons. A search of the NBRF protein data base revealed that within the predicted amino acid sequence are two of the evolutionarily conserved redox sites typified by thioredoxin of bacteria. Of this family of proteins, the recently sequenced rat genes encoding protein-disulfide isomerase (PDI) and form I phosphoinositide-specific phospholipase C (PIPLC) showed homology extending over the length of all three proteins (i.e., between BS2, PDI, and PIPLC). Although this homology includes the acidic C-terminus characteristic of proteins localized to the lumen of the enoplasmic reticulum, the BS2 product is predicted to possess multiple sites for N-linked glycosylation while PDI and PIPLC have none. Possible roles of the BS2 gene product in trypanosome physiology are discussed.

The complex life cycle of pathogenic trypanosomes involves developmental stages in both a mammalian host and the insect vector *Glossina*. Survival and growth in these very different environments demand not only gross morphological changes but also adaptive changes in important biochemical pathways. For example, transition from the bloodstream form of the African *Trypanosoma brucei* to the procyclic form found in the insect involves the rapid shedding of their dense surface coat (Vickerman, 1985; Overath et al., 1983). This developmental switch is also accompanied by a change from utilization of the glycolytic pathway as the principal energy source to activation of a greatly expanded mitochondrial system of oxidative respiration (Vickerman, 1962, 1965; Opperdoes, 1987). These morphologic and metabolic changes are associated with

† This work was supported by grants from the NIH (AI21025) and the John D. and Catherine T. MacArthur Foundation. M.L.M. was the recipient of an NIH postdoctoral fellowship (AI07558). J.C.B. holds a Burroughs Wellcome Scholarship in Molecular Parasitology.
‡ The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02865.
* Author to whom correspondence should be addressed.
§ Present address: Department of Microbiology, University of Iowa, Iowa City, IA 52242.
∥ Present address: Division of Chemistry 147-75, California Institute of Technology, Pasadena, CA 91125.

0006-2960/89/0428-6440$01.50/0

the precise up or down regulation of the corresponding genes (Jasmer et al., 1985; Overath et al., 1983; Osinga et al., 1985).

Thioredoxin is a ubiquitous protein which plays a role in many biological activities [for a review, see Holmgren (1985)]. First characterized in *Escherichia coli* (Holmgren et al., 1975), it has since been isolated from various sources such as spinach chloroplasts (Maeda et al., 1984) and rat liver (Luthman & Holmgren, 1982). It is generally a small protein with an active site including two cysteine residues which readily exist either reduced or oxidized, forming an intramolecular disulfide bridge. Thioredoxin and related proteins participate in various biological functions through the reversible oxidation of this redox-active center.

Recently, two rat proteins, protein-disulfide isomerase (PDI; Edman et al., 1985) and form I phosphoinositide-specific phospholipase C (PIPLC; Bennett et al., 1988), have been identified and found to contain two domains, each with clear homology to thioredoxin. Both proteins are about 55–56 kilodaltons (kDa) as compared with ~12 kDa for thioredoxin. PDI catalyzes the formation of cysteine disulfide bonds during the process of protein folding in vivo and in vitro. It is also known to possess a C-terminal signal (Lys-Asp-Glu-Leu-COOH) which causes it to be localized in the lumen of the endoplasmic reticulum where, presumably, it carries out its catalytic role in protein folding. PIPLC form I is one of several isoenzymes involved in signal transduction through hydrolysis of phosphatidylinositol 4,5-bisphosphate to the second messengers 1,2-diacylglycol and inositol 1,4,5-triphosphate (Bennett et al., 1988). The other PIPLC forms have different structures and may function in association with different subcellular compartments and/or other protein cofactors.

In this report, we describe the cloning and characterization of a developmentally regulated trypanosome gene, arbitrarily termed BS2 (for bloodstream-specific gene 2), which encodes a product with clear homology to both rat PDI and PIPLC form I, although it differs in probably being a glycoprotein. Possible reasons for its developmental regulation will de discussed.

Materials and Methods

Trypanosomes and Nucleic Acid Preparation. *Trypanosoma brucei brucei* strain 427 was used throughout this work (Cross, 1975). Growth of the bloodstream-form parasites in animals and purification of nucleic acids were as described previously (Muhich & Boothroyd, 1988). Procyclic-form parasites (cooresponding to the form found in insects) were grown in axenic culture in vitro essentially as described (Brun & Schonenberger, 1981).

Cloning of the BS2 Gene. A cosmid library (Boothroyd & Beals, 1987) of *T. brucei* was screened by using a probe to cDNA 14 (see Results). From this cosmid, a 2.8-kb *Eco*RV/*Bam*HI fragment was isolated containing the BS2 gene and inserted into the *Eco*RV/*Bam*HI sites of Bluescript KS+ plasmid vector (Stratagene Inc.).

DNA Sequence. The DNA sequence (Figure 3) was determined primarily by the chain termination method (Sanger et al., 1977; Sanger & Coulson, 1978) with some regions confirmed by the chemical modification method (Maxam & Gilbert, 1980). Single-stranded DNA templates for dideoxy sequencing were generated either from nested deletions in Stratagene's Bluescript Vectors or from clones of fragments directly subcloned into M13 mp10, mp11, mp18, or mp19 vectors (Yanisch-Perron et al., 1985). All sequence was determined on both strands.

Computer-Assisted Homology Search and Alignment. The FASTP computer algorithm (as contained in the IBI/Pustell Cyborg Software; International Biotechnologies, Inc.) was used to compare the derived amino acid sequence of the BS2 gene to those found in the NBRF (National Biochemical Research Foundation) protein library. Homology to rat PIPLC (not then in the NBRF data base) was noted independently on the basis of recent publication of its thioredoxin-like domains (Bennett et al., 1988).

Genomic DNA Blots and Hybridization. *T. brucei* nuclear DNA (1.5 μg) was digested with various restriction endonucleases, electrophoretically separated on a horizontal 0.7% agarose gel, and transferred following sequential (0.5 h each) treatments in 0.5 M NaOH/1.5 M NaCl, then 0.5 M Tris (pH 7.4)/3 M NaCl, and, finally, 1 M ammonium acetate/0.02 M NaOH. Nick translation (Rigby et al., 1977) and hybridization (Muhich et al., 1983) of the probe were essentially as previously described.

RNA Blot Analysis. Whole cell RNA (10 μg) from either bloodstream- or culture-form *T. brucei* was electrophoretically separated on a horizontal 1.2% agarose/formaldehyde denaturing gel. The gel was transferred to nitrocellulose and probed as previously described (Muhich & Boothroyd, 1988). Briefly, transfer was to 0.22 μm nitrocellulose paper with 20× SSC (1× SSC = 0.15 M sodium chloride/0.015 M sodium citrate). The filter was baked in a vacuum oven at 80 °C for 2 h followed by hybridization with radiolabeled probe at 42 °C in HMA-50: 5× SSC, 5× Denhardt's (1× Denhardt's = 0.02; Ficoll 400, 0.02% bovine serum albumin, and 0.02% polyvinylpyrrolidine), 50% deionized formamide, 0.2% SDS, and 0.5 mg/mL calf thymus DNA. Washes were at 63 °C in 0.1× SSC/0.1% SDS.

RNase Protection Analysis. Uniformly labeled antisense RNA probes of high specific activity were synthesized by using either T3 or T7 RNA polymerase as described (Melton et al., 1984). Probe (>2000 cps) was coprecipitated with 10–15 μg of bloodstream-form trypanosome whole cell RNA plus 50 μg of *E. coli* tRNA as carrier RNA. Subsequent treatments were as previously described (Muhich & Boothroyd, 1988). Protection products were analyzed on a 5% acrylamide/7 M urea gel.

Results

Identification and Isolation of the BS2 Gene. As part of our studies on gene expression in trypanosomes, we have been studying a single-copy gene whose expression is apparently constitutive (arbitrarily termed "cDNA 14": D. A. Campbell, M. P. Hsu, M. L. Muhich, and J. C. Boothroyd, unpublished results). Cosmids containing this gene had been previously isolated and analyzed (to be described elsewhere). During the course of those studies, we noted that there was a gene located a short distance upstream of cDNA 14 which, as will be detailed below, became interesting to us for its developmental regulation and apparent coding function. We term this gene BS2 for bloodstream-specific gene 2 in the absence of definitive knowledge of its function.

Genomic Organization of BS2. Figure 1 presents the restriction map for the BS2 gene as determined from the cloned gene and confirmed by Southern blot analysis of genomic DNA (Figure 2).

The data presented in Figure 2 also indicate that unlike many trypanosome genes studied to date, which are frequently arranged as closely spaced tandem reiterations, the BS2 gene is apparently present in trypanosomes in a single copy per haploid genome. When a variety of restriction endonucleases were used, the genomic Southern blots revealed only one band except for enzymes which cut within the probe. Although it is possible that there is a duplication of a very large segment

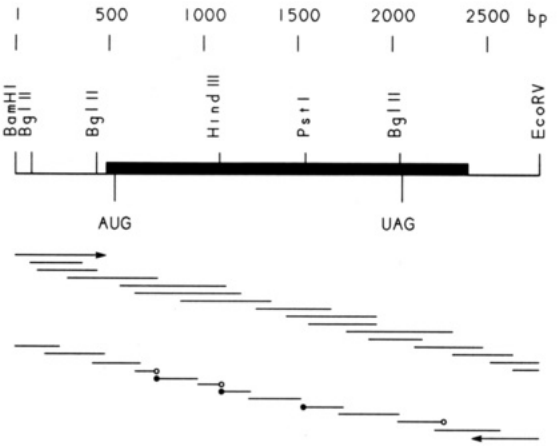

FIGURE 1: Restriction map and sequencing strategy for the trypanosome BS2 gene. The solid box represents the protein-coding exon (splice acceptor to polyadenylation site) with the deduced start and stop codons (see text) also indicated. Below this is the strategy used to determine the sequence with the arrow indicating the 5′–3′ direction for the first of each group of dideoxy sequencing reactions. The lines with open and closed circles indicate the sequence obtained by the chemical modification method following 5′ or 3′ end-labeling of the antisense strand, respectively.

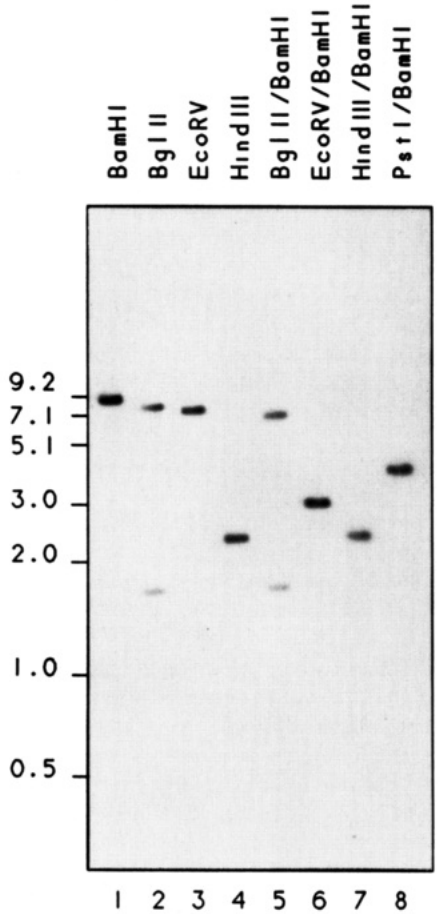

FIGURE 2: Genomic organization of the trypanosome BS2 gene. Samples (1.5 μg) of genomic DNA were digested with the indicated restriction endonucleases, electrophoresed on a 0.7% agarose gel, blotted to nitrocellulose, and hybridized to a nick-translated *Pst*I/*Eco*RV fragment (from positions 1515–2750 in Figure 1). The markers are from the kilobase ladder from BRL.

of the genome (so that all digests give only a single-gene pattern), such duplications have not been observed for other trypanosome genes (Boothroyd, 1989).

Nucleotide Sequence of BS2. The complete nucleotide sequence of the BS2 gene is given in Figure 3. A single

```
   1 GGATCCACAA  CTTTACCCTA  TAGGATAGGG  GGATGTAGTG  AAGAGTAGCA  GTACCGATAA
  61 CAAAATGAAG  ATCTTCGAGC  GCGAGTCACA  CTTCTTGGAG  CCTTCCCCGT  CACTTCGAGC
 121 AGGTGCCAAC  AGGTTTTGAT  CATCTCACTG  TTTGTTAAAG  GTGTTTACAG  AAAGTGTGGT
 181 GCAGTTGAGT  TGAGTCGGGT  CTACTGAGTT  TACTCCTCAG  TCAAAAAATG  AGAGGGGTTG
 241 AGGGGTCGTC  ACGCACCTGA  TGTTGGCTGA  CGGTCCCCTT  TACAGGGCAT  CCGGGGCGAA
 301 TACAGGCGGT  TGTATGCTGC  TGGATCAGTG  ACTCTTGTGA  TACATTTATT  TTTTCGTGCC
 361 ACCGCGAGTG  CTGGCATTAC  GTTGGTGTCG  TACTTTCCTG  TCACTGCATG  CCGCTCCGAA
 421 GATCTCTCGT  GTCAAAATTT  GCTTCGCACC  CCTTTTCTCC  TCTTTGCTAG  GTGAACGTTT
 481 GTTATACTGT  TTACTTTTTT  TTTATTAAGT  TAATAAAAAG  CGTCACG ATG CGC GCT ATT TTT
                                                          Met Arg Ala Ile Phe
                                                          ***

543  CTT GTT GCT CTT GCG CTC GCC ACC ATG CGC GAG TCG ACT GCG GAA TCC CTG AAG CTC ACC
     Leu Val Ala Leu Ala Leu Ala Thr Met Arg Glu Ser Thr Ala Glu Ser Leu Lys Leu Thr

603  AAA GAG AAC TTC AAC GAG ACC ATC GCA AAG AGT GAA ATA TTT CTC GTG AAG TTT TAT GTT
     Lys Glu Asn Phe Asn Glu Thr Ile Ala Lys Ser Glu Ile Phe Leu Val Lys Phe Tyr Val

663  GAC ACC TGT GGA TAC TGC CAG ATG CTC GCA CCA GAG TGG GAG AAG GCG GCC AAT GAA ACT
     Asp Thr Cys Gly Tyr Cys Gln Met Leu Ala Pro Glu Trp Glu Lys Ala Ala Asn Glu Thr
              ●           ●

723  ATT GAC AAT GCA CTT ATG GGC GAA GTC GAC TGT CAC TCA CAA CCT GAA TTG GCG GCT AAC
     Ile Asp Asn Ala Leu Met Gly Glu Val Asp Cys His Ser Gln Pro Glu Leu Ala Ala Asn

783  TTT TCC ATC AGG GGA TAC CCA ACA ATC ATT CTC TTT CGT AAT GGC AAG GAG GCG GAG CAC
     Phe Ser Ile Arg Gly Tyr Pro Thr Ile Ile Leu Phe Arg Asn Gly Lys Glu Ala Glu His

843  TAC GGT GGT GCC CGC ACG AAG GAC GAT ATT ATC AAG TAC ATT AAA GCG AAC GTT GGA CCG
     Tyr Gly Gly Ala Arg Thr Lys Asp Asp Ile Ile Lys Tyr Ile Lys Ala Asn Val Gly Pro

903  GCA GTG ACG CCA GCG TCC AAT GCT GAA GAG GTG ACA AGG GCA AAG GAG GAG CAT GAT GTA
     Ala Val Thr Pro Ala Ser Asn Ala Glu Glu Val Thr Arg Ala Lys Glu Glu His Asp Val

963  GTG TGC GTT GGC CTG ACA GCC AAC AAT AGC ACT TCA CTT TCT ACC ACA CTG GCC GAG GCC
     Val Cys Val Gly Leu Thr Ala Asn Asn Ser Thr Ser Leu Ser Thr Thr Leu Ala Glu Ala

1023 GCG CAG TCG TTC AGA GTC AGC CTC AAA TTC TTT GAA GCA GAA CCT AAG CTT TTC CCC GAC
     Ala Gln Ser Phe Arg Val Ser Leu Lys Phe Phe Glu Ala Glu Pro Lys Leu Phe Pro Asp

1083 GAG AAG CCC GAG ACC ATT GTG GTG TAC AGG AAG GGA GGC GAA AAG GAG GTT TAC GAT GGC
     Glu Lys Pro Glu Thr Ile Val Val Tyr Arg Lys Gly Gly Glu Lys Glu Val Tyr Asp Gly

1143 CCC ATG GAA GTA GAG AAG TTG ACA GAA TTC CTT CAA ATT TCT CGC GTT GCC TTT GGG GGT
     Pro Met Glu Val Glu Lys Leu Thr Glu Phe Leu Gln Ile Ser Arg Val Ala Phe Gly Gly

1203 GAA ATT ACC CCG GAG AAC TAT CAA TAC TAC TCA GTT ATA AAA CGG CCC GTA GGA TGG GCT
     Glu Ile Thr Pro Glu Asn Tyr Gln Tyr Tyr Ser Val Ile Lys Arg Pro Val Gly Trp Ala

1263 ATG GTC AAG CCT AAT GAA ACT GCC TCG ATC GAG TTG AAA GAA AGC CTC ACA GAA GTG GGT
     Met Val Lys Pro Asn Glu Thr Ala Ser Ile Glu Leu Lys Glu Ser Leu Thr Glu Val Gly

1323 AAA AAA ATG CGA TCC CAT ATG GTA GTT CTC TGG GTT AAC ATT TCG AAA CAT CCG GTT TGG
     Lys Lys Met Arg Ser His Met Val Val Leu Trp Val Asn Ile Ser Lys His Pro Val Trp

1383 AGG GAC TTC GGT GTC CCG GAG GAC GCC AAG TAC CCC GCC TTT TTG GCT ATC CAC TGG GGC
     Arg Asp Phe Gly Val Pro Glu Asp Ala Lys Tyr Pro Ala Phe Leu Ala Ile His Trp Gly

1443 GCC AAT TAC TTG CAT TCC ACA GCG GAG GTT GTG ACC CGC GAG TCC CTT GAG AAG TTC ATC
     Ala Asn Tyr Leu His Ser Thr Ala Glu Val Val Thr Arg Glu Ser Leu Glu Lys Phe Ile

1503 CTG GAG TTT GCT GCA GGT CGT GTC GAG CCC ACC ATC AAG TCC CTC CCT GTC CCC GAG GTG
     Leu Glu Phe Ala Ala Gly Arg Val Glu Pro Thr Ile Lys Ser Leu Pro Val Pro Glu Val

1563 GAG ACT GTT GAT GGC AAA ACC ACC ATC GTG GCA AAA ACA ATG CAG AAG CAC CTT ACG TCG
     Glu Thr Val Asp Gly Lys Thr Thr Ile Val Ala Lys Thr Met Gln Lys His Leu Thr Ser

1623 GGA AAA GAC ATG CTC ATC CTT TTC TTT GCC CCT TGG TGC GGT CAC TGC AAG AAT TTC GCT
     Gly Lys Asp Met Leu Ile Leu Phe Phe Ala Pro Trp Cys Gly His Cys Lys Asn Phe Ala
                                                 ●           ●

1683 CCA ACC TTT GAT AAG ATC GCG AAG GAG TTC GAT GCA ACC GAC CTT ATC GTT GCC GAG CTT
     Pro Thr Phe Asp Lys Ile Ala Lys Glu Phe Asp Ala Thr Asp Leu Ile Val Ala Glu Leu

1743 GAT GCC ACA GCC AAT TAC GTT AAC TCG TCT ACT TTT ACA GTT ACA GCA TTT CCG ACG GTG
     Asp Ala Thr Ala Asn Tyr Val Asn Ser Ser Thr Phe Thr Val Thr Ala Phe Pro Thr Val

1803 TTC TTC GTC CCG AAT GGA GGG AAA CCG GTT GTT TTC GAA GGT GAA CGC TCA TTT GAA AAC
     Phe Phe Val Pro Asn Gly Gly Lys Pro Val Val Phe Glu Gly Glu Arg Ser Phe Glu Asn

1863 GTC TAC GAA TTT GTT CGC AAA CAT GTG ACA ACG TTT AAG GTG AGT GAG AAA CCA GCT AAT
     Val Tyr Glu Phe Val Arg Lys His Val Thr Thr Phe Lys Val Ser Glu Lys Pro Ala Asn

1923 GTG ACA GAG GAG AAG AAA AGC GAG GAA GAG AAC AAG AGC AGT AAG AGC AAT GAA AGT AAT
     Val Thr Glu Glu Lys Lys Ser Glu Glu Glu Asn Lys Ser Ser Lys Ser Asn Glu Ser Asn

1983 GAT AGC AAT GAG AGC AAT GTT GAT AAA CAA GAT CTG TAG CAAGATTAA GAAAACGTTT
     Asp Ser Asn Glu Ser Asn Val Asp Lys Gln Asp Leu ---

2041 GGCCCAGTTT  CAAATCCTTA  TCCACTTAGC  TATTTTCGCT  TTTCGGCTCC  GATGATGTGG
2101 GCAACAATGG  TTTGCAAGTT  GCCGGATGGA  AGACAAAGTG  AAGAACCCCT  TATATATGAA
2161 TAGAAGGGGT  CGTTGGCGGG  GGTGGGTTAT  GGGGCTAACG  TAATGCCAGG  TAATAGAGAT
2221 GATAGTGGTA  AAAGCGGGGT  ATCTTTAAGA  AATTGAAATG  TCGGAAGGGA  GGGGGGGGGG
2281 ATAAATAACG  GCGGAGGGTC  GGGGGTGCTC  CGCGCCTCTA  GAGTCCTACT  CGCTGTAGAT
2341 AATAGGTATA  TAAAAAGATT  TGAAAGGGGC  GACTGGTAAG  AAACTCATAT  TCATGCCTTT
2401 CTATGTGGGC  TTTGGCTGGT  TCCTTGCGTC  ACCGTTTCTT  CATCTTCAGC  TTGAACTCCT
2461 CCTTTTTTGC  TTAGGAAGGT  TGACATATTT  TGTGTGCGAC  AGCGCTCATT  TTTCTACATA
2521 ATAGTGCGTT  TGGGTGGCGT  GGACGTTGGA  TGGCGTTTTC  GTTTCTACGG  ATTCACCCCT
2581 CCCACAAGCA  ATATACAGGG  CCCCTGTGTT  AAGGGGAAGC  CGACGATACT  TTACCCAGGA
2641 GGTGAAGGAG  ATAAGTTGAA  TTTACCGATT  TGTTTGGGGA  TGGGATAATC  CCAAACTCTG
2701 TGAATCGCTG  CGCGTCTTGA  TGCCTTGTTG  AAAGTTTGAG  GCTGGGGGAT  ATCC
                                                              EcoRV
```

FIGURE 3: Nucleotide and predicted amino acid sequence of the trypanosome BS2 gene. The region containing the trypanosome BS2 gene was completely sequenced from the genomic clone shown in Figure 1 (numbering from the *Bam*HI site). The putative sites for the splice acceptor (▼), start codon (***), stop codon (---), and poly(A) addition (▲) are marked as are the two putative redox-active sites (dots at the relevant cysteines; see text and subsequent figures). Potential N-linked glycosylation sites are underlined.

contiguous open-reading frame is evident. Assuming that the most upstream AUG is the start codon, the sequence predicts a primary translation product of ~55.5 kDa. Codons 3–14 encode hydrophobic residues, suggesting that this protein may possess a N-terminal signal peptide. There are 12 potential N-linked glycosylation sites (Figure 3; Asn-Xxx-Ser/Thr) with

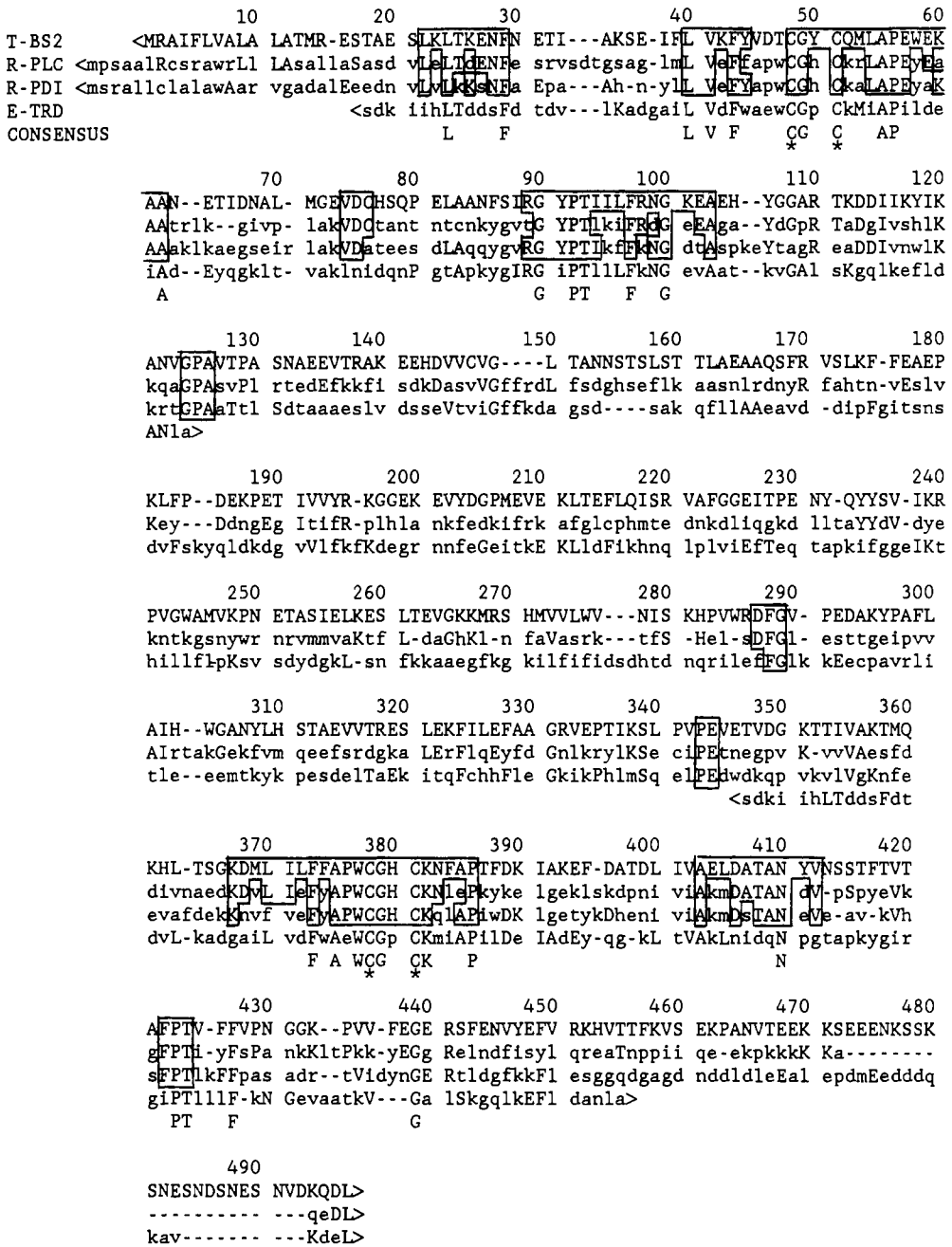

FIGURE 4: Homology between the complete sequences of trypanosome BS2, rat PDI (Edman et al., 1985), form I PIPLC (Bennett et al., 1988), and *E. coli* thioredoxin (Holmgren et al., 1975). The thioredoxin sequence is given twice, once below each of the two regions containing a thioredoxin-like active site. Identity to the BS2 sequence is indicated by upper case letters: lower case letters indicate nonidentity. Regions containing two or more adjacent residues common to BS2, PDI, and PIPLC are boxed. Residues common to all four proteins are indicated below each line of compilation. Asterisks indicate the cysteines in the two putative active sites.

a particularly notable cluster near the C-terminus of the predicted protein. These aspects will be further discussed below.

Homology to PDI and PIPLC. The protein sequence predicted for the BS2 gene was compared with the NBRF data base and found to have significant homology with the highly conserved but much shorter thioredoxin (about 108 residues) of bacteria [e.g., *E. coli* (Holmgren et al., 1975)] and eukaryotes [e.g., spinach chloroplast (Maeda et al., 1984)]. Greatest similarity, however, was observed with the sequence for rat PDI (Edman et al., 1985; Figure 4). This homology received a score of 511 using the FASTP algorithm, which is greater than 10 standard deviations from the mean score and therefore judged highly significant. Most recently, the sequence of the form I PIPLC of rat has been reported (Bennett et al., 1988). As this also has two thioredoxin-like domains, we compared it with BS2 using FASTP. As expected, high similarity was again observed (Figure 4) with a score of 534. The molecular weight of the primary translation product for all three is predicted to be ~55K–56K.

The homology between PDI, PIPLC, and BS2 extends over the entire length of all three proteins but is most obvious at and around the two (thioredoxin-like) active sites centered on residues 50 and 380 of BS2 as well as three other regions centered on residues 26, 96, and 409. In all these regions (and to a lesser extent throughout the three proteins), the nonidentical residues are generally conserved substitutions. The conserved regions are found at similar positions relative both to the protein termini and to each other (resulting in the need for a minimum of gaps to be introduced in obtaining optimum alignment). This strengthens the implication of evolutionary relatedness between all three proteins. The final region of

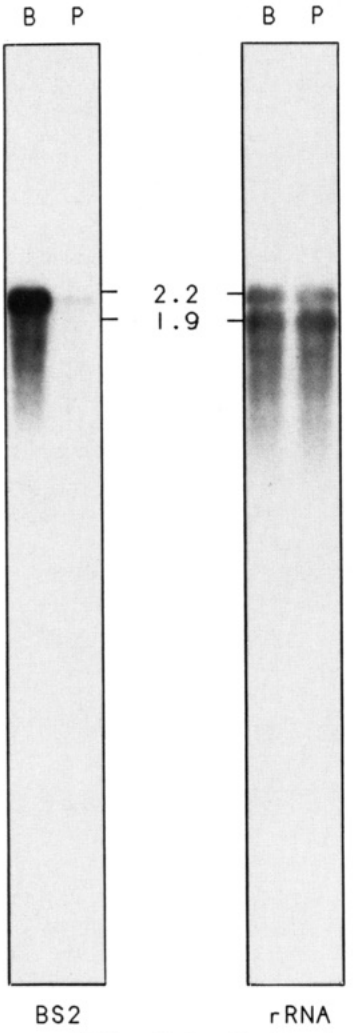

FIGURE 5: Stage specificity and size of mature transcript. (Left) Northern blot analysis of whole cell RNA (10 μg) from bloodstream (B) or procyclic (P) culture forms hybridized with a nick-translated *Pst*I/*Bgl*II fragment of the BS2 gene (from positions 1515–2014 in Figure 1). (Right) The filter used in the left panel was cleared of counts by boiling and rehybridized with a probe for two of the ribosomal RNAs [pTBR1 (Campbell et al., 1987)] as a control to show that similar amounts of RNA were indeed loaded. The latter rRNAs also serve as size markers (Campbell et al., 1987).

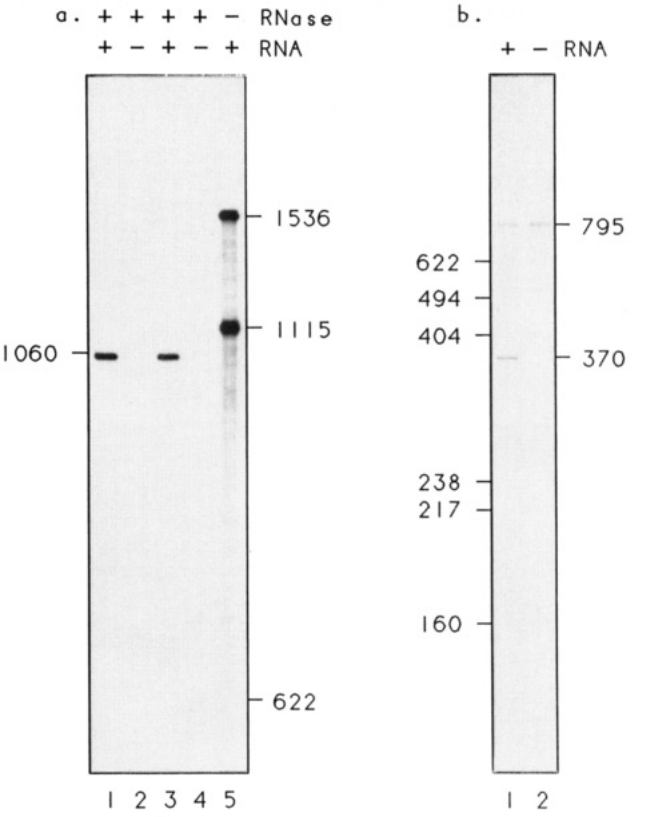

FIGURE 6: Mapping of the protein-coding exon. (a) Identification of the splice acceptor site. A uniformly labeled antisense RNA probe was generated by transcription off the T7 promoter of a "pGEM2" subclone comprising the *Bam*HI to *Pst*I fragment of the BS2 gene (positions 1–1515, lanes 1 and 2) or *Bgl*II to *Pst*I (positions 422–1515, lanes 3 and 4). This was hybridized to 13 μg of total bloodstream-form RNA (lanes 1 and 3) or no RNA (lanes 2 and 4) and digested with RNase T1, and the products were analyzed by polyacrylamide gel electrophoresis (5% acrylamide/7 M urea) and autoradiography. Lane 5 contains the two untreated probes. Size markers (in addition to the two probes) were pAT153 digested with *Msp*I (co-run but not shown). (b) Identification of the poly(A) addition site. As for part a, except transcription was off the T3 promoter of a bluescript subclone comprising the *Bgl*II to *Eco*RV sites (positions 2014–2750 which, including 59 nt of vector, yield an antisense RNA probe of 795 nt).

strong similarity is the C-terminus of each protein which in all three cases is comprised of predominantly charged residues.

Transcript Size and Developmental Regulation. Equal amounts of bloodstream-form and procyclic culture-form total RNA were analyzed by Northern blot analysis to determine the size and relative amounts of BS2 transcript in each. The results (Figure 5) show that the BS2 mRNA is about 2100 nucleotides (nt) long and that steady-state levels of the mature transcript are substantially greater in bloodstream than in procyclic culture forms of the organism (by densitometric analysis, 15-fold more in bloodstream RNA). To ensure that equal amounts of RNA were loaded, the blot was cleared by boiling and reprobed with an rRNA probe (Figure 5, right panel). Approximately equal signals were obtained for the RNA from each stage, indicating that this was indeed the case. [Note that, in other experiments, we have also analyzed this RNA for the levels of the tubulin transcripts; as reported by others (Van der Ploeg et al., 1985), tubulin mRNA is about 2–4-fold more abundant in bloodstream forms so that, even relative to this transcript, the BS2 mRNA is significantly more abundant in the bloodstream vs procyclic forms.]

Mapping the 5′ and 3′ Ends of the Protein-Coding Exon. In trypanosomes, all mRNAs so far examined are generated by a process known as trans-splicing [reviewed in Borst (1986) and Boothroyd (1989)]. Briefly, this refers to the synthesis of a mRNA through the splicing together of 2 discrete RNA molecules; one encodes the common 39-nucleotide miniexon sequence found at the 5′ end of most if not all trypanosome mRNAs while the other includes the complete sequence of the protein-coding exon for the gene in question. There are no known instances where a conventional "cis" intron has been found in any trypanosome gene. For a given mRNA, therefore, identification of the splice acceptor site at the 5′ end (i.e., the point of addition of the miniexon sequence) and the poly(A) addition site at the 3′ end should accurately define the complete protein-coding exon.

RNase protections were used to map this region for the BS2 gene (Figure 6). The 5′ end was mapped (Figure 6a) by using two antisense RNA probes spanning positions 1–1515 and 422–1515 (*Bam*HI to *Pst*I and *Bgl*II to *Pst*I, respectively; Figure 1). The protected product in both cases was about 1060 nt which indicates that the splice acceptor site lies near position 455 of the sequence presented in Figure 2. Even though trypanosomes apparently trans-splice their transcripts, their splice sites conform to the eukaryotic consensus sequence of a polypyrimidine stretch followed by AG (Borst, 1986). Only

one such site (position 470/471) is found in the region predicted by the RNase protection data shown in Figure 6, and this has been marked in Figure 3.

The 3′ end was mapped (Figure 6b) using antisense RNA spanning position 2014 (*Bgl*II) to 2750 (*Eco*RV). The protected species of about 370 nt localized the 3′ end of the protein-coding exon to about position 2380. If BS2 is polyadenylated, this should correspond to the poly(A) addition site (for which no consensus site has been identified in trypanosomes). These 5′ and 3′ ends of the protein-coding exon predict a transcript of somewhat more than 1950 nt [comprising 39 nt of miniexon sequence, 1910 nt of protein-coding exon, and the possibility of an unknown amount of poly(A) tail]. This is in approximate agreement with the size (2100 nt) of the transcript as determined by Northern blot analysis (Figure 5).

Discussion

A trypanosome gene with homology to two rat proteins containing thioredoxin-like domains has been identified and characterized. Steady-state levels of this transcript are substantially greater in bloodstream-form trypanosomes than in the procyclic culture forms which mimic the forms found in the insect vector.

The residues in and around the active site for each thioredoxin-like domain are almost perfectly conserved in all three proteins as are the adjacent residues. Other regions of homology further confirm the evolutionary relatedness of the three proteins. However, the disparate enzymatic activities of the two rat proteins preclude any deduction of the likely function of BS2 based on structural homology alone. Indeed, the PDI gene product apparently possesses several activities, most notably the β-subunit of prolyl 4-hydroxylase (Pihlajaniemi et al., 1987), cellular thyroid hormone binding (Yamauchi et al., 1987), glycosylation site binding protein (Geetha-Habib et al., 1988), and glutathione-insulin transhydrogenase (Morris & Varandani, 1988), further obscuring the likely function of the BS2 gene product in trypanosomes. Although other proteins containing active disulfide centers have been described including the trypanosomal analogue of glutathione reductase ("trypanothione reductase"; Shames et al., 1988), such proteins show no similarity to BS2 beyond the presence of two cysteine residues (at different spacing and with different adjacent residues) in the presumed active site.

Recently, the structure of a trypanosome PIPLC (the so-called GPIPLC with specificity for the glycolipid anchor of the surface antigens) has been fully determined (M. Carrington, R. Buelow, H. Reinke, P. Overath, and P. Englund, personal communications): no significant similarity between this later protein and BS2 is apparent, excluding this precise function for BS2. However, in mammals there are several PIPLC enzymes, each with different cofactor requirements and/or structure. Therefore, BS2 could encode a PIPLC with different specificity to the trypanosome GPIPLC described above. Clearly, isolation and biochemical characterization of the BS2 protein will be required before any definitive function can be ascribed.

As already noted, an additional region of similarity is readily apparent between BS2, PDI, and rat PIPLC at their respective C-termini. In all three, this region is comprised of predominantly charged (generally acidic) residues. Munro and Pelham (1987) have suggested for PDI that such C-termini and, in particular, the four residues Lys-Asp-Glu-Leu may be the signal for localization of this protein to the endoplasmic reticulum (rather than being transported to the surface and/or secreted). The corresponding segment of BS2 is Lys-Gln-Asp-Leu, which, allowing for the evolutionary distance separating mammals from trypanosomes, may indicate that BS2 is also localized to the endoplasmic reticulum. (The C-terminus of the rat form I PIPLC, Gln-Glu-Asp-Leu, has not been studied in this regard, and its precise subcellular localization is not known.)

The one striking difference between BS2 and the rat proteins is that BS2 has several possible N-linked glycosylation sites (Figure 3) whereas none are present in the two rat proteins. Enzymes with multiple N-linked glycosylations are rare, yet the thioredoxin-like domains clearly argue for enzymatic activity. If indeed glycosylated, this could affect not only the enzymatic function of the BS2 protein but also its localization.

The developmental regulation of the BS2 transcript argues for a similar modulation of its encoded protein. As the physiology of the two forms of the parasite differs enormously, many possible pathways exist which might demand more of the BS2 activity in bloodstream vs procyclic forms (e.g., glycolysis and generation/maintenance of a surface coat). It is important that the BS2 protein be purified and its activity and subcellular localization determined as it appears likely to be a key protein in the physiology and development of this parasite. Its predicted heavy glycosylation and possible localization to the lumen of the endoplasmic reticulum may have general implications for protein trafficking in eukaryotic cells.

Acknowledgments

We thank David Campbell (UCLA) for the cosmid clone containing BS2 and for characterization of cDNA 14 which led to its discovery, Dalia Perelman for finishing off the sequence, Thomas Beals for the genomic library, Maurine Hobbs for the procyclic culture-form RNA, Chris Walsh (Harvard) for a preprint of the trypanothione reductase work, R. Buelow (Stanford) and P. Englund (Johns Hopkins) for information on the GPIPLC sequence, and C. C. Wang (UCSF) for helpful comments on thioredoxins.

Registry No. PDI, 37318-49-3; PIPLC, 63551-76-8; DNA (*Trypanosoma brucei* gene BS2), 121210-43-3; protein (*Trypanosoma brucei* gene BS2 reduced), 121210-47-7; RNA (*Trypanosoma brucei* gene BS2 protein messenger), 121210-50-2.

References

Bennett, C. F., Balcarek, J. M., Varrichio, A., & Crooke, S. T. (1988) *Nature 334*, 268–270.

Boothroyd, J. C. (1989) in *Nucleic Acids and Molecular Biology* (Eckstein, F., & Lilley, D. M. J., Eds.) Springer-Verlag, Berlin (in press).

Boothroyd, J. C., & Beals, T. P. (1987) in *Host–Parasite Cellular and Molecular Interactions in Protozoal Infections* (Chang, K. P., & Snary, D., Eds.) pp 9–18, Springer-Verlag, Heidelberg.

Borst, P. (1986) *Annu. Rev. Biochem. 55*, 701–732.

Brun, R., & Schonenberger, M. (1981) *Z. Parasitenkd. 66*, 17–24.

Campbell, D. A., Kubo, K., Clark, G., & Boothroyd, J. C. (1987) *J. Mol. Biol. 196*, 113–124.

Cross, G. A. M. (1975) *Parasitology 71*, 393–417.

Edman, J. C., Ellis, L., Blacher, R. W., Roth, R. A., & Rutter, W. J. (1985) *Nature 317*, 267–270.

Geetha-Habib, M., Noiva, R., Kaplan, H. A., & Lennarz, W. J. (1988) *Cell 54*, 1053–1060.

Holmgren, A. (1985) *Annu. Rev. Biochem. 54*, 237–271.

Holmgren, A., Soderberg, B. O., Ecklund, H., & Branden, C. I. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 2305–2309.

Jasmer, D. P., Feagin, J. E., & Stuart, K. (1985) *Mol. Cell. Biol. 5*, 3041–3047.

Luthman, M., & Holmgren, A. (1982) *Biochemistry 21*, 6628-6633.
Maeda, K., Tsugita, A., Dalzoppo, D., Vilbois, F., & Schurmann, P. (1984) *Eur. J. Biochem. 154*, 197-203.
Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499-560.
Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K., & Green, M. R. (1984) *Nucleic Acids Res. 12*, 7035-7056.
Morris, J. I., & Varandani, P. T. (1988) *Biochim. Biophys. Acta 949*, 169-180.
Muhich, M., & Boothroyd, J. C. (1988) *Mol. Cell. Biol. 8*, 3837-3846.
Muhich, M. L., Simpson, L., & Simpson, A. M. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4060-4064.
Munro, S., & Pelham, H. R. B. (1987) *Cell 48*, 899-907.
Opperdoes, F. R. (1987) *Annu. Rev. Microbiol. 41*, 127-151.
Osinga, K. A., Swinkels, B. W., Gibson, W. C., Borst, P., Veeneman, G. H., VanBoom, J. H., Michels, P. A. M., & Opperdoes, F. R. (1985) *EMBO J. 4*, 3811-3817.
Overath, P., Czichos, J., Stock, U., & Nonnengaesser, C. (1983) *EMBO J. 2*, 1721-1728.
Pihlajaniemi, T., Helaakoski, T., Tasanen, K., Myllyla, R., Huhtala, M. L., Koivu, J., & Kivirikko, K. I. (1987) *EMBO J. 6*, 643-649.
Rigby, P. W. J., Dickman, M., Rhodes, C., & Berg, P. (1977) *J. Mol. Biol. 113*, 237-251.
Sanger, F., & Coulson, A. R. (1978) *FEBS Lett. 87*, 107-110.
Sanger, F., Nickeln, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463-5467.
Shames, S. L., Kimmel, B. E., Peoples, O. P., Agabian, N., & Walsh, C. T. (1988) *Biochemistry 27*, 5014-5019.
Van der Ploeg, L. H. T., Giannini, S. H., & Cantor, C. R. (1985) *Science 228*, 1443-1446.
Vickerman, K. (1962) *Trans. R. Soc. Trop. Med. Hyg. 56*, 487-495.
Vickerman, K. (1965) *Nature 208*, 762-766.
Vickerman, K. (1985) *Br. Med. Bull. 41*, 105-114.
Yamauchi, K., Yamamoto, T., Hayashi, H., Koya, S., Takikawa, H., Toyoshima, K., & Horiuchi, R. (1987) *Biochem. Biophys. Res. Commun. 146*, 1485-1492.
Yanisch-Perron, C., Vieira, J., & Messing, J. (1985) *Gene 33*, 103-119.

Structure of Synthetic Unmethylated 16S Ribosomal RNA as Purified RNA and in Reconstituted 30S Ribosomal Subunits†

Gwen Ericson, Kairev Chevli, and Paul Wollenzien*

E. A. Doisy Department of Biochemistry and Molecular Biology, St. Louis University Medical Center, St. Louis, Missouri 63104

Received January 4, 1989; Revised Manuscript Received April 17, 1989

ABSTRACT: 16S ribosomal RNA was made by in vitro transcription of a cloned gene, and its structure was compared to authentic 16S ribosomal RNA. The comparison was made by subjecting the two types of 16S rRNA to chemical reagents that react specifically with unpaired bases and determining the extent of reaction by reverse transcription and gel electrophoresis of the cDNA. In solution, the rRNAs were indistinguishable in their pattern of reactivity, except for a difference at A1408 and many differences in the region between residues 470 and 562. When the 16S rRNAs were reconstituted into 30S ribosomal subunits, these reactivity differences were absent. When the synthetic 16S rRNA was reconstituted into 30S subunits and then extracted and tested in solution, its pattern of chemical reactivity in the 470-562 region was the same as authentic 16S rRNA, but differences in the 1408-1410 region persisted. This study indicates that the synthetic 16S rRNA has a secondary structure in solution similar to a native secondary structure except in two regions, one apparently incorrectly folded during synthesis and the other in which nucleotides which are normally methylated in authentic 16S rRNA may be responsible for a structural difference.

Purified 16S rRNA in solution has been shown to have a secondary structure consistent with the structure predicted from comparative sequence analysis (Noller & Woese, 1981; Van Stolk & Noller, 1984; Moazed et al., 1986; Baudin et al., 1987). This indicates that the base-pairing arrangement necessary for its biological function is stable without the presence of the ribosomal proteins, except in several local regions. However, since the source for 16S rRNA is the ribosome or ribosomal subunits, 16S rRNA may have a specific conformation because of its prior assembly into ribosomal particles. It is known that subjecting the 16S rRNA to organic solvents or heating under conditions of low ionic conditions renders the molecule difficult to reconstitute into 30S particles (Barritault et al., 1979). This raises the possibility that the native structure of the RNA molecule is quasi-stable and, once disrupted, not easily regained during in vitro reconstitution.

A second question pertaining to the behavior of the 16S rRNA concerns the role of posttranscriptional methylations in determining its structure and functional activity. In the *Escherichia coli* 16S rRNA, there are 10 posttranscriptional methylations (Noller & Woese, 1981; Noller, 1984); these are representative of the modifications that occur in eubacterial

† This research was supported by U.S. Public Health Service Grant GM35410 and by a Biomedical Research Support grant from St. Louis University School of Medicine.

* Address correspondence to this author at the Department of Biochemistry, St. Louis University Medical Center, 1402 S. Grand Blvd., St. Louis, MO 63104.

0006-2960/89/0428-6446$01.50/0