Expression of antigenic regions of a trypanosome variable surface glycoprotein in *E. coli* using Bal-31 nuclease digestion

Michael J. Lenardo*⁺, Steven T. Brentano* and John E. Donelson*

*Department of Biochemistry, and ⁺Division of Hematology, Department of Medicine, School of Medicine, University of Iowa, Iowa City, IA 52240, USA

Received 17 February 1984; Revised and Accepted 21 May 1984

ABSTRACT

We have used the expression of a trypanosome variable surface glycoprotein (VSG) in *E. coli* to produce VSG serotype-specific antisera which have none of the cross-reacting specificities characteristic of antisera prepared against purified VSGs. This was accomplished by treating restriction fragments of VSG cDNAs with Bal-31 nuclease to facilitate expression of their open reading frames in the *E. coli* expression vector, pMR100 (4). The resultant VSG- β -galactosidase fusion proteins possess various antigenic regions of the original VSG. This provides a rapid means for producing VSG-specific antisera for reagent use and has the capability of large scale production of antigen for immunological investigation.

INTRODUCTION

The major antigen on the surface of African trypanosomes, the causal agents of African sleeping sickness, is the variable surface glycoprotein (VSG) (1). On any given organism, it is serologically homogeneous and constitutes about 5% of the total protein of the cell (2). By sequentially changing the VSG in their surface coat African trypanosomes avoid destruction by their host's immune system (2).

We sought efficient expression of trypanosome VSG cDNA sequences in *E. coli* for several reasons. First, *E. coli* can be used to generate large quantities of homogeneous VSG proteins when large scale isolations may not be possible by other techniques. For example, trypanosomes at the metacyclic stage are obtained by laborious dissection of salivary glands from the tsetse fly (3). The small number of organisms obtained in this manner are a mixture of serotypes expressing different VSGs (3). This is significant because immunological studies show that metacyclic organisms - the stage at which trypanosomes are inoculated by the tsetse fly into the vertebrate host - express only a restricted set of VSG antigenic types (7, 19). For some trypanosome species this restricted set is stable enough to raise the prospect of a "cocktail" vaccine against

the disease (19). Expression of these types in *E. coli* would facilitate the immunological investigation of this possibility. Second, the isolation of VSG genes or cDNAs is relatively simple since they can be identified by characteristic C-terminal homology sequences present in all VSG genes sequenced to date (10). However, if the cDNA library is constructed from a mixture of trypanosome serotypes, the correlation of a specific serotype with a given cDNA sequence is much more difficult. This would be simplified if serotype-specific antisera could be generated for this purpose using *E. coli* fusion proteins bearing VSG segments. Third, specific antigenic regions can be selectively expressed by the method described below allowing study of the antigenic structure of the VSG molecule using recombinant DNA techniques.

Several laboratories have constructed plasmids that facilitate the expression of eukaryotic proteins as fusions with *E. coli* β -galactosidase (4-6, 14). We report here the use of the plasmid pMR100, developed by Gray et al. (4), to obtain efficient expression of trypanosome VSG cDNAs in *E. coli*. This was accomplished by digesting cDNA restriction fragments with Bal-31 nuclease to randomize their ends and remove other hindrances to their expression. We show that fusion proteins from clones generated by this technique bear antigenic regions of a VSG. Moreover, small regions of the VSG coding sequence, when expressed as part of a fusion protein, are shown to elicit VSG serotype-specific antisera. These antisera do not contain cross-reacting specificities directed against carbohydrate moieties in the C-terminal amino acid homology region which is characteristic of antisera raised against purified VSGs.

MATERIALS AND METHODS

Bal-31 digestion of trypanosome cDNAs and construction of recombinant expression segments.

The full-length VSG cDNA (42) used in these studies was isolated from a *Trypanosoma b. brucei* IaTat 1.2 cDNA library by differential hybridization and its complete sequence was determined by Murphy et al. (8,41). *Trypanosoma b. rhodesiense* cDNAs were isolated in similar fashion and either partially or completely sequenced (40). Restriction fragments from these cDNAs were separated on preparative 6% polyacrylamide gels and electroeluted (9). Bal-31 nuclease (Bethesda Research Labs) digestions were carried out in buffer described by Legerski (11). A time course reaction was carried out at 37°C with 2 μ g of DNA and 0.71 unit of

Bal-31 nuclease in a volume of 20 μ l. At various intervals, 5 μ l aliquots were transferred to tubes containing 4.5 μ l of 0.2 mM EGTA and electrophoresed on 10% acylamide gels to estimate optimal digestion times. The digest was repeated using 2 μ g of DNA for the optimal time followed by phenol extraction and ethanol precipitation of the DNA. Ends of the nuclease-digested DNA were repaired with Klenow fragment of DNA Polymerase I (New England Biolabs). The plasmid pMR100 (a gift of Dr. M. Rosbash) was digested with Sma I (New England Biolabs) in conditions recommended by the supplier. The reaction mixture was brought to pH 8.0 by addition of an equal volume of 1 M Tris-Cl pH 8.0 and incubated at 65°C with 5 units/ μ g DNA of bacterial alkaline phosphatase (Bethesda Research Labs) for 2 hours followed by phenol extraction and ethanol precipitation. Ligations were performed under standard conditions (12) using T4 DNA ligase (New England Biolabs). Transformations were done according to Mandel and Higa (13) using E. coli strain LG90 (14) (a gift of Dr. M. Rosbash) which was made competent by freezing in 10 mM CaCl₂ and 14% glycerol (15). Transformants were selected on MacConkey agar (Difco) with 50 μ g/ml ampicillin (Omnipen, Wyeth) added.

Southern blotting and nucleotide sequencing

DNAs from small scale plasmid isolations (16) were digested with Bam HI (New England Biolabs) which excises the inserts and a few base pairs of flanking sequence from the vector (see Figure 1), and electrophoresed on 1.5% agarose (Seakem) gels. Bidirectional transfer (17) of the DNA to nitrocellulose filters (Schleicher and Schuell) was followed by hybridization of the blots at 65°C in 4 X SET (0.6 M NaCl, 0.12 M Tris-Cl, pH 7.5, and 8.0 mM EDTA) for 12 hrs. The appropriate cDNA probes were radioactively labelled with [α^{32} P]-dATP (ICN Pharmaceuticals, Inc.) by nick translation (18). Sequencing was carried out by the method of Maxam and Gilbert (20).

Antibody preparation

Antisera against purified IaTat 1.2 VSG was raised in female New Zealand white rabbits by injecting 100 μ g of protein (gift of Dr. George Cain) in complete Freund's adjuvant subcutaneously, followed in 2 weeks with a boost of an equal dose in incomplete Freund's adjuvant. Serum was harvested one week after the second injection. Ouchterlony immunodiffusion gels demonstrated a homologous reaction between the antisera and the IaTat 1.2 VSG. The serum was also shown to specifically immunoprecipitate iodinated IaTat 1.2 VSG. Antisera against the E. coli fusion

proteins were prepared in a similar fashion. Because of the large sizes of the fusion proteins (>120,000 daltons), they could be separated electrophoretically from nearly all *E. coli* proteins (see Figure 3). Slices from 7.5% preparative SDS-acrylamide gels (21) containing between 100 and 200 µg of fusion protein were macerated in phosphate-buffered saline using a 7.0 ml Dounce homogenizer and used for immunizations with Freund's complete adjuvant.

Western blotting

Overnight cultures of Lac⁺ transformants of *E. coli* strain LG90 were grown in 5 mls of LB media with 50 µg/ml of ampicillin at 37°C with shaking. These were lysed in sample buffer and electrophoresed on 7.5% SDS-acrylamide gels (21). Western blotting was performed according to Towbin et al. (22) using a 1:100 dilution of anti-IaTat 1.2 VSG antisera and probing with a 1:133 dilution of ¹²⁵I-Staph A protein (New England Nuclear) that possessed a specific activity of 2-10 uCi/ug and a concentration of 100 uCi/ml. Overlay solutions of antibody and probe were preincubated with LG90 lysate for 2 hours at 4°C.

Blots against VSG proteins were carried out in the same manner using 2 µg of purified IaTat 1.2 VSG, IaTat 1.10 VSG (gifts from Dr. George Cain), or *E. coli* β-galactosidase (Sigma) per lane. Dilutions of antisera were: anti-IaTat 1.2 - 1:200; anti-B3.55 - 1:50; anti-B3.56 - 1:50 and preimmune serum - 1:100. ¹²⁵I-staph A probe was diluted 1:200. Autoradiography was performed on all blots using Kodak XAR film with a Cronex intensifying screen (Dupont) at -70°C.

Scanning densitometry - *E. coli* lysates electrophoresed on 7.5% SDS-acrylamide gels and stained with Coomassie Brilliant Blue or silver stain were scanned on a RFT II scanning densitometer, Transidyne General Corp.

Computer analysis - Analysis of nucleotide sequence data was performed on an IBM personal computer using the PCS software package described by Lagrimini et al. (23).

RESULTS

Use of Bal-31 nuclease to produce open reading frame fragments.

The general strategy for cloning antigenic coding regions of trypanosome VSG cDNAs into the vector pMR100 (4) is shown in Figure 1 for the case of the IaTat 1.2 cDNA. This cDNA contains the coding sequence for the entire nascent VSG (496 amino acids) including the amino terminal signal peptide (31 amino acids) and the C-terminal hydrophobic peptide (17 amino acids) which are not present in mature VSG (24). The initiation

FIGURE 1. The cloning strategy. The 1063 base pair fragment to be sub-cloned lies between the Pst I site at one end of the cloned IaTat 1.2 cDNA and the first internal Pst I site. It was treated with Bal-31 nuclease and cloned as described in text. P = Pst I, H³ = Hind III, B = Bam HI, S = Sma I, Lac P, O = a tandem repeat of the promoter and operator from the Lac operon. Lac Z' = a deleted version of the lac Z gene. For clarity, the diagram of pMR100 does not show the λ CI gene segment which is between the lac promoters and cloning site and contains the start codon of the fusion protein or the lac I' gene segment between the cloning site and lac Z' gene (see. ref. 4, figure 1). Vertical hatching = dG:dC tail, diagonal hatching = protein coding sequence, open boxes = 5' and 3' untranslated regions, and solid boxes = pBR322 sequence.

codon and the orientation of the open reading frame in the cDNA are shown along with the locations of two internal Pst I sites in the cDNA. The Pst I-Pst I fragment bearing the initiation codon is 1063 base pairs long and contains 33 base pairs of dG:dC tail, 56 base pairs of 5' nontranslated region, and 974 base pairs of VSG coding sequence (8). The approximately 324 codons that it contains specify the N-terminal three-fifths of the VSG sequence and none of the C-terminal homology region (10). Since antigenic variability is believed to involve the N-terminus of the VSG molecule (2), this restriction fragment was used as substrate for the expression experiments.

Because the restriction fragment contained an in-phase termination codon immediately preceding the initiation codon and its Pst I termini were incompatible with suitable cloning sites in pMR100, the fragment was digested with Bal-31 nuclease. An average shortening of 350 base pairs was considered optimal to facilitate expression of the open reading frame in the fragment since the in-phase termination codon was 86 base pairs from the end of the fragment, and the dG:dC tail at this end would be degraded several times slower than the other end of the fragment (31). A time course demonstrated that the time required to remove about 350 base pairs was 6 minutes (data not shown). To extend the analysis, Bal-31 nuclease was also used to prepare restriction fragments for cloning into pMR100 of three *T. b. rhodesiense* cDNAs for which only partial DNA sequence data was available. In these cases, Bal-31 nuclease removed several hindrances to the expression of the open reading frame each contained, such as incompatible restriction termini, dG:dC tails and in-phase termination codons. It also randomized both ends of these fragments such that the reading frame of a fraction (1/18 or 5.6%) of the Bal-31 digested molecules would match the reading frame in the vector when blunt-end ligated into the Sma I site of pMR100. Following Bal-31 digestion, fragments were treated with the Klenow fragment of *E. coli* DNA Polymerase I in the presence of dXTPs to fill in 3' recessed ends and remove 3' protruding ends.

Using this strategy, the Pst I-Pst I fragment of the IaTat 1.2 cDNA and the three *T. b. rhodesiense* cDNAs were cloned into the Sma I site of pMR100 and scored for lac⁺ recombinants. Table I shows that in all four cases cloning of the Bal-31 nuclease digested fragment produced a little more than one-half the expected number of lac⁺ clones (5.6%). Also evident is the fact that sufficiently large numbers of lac⁺ colonies can be obtained after Bal-31 nuclease treatment if 10³ or more transformants are obtained despite the expected low efficiency of blunt-end ligation. Control experiments in which the phosphatase treated vector was ligated to itself without an insert revealed many fewer transformants and only rare lac⁺ colonies (<0.1%) (not shown).

Analysis of recombinant molecules produced from Bal-31 digestion of the IaTat 1.2 Pst-Pst fragment

Since the complete sequence of the full-length IaTat 1.2 VSG cDNA is known (41), we were able to identify the structures of the IaTat 1.2 cDNA recombinants as well as investigate the immunological characteristics of

TABLE I. β -Galactosidase phenotype in the expression of several trypanosome VSG cDNAs.

Subspecies	Serotype ¹	Restriction fragment size (bp)	lac ⁺	Total transformants	%
T. b. brucei	IaTat 1.2	1063	56	1949	2.9
T. b. rhodesiense	WraTat 1.27	622	16	516	3.1
T. b. rhodesiense	(6R25)	1505	51	1315	3.9
T. b. rhodesiense	(9L9)	1700	53	1277	4.3

¹In cases where the VSG cDNA has not been assigned a serotype it is identified by a lab reference number in parentheses. All VSG cDNAs are from this laboratory (8,40).

their fusion proteins. We chose five lac⁺ colonies from the IaTat 1.2 cDNA transformants for further study. Southern blot analysis revealed that one, A.47, did not contain an insert that hybridized with the IaTat 1.2 cDNA probe. The other four contained inserts ranging from 90 to 600 base pairs. The four recombinant plasmids bearing inserts detected by the Southern blot were subjected to DNA sequencing to determine the 5' boundaries of the open reading frame insertions. This data is summarized in Figure 2 (top). In the upper left corner is shown the in-phase termination codon preceding the initiator methionine which must be removed by the Bal-31 digestion to obtain expression of the Bal-31 treated fragment. The inserts form a nested series of end-deleted fragments derived from the IaTat 1.2 Pst I-Pst I fragment (P1063). In all cases the reading frame begins after the in-phase stop codon. Furthermore, the positions of the deleted fragments on the map of IaTat 1.2 P1063 is biased towards the end bearing the G:C tail. This reflects the increased resistance of the G:C tail to Bal-31 nuclease digestion. Figure 2 (bottom) shows the deduced amino acid sequences at the upstream boundary of each insert. In each case, the 5' boundary the VSG reading frame is in phase with that of vector. In the two cases where it was possible to determine the complete sequence of the insert on the sequencing gel (A.55 and A.57), the 3' end of the sense strand of the insert also forms a continuous reading frame with the vector. The approximate 3' termini of other fragments are inferred from the size of the insert on the Southern blot. This analysis also illustrates that non-overlapping regions of the protein coding sequence (for example, A.57 vs. A.49) can be readily generated by the process of random degradation.

FIGURE 2. Four examples of the nested series of fragments produced by Bal-31 nuclease digestion. The upper part shows the in-phase termination codon which precedes the initiator methionine. P = *Pst* I. P1063 denotes the 1063 base pair *Pst* I-*Pst* I fragment of the IaTat 1.2 cDNA. Hatched box shows dG:dC tail. The arrows show the regions subcloned into pMR100. Under the endpoints of the arrows are listed the codon numbers in the sequence of the cDNA. Single arrowheads denote a precise endpoint was determined, double arrowheads indicate endpoints inferred from the size of the insert and position of the beginning of the fragment. The bottom portion shows the amino acid sequence at the boundaries of the insert as determined by DNA sequencing. The single letter amino acid code is used. λ cl and lac I'-Z' show regions of the pMR100 fusion which are present on each side of the cloning site (see ref. 4, figure 1). Solid dots denote amino acids specified by a hybrid codon.

Analysis of fusion proteins and detection of antigenic sites of IaTat 1.2 in fusion proteins.

Whole cell lysates were run on 7.5% SDS-acrylamide gels to investigate the protein products of these clones. An example is shown in Figure 3. Lane 1 shows the protein encoded on plasmid pMR200, a rare revertant in which the frameshift in pMR100 has been eliminated (4). The four plasmids bearing inserts direct the synthesis of fusion proteins whose sizes directly correlate with the size of the cDNA insert: A.49, ~145,000 daltons (lane 3); A.55, ~141,000 daltons (lane 4); A.56, ~153,000 daltons (lane 5); and A.57, ~133,000 daltons (lane 6). As expected, the protein in strain A.47 (lane 2) which does not appear to have a cDNA insert comigrates with the revertant pMR200 protein. This protein may be the result (as was pMR200) of a rare spontaneous reversion of the frameshift

FIGURE 3. SDS-polyacrylamide gel of lysates of bacterial clones containing the fusion proteins. The gel is stained with Coomassie blue. The bacterial clones contained the following plasmids: 1 = pMR200; 2 = A.47; 3 = A.49; 4 = A.55; 5 = A.56; 6 = A.57. Upper arrow indicates position of the β subunit of *E. coli* RNA polymerase ($M_r = 155,000 \pm 15,000$). Lower arrow indicates position of pMR200 protein ($M_r = 130,000$).

mutation. More likely, it is the product of an exonuclease contamination of the *Sma* I enzyme or T4 DNA ligase which removed a single base pair and re-established the continuity of vector's reading frame (4). Similar analyses of lac^+ clones containing the *T. b. rhodesiensis* cDNAs (see Table I) showed the sizes of the fusion proteins correlated directly with the lengths of the Bal-31 treated cDNA insert (data not shown).

To determine whether the fusion proteins contain antigenic regions of the IaTat 1.2 VSG, duplicate gels were transferred to nitrocellulose and probed with antisera prepared against VSG protein purified from IaTat 1.2 organisms. The results in Figure 4 show that all recombinant plasmids which contain the IaTat 1.2 cDNA sequence also generate fusion proteins which bear antigenic determinants of the IaTat 1.2 VSG. The magnitude of the reactivity with the antibody appears to be proportional to the size of the insertion. Also notable is the fact that since the inserts of clones A.49 and A.57 (see Figure 2) do not overlap, different antigenic regions of the VSG protein encoded by the cDNA are represented among the clones examined. In all cases, smaller immunoreactive proteins are also observed. These are most likely due to degradation of the fusion proteins. Support for this conclusion comes from Western blots of samples which have been stored for several weeks with several freezings and thawings.

FIGURE 4. Autoradiogram of a Western blot of fusion proteins using antiserum directed against IaTat 1.2 VSG. The gel is identical to that shown in figure 3; arrows are also the same. Control experiments with pre-immune serum revealed no antibody reactivity on equivalent strips from the same gel.

These reveal a greater abundance of the smaller bands and less of the full-length fusion protein (data not shown).

Generation of VSG serotype-specific antisera using *E. coli* fusion proteins.

The value of *E. coli* generated VSG protein in the production of antiserum to identify the serotype of a given cDNA or in the long-range goal of generating immunity to organisms of a given VSG type depends on the ability of VSG- β -galactosidase fusion proteins to act as immunogens and stimulate VSG specific antibody responses. To investigate this possibility we immunized rabbits with two of the IaTat 1.2 fusion proteins, A.55 and A.56 (derived from clones A.55 and A.56). These fusion proteins contain approximately 92 and 199 amino acids of the IaTat 1.2 VSG coding sequence, respectively. Using these antisera and antisera against the purified IaTat 1.2 VSG protein we probed Western blots of IaTat 1.2 and IaTat 1.10 VSG proteins as well as *E. coli* β -galactosidase. The second panel in Figure 5 (labelled anti-1.2) shows that antiserum prepared against purified IaTat 1.2 VSG cross-reacts strongly with the IaTat 1.10 VSG, despite the fact that corresponding coding sequences have

FIGURE 5. Autoradiogram of Western blots using preimmune serum (control) or antiserum against either purified IaTat 1.2 VSG (anti-1.2), fusion protein A.55 (anti-55) or fusion protein A.56 (anti-56). Lanes marked 2 contain 2 μ g of purified IaTat 1.2 VSG; lanes marked β contain 2 μ g of *E. coli* β -galactosidase; lanes marked 10 contain 2 μ g of purified IaTat 1.10 VSG. Upper and lower arrows indicate the migration of β -galactosidase (Mr ~120,000) and VSG (Mr ~60,000), respectively on stained strips.

little or no homology (S. Brentano, unpublished observations). This cross-reactivity is often found in antisera prepared against purified VSG proteins. Evidence suggests that the cross-reacting determinant is a common oligosaccharide moiety at the C-termini which is not exposed when the VSG is in place on the surface of the organism (28). This should not affect the immunologic specificity of anti-VSG antisera generated by the fusion proteins since normal glycosylation of eukaryotic glycoproteins does not occur in *E. coli* (35). As shown in the third and fourth panels of Figure 5, antisera against the fusion proteins A.55 (anti-55) and A.56 (anti-56) react with purified IaTat 1.2 protein and with β -galactosidase but not with the IaTat 1.10 protein. Several characteristics of these antisera are noteworthy. The antisera appear more reactive with the IaTat 1.2 VSG than with β -galactosidase even though the fusion protein contains about 1000 amino acids of β -galactosidase and 100 or 200 amino acids, respectively, of VSG. Also these anti-sera are only 4-8 fold less in titre than antisera generated against purified VSG protein (see Methods for dilutions). Finally, it is of interest that the anti-56 serum appears more avid than the anti-55 serum. This may simply indicate differences

in the immune response of the two animals immunized but may also reflect the fact that the A.56 fusion protein contains about twice the amount of VSG coding sequence as the A.55 fusion protein. Control experiments using antisera raised against *E. coli* β -galactosidase showed no cross-reactivity against the IaTat 1.2 or IaTat 1.10 VSGs.

Efficiency of expression of trypanosome VSG cDNAs in *E. coli*.

In the experiments described here, the steady-state accumulation of the fusion protein in clone A.56 is representative of the clones analyzed (see Fig. 3A). Since it is close in size to the β subunit of *E. coli* RNA polymerase, the relative abundance of the two proteins was compared by their intensity on gels stained with Coomassie Blue and gels stained with silver stain. Based on densitometer tracings and the known abundance of the β subunit of RNA polymerase (25), we estimate this fusion protein constitutes roughly 1-2% of the total soluble protein of the host bacterium (data not shown). This is approximately 6,000-12,000 monomers/cell, about the same order of magnitude as the level of expression of other proteins, such as β -globin, using a lac promoter (14, 26, 27).

DISCUSSION

We describe here the use of recombinants of the *E. coli* expression vector pMR100 to generate bacteria in which 1-2% of the total soluble protein is a fusion product between β -galactosidase and trypanosome VSGs. In preliminary studies we used the expression vectors of Talmadge and Gilbert (32) which have an insertion site just downstream from the β -lactamase promoter of plasmid pBR322. In these cases, we could demonstrate the expression of full-length VSG cDNAs in *E. coli* but the amount of the VSG produced was extremely low. Two other laboratories have reported fortuitous read-through expression from this same β -lactamase promoter of VSG cDNAs cloned into the *Pst* I site of pBR322; in these cases the amount of synthesized protein was also very low and was only detected by sensitive immunological assays (29, 30). Using the more efficient lac UV5 promoters on pMR100, larger amounts of fusion protein, suitable for raising antisera, could be readily obtained.

A level of 1-2% of the soluble cellular protein is substantially less than theoretically possible from the lac UV5 promoter (14, 26). As has been pointed out by Guarente et.al. (14), the levels of expression obtained in a typical construction using the lac promoter to express a eukaryotic protein is roughly an order of magnitude lower than would be predicted from the known promoter efficiency and the copy number of Col

El-derived plasmids in *E. coli*. In our case several factors may contribute to lowering the maximal level of expression. One factor is that the efficiency of the tandem lac promoters on pMR100 may be somewhat reduced because upstream translation initiation near the lac promoter interferes with initiation at CI which forms the N-terminus of the fusion protein (26). Another factor may be the distinctly biased use of specific codons in the VSG genes compared to that of *E. coli* genes (30, 33). Since VSGs constitute one of the most abundant proteins of the trypanosome (3), their codon preference may be biased towards efficient translation in that organism. A third factor is proteolysis of the fusion proteins which might be lessened by the use of bacterial hosts with mutations affecting proteolytic degradation pathways (34). Still another possible cause of diminished expression is potential toxicity of the VSG- β -galactosidase to *E. coli* but we observed no difference in growth rates between the bacteria containing pMR200 and those containing the recombinant plasmids. Other factors such as reduced stability of the chimeric mRNAs or poor solubilization of the fusion protein may also lead to decreased apparent expression of the fusion protein.

Open reading frame expression vectors such as pMR100 provide an efficient method for the immunological investigation of cDNA coding sequence. However, when cloning a variety of different cDNAs, many obstacles to expression are encountered such as restriction fragment ends dissimilar to the vector cloning site, in-phase stop codons, oligo (dG):(dC) tails and proper alignment of the desired open reading frame. We have exploited limited Bal-31 nuclease digestion to overcome these problems and found that it significantly enhances the speed and convenience of using these expression vectors. Its advantages include general applicability, ease of use, little or no dependence on DNA sequence information, and the ability to remove a variety of hindrances to the expression of open reading frames. It eliminates the need to design vectors with large numbers of restriction sites in different reading frames and is best suited for cases in which specific endpoints of the insert are not crucial. For small restriction fragments (<1 kilobase) it may be superior to previously described methods of preparing open reading frame DNA by sonication (4,5) or DNase treatment (5). While this work was being completed we noted that others have also successfully used Bal-31 nuclease to achieve randomization of reading frames in creating fusions between β -galactosidase and genes of *Saccharomyces cerevisiae* when they

are contained in the same plasmid (36-38).

The fusion proteins bearing only 100-200 amino acids of a VSG sequence on an ~1000 amino acid carrier protein generated an effective antibody response to VSG determinants. In principle, this is similar to generating antisera using short synthetic peptides (39). We were surprised to find the antisera generated by the fusion proteins reacted more strongly with the IaTat 1.2 VSG than with *E. coli* β -galactosidase, despite the fact that only 1/6 - 1/10 of the amino acid residues in the fusion proteins are derived from the inserted VSG coding sequences. This suggests that VSG sequences may be more immunogenic than *E. coli* β -galactosidase, which could reflect the fact that the VSG has evolved as the major antigenic structure on the surface of the trypanosome. It is notable also that the reactivity of anti-IaTat 1.2 VSG antisera with the fusion proteins appeared proportional to the length of IaTat 1.2 coding sequence in the fusion protein. This may indicate either that larger proteins encoded by these inserts have a greater number of antigenic determinants or that they adopt a conformation which interacts more favorably with the antiserum.

The expression of antigenic regions of VSG proteins in *E. coli* provides an approach to three areas of investigation in the molecular biology of trypanosomes. First, it will be useful in correlating cDNAs with trypanosome serotype. This can be achieved by using the fusion protein to generate a VSG serotype-specific antisera as is shown in this report. Alternatively, polyvalent or monoclonal antibodies of known serotype could be used to identify the homologous fusion protein. Recently, we have made such a serotype identification of a VSG- β -galactosidase fusion protein generated from a *T. rhodesiense* cDNA using monoclonal antibodies against metacyclic organisms (K.M. Esser and M. Lenardo, unpublished results). Second, the clone A.57, which encodes only 29 amino acids of the VSG, emphasizes the usefulness of the approach for dissecting a large antigenic structure into smaller components. This combined with the fact that non-overlapping segments are generated will allow small antigenic regions of VSG molecules to be selectively expressed. For example, it will be possible to test whether biologically important antigenic determinants in different VSGs occur in approximately the same location within the polypeptide chain. This would suggest antigenically unique VSGs share a common structure from which emerge distinctive antigenic features. Finally, as more becomes known about the antigenic types present at the metacyclic stage of development, we hope

to gain information on the usefulness of *E. coli* protein fusion products in establishing protective immunity against African trypanosomiasis.

ACKNOWLEDGMENTS

This work was supported by grant number DAMD 17-82-C-2228 and NIH grant AI 18954. We would like to thank Klaus Esser for helpful discussions, Michael Rosbash for supplying the vector-host system and both Michael Rosbash and Leonard Guarente for restriction map data. M.J.L. would like to thank Dr. John Hoak for his generous support and encouragement. MJL was supported by Public Health Service grant No. 2 T32 HL07344-06.

REFERENCES

1. Cross, G.A.M. (1975) *Parasitology* **71**:393-419.
2. Borst, P. and Cross, G.A.M. (1982) *Cell* **29**:291-303.
3. Esser, K.M., Schoenbecher, M.J. and Gingrich, J.B. (1982) *J. Immunol.* **129**:1715.
4. Gray, M.R., Colot, H.V. Guarente, L. and Rosbash, M. (1982) *Proc. Natl. Acad. Sci.* **79**:6598.
5. Rther, U., Konen, M. Sippel, A.E. and Mller-Hill, B. (1982) *Proc. Natl. Acad. Sci.* **79**:6852-6855.
6. Weinstock, G.M., Rys, C.A.P., Berman, M.L., Hampar, B., Jackson, D., Silhavy, T.J., Weisemann, J. and Zweig, M. (1983) *Proc. Natl. Acad. Sci.* **80**:4432-4436.
7. Hadjuk, S. and Vickerman, K. (1981) *Trans. R. Soc. Trop. Med. Hyg.* **75**:145-146.
8. Donelson, J.E., Murphy, W.J., Brentano, S.T., Rice-Ficht, A.C. and Cain, G.D. (1983) *J. Cell. Biochem.* **23**:1-12.
9. Smith, H.O. (1980) *Meth. Enzymol.* **65**:371.
10. Rice-Ficht, A., Chen, K., and Donelson, J. (1981) *Nature (London)* **294**:53-57.
11. Legerski, R.J., Hodnett, J.L. and Gray, H.B. Jr. (1978) *Nucleic Acids Research* **5**:1443.
12. Bolivar, F. and Backman, K. (1979) *Meth. Enzymol.* **68**:245-267. Eds. S.P. Colowick and N.O. Kaplan, Academic Press, New York.
13. Mandel, M. and Higa, A. (1970) *J. Mol. Biol.* **53**:154.
14. Guarente, L., Lauer, G. Roberts, T.M. and Ptashne, M. (1980) *Cell* **20**:543.
15. Morrison, D.A. (1979) *Meth. Enzymol.* **68**:326-331. Eds. S.P. Colowick and N.O. Kaplan. Academic Press, New York.
16. Holmes, D.S. and Quigley, M. (1981) *Anal. Biochem.* **114**:193-197.
17. Smith, G.E. and Summers, M.D. (1980) *Anal. Biochem.* **109**:123-129.
18. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) *J. Mol. Biol.* **113**:237.
19. Crowe, J.S., Barry, J.D., Luckins, A.G., Ross, C.A. and Vickerman, K. (1983) *Nature* **306**:389-391.
20. Maxam, A.M. and Gilbert, W. (1980) *Meth. Enzymol.* **65**:499-560.
21. Laemmli, U.K. (1970) *Nature (London)* **227**:680-685.
22. Towbin, H., Stoehelin, T. and Gordon, J. (1979) *Proc. Natl. Acad. Sci. USA* **76**:4350-4354.
23. Lagrimini, M., Brentano, S. and Donelson, J. (1984) *Nucleic Acids Res.*, **12**:604-614.

24. Boothroyd, J.C., Cross, G.A.M., Hoeijmakers, J.H.H. and Borst, P. (1980) *Nature* (London) **228**:624-626.
25. Burgess, R. (1969) *J. Biol. Chem.* **22**:6168-6176.
26. Backman, K. and Ptashne, M. (1978) *Cell* **13**:65-71.
27. Backman, K., Ptashne, M. and Gilbert, W. (1976) *Proc. Natl. Acad. Sci.* **73**:4174-4178.
28. Holder, A.A. and Cross, G.A.M. (1981) *Mol. Biochem. Parasitol.* **2**:135-150.
29. Marcu, K.B. and Williams, R.O. (1981) in *Genetic Engineering*, Setlow, Jane K. and Hollaender, A. Eds., Vol. 3, pp. 129-154, Plenum Publishing Corporation, New York.
30. Parsons, M., Smit, J., Nelson, R.G., Stuart, K. and Agabian, N. (1984) *Mol. Biochem. Parasit.* **10**:207-216.
31. Guo, L.-H., Yang, R.C.A. and Wu, R. (1983) *Nuc. Acids. Res.* **11**:5521-5540.
32. Talmadge, K. and Gilbert, W. (1980) *Gene* **12**:235-241.
33. Boothroyd, J.C., Paynter, C.A., Coleman, S.L., Cross, G.A.M. (1982) *J. Mol. Biol.* **157**:547-556.
34. Schoemaker, J.M. and Markovitz, A. (1981) *J. Bac.* **147**:46-56.
35. Pestka, S. (1983) *Arch. Biochem. Biophys.* **221**:1-37.
36. Ross, M. and Botstein, D. (1983) *Meth. Enzymol.* **101**:167. Eds. S.P. Colowick and N.O. Kaplan Academic Press, N.Y.
37. Guarente, L. (1983) *Meth. Enzymol.* **101**:181. Eds. S.P. Colowick and N.O. Kaplan Academic Press, N.Y.
38. Casadaban, M.J., Martinez-Arias, A., Shapira, S., and Chou, J. (1983) *Meth. Enzymol.* **100**:293-308. Eds. S.P. Colowick and N.O. Kaplan Academic Press, New York.
39. Lerner, R.A., Green, N., Alexander, H., Liu, F.-T., Sutcliffe, J.G., and Shinnick, T.M. (1981) *Proc. Natl. Acad. Sci.*, **78**:3403-3407.
40. Kelly, G., Lenardo, M., Rice-Ficht, A., Esser, K.M. and Donelson, J.E. (1984) Sequence Analysis of cDNAs for Surface Glycoproteins from Metacyclic Trypanosomes. *J. Cell. Biochem. Suppl.* **8B** 13th Annual UCLA Symposia p. 149 (abstr.).
41. Murphy, W.J., Brentano, S., Rice-Ficht, A., Dorfman, D., Cain, G. and Donelson, J. (1984) *J. Mol. Biol.*, in press.
42. Abbreviations used are: IaTat, Iowa Trypanosome antigenic type; cDNA-complementary DNA.