Subcloning and Transformation of a Portion of the Gene Encoding the ISG-75 N-terminal Domain of *Trypanosoma brucei*

Robert F. Cornell II, Derek P. Nolan, Paul Voorheis

ABSTRACT:

Objectives: African trypanosomiasis is a parasitic disease caused by *Trypanosoma brucei* and transmitted by the tsetse fly. Trypanosomes evade the immune system by altering their surface structure, which contains a layer of 10^7 variant surface glycoprotein (VSG) that partially shield a number of underlying invariant surface glycoproteins (ISGs).[1,2] Understanding invariant surface glycoproteins has been the recent focus of study for vaccine development and chemotherapy. Since its discovery, ISG-75 has received little study despite the fact its sequence is unique and not related to any other ISG. The focus of this project was to subclone the portion of the gene encoding the 440 amino acid N-terminal domain of ISG-75. The recombinant domain was purified and used to raise antibodies for use as a reagent to determine the orientation of ISG-75 in live bloodborne trypanosomes. The orientation of this protein is important, both for academic reasons in the attempt to understand its function and practically, as a possible target for parasite control by chemotherapy or immune control via a single chain. This involves the production of a single domain recombinant antibody that would be small enough to pass the VSG barrier and reach its ISG target. Understanding which part of the protein is exposed externally on the cellular surface is critical in selecting the portion of the protein to target. **Methods:** Polymerase Chain Reaction (PCR) was performed using genomic DNA to obtain the N-terminal domain of ISG-75. The gene fragment encoding the N-terminal domain of ISG-75 and the pET-21a vector were successfully double digested with the restriction enzymes, Nde1 and Not1. The insert and vector were cut from 1% agarose gels and extracted from the gel material. They were then ligated together and transformed into *E. coli*. **Results:** The N-terminal domain measured 1354 basepairs after PCR. After incubating the insert and vector with the restriction enzymes the N-terminal domain was visible in the appropriate region. Two smaller band fragments 600 and 754 basepairs in length were also identified. The pET-21a vector was found in the appropriate region and measured 5443 basepairs in length. *E. coli* colony counts where performed on both the ligation mixture and control LB agar plates. Thirteen colonies were found on the ligation mixture plate.

INTRODUCTION

Trypanosoma brucei, a parasitic protozoan hemoflagellate, is the causative agent of sleeping sickness in humans and nagana disease in cattle.[3] This parasite has medical, veterinary, and economic significance.[4] Humans are infected with *T. brucei* during a blood meal by the tsetse fly. The parasite migrates through the bloodstream and into the lymphatic system, multiplying by binary fission.[5] The gene encoding the variant surface glycoprotein (VSG) is switched every eight to ten days from amongst some 1000 existing gene variants. This allows the parasite to evade the humoral immune response. Due to the large number of VSG genes and their rapid rate of switching the VSG is thought to be unsuitable as a vaccine target.[6-8]

This has shifted the focus of study towards the invariant surface glycoproteins (ISGs) in the hope that their invariant nature will make them better vaccine candidates than a constantly changing VSG. Gene cloning and sequencing experiments indicate that most ISGs are composed of a large extracellular domain, a single transmembrane α-helix, and a small intracellular domain. They are also uniformly distributed across the cell surface with an abundance of 0.5% of that of the VSG. ISGs were first identified using non-penetrating surface-labeling techniques. First, surface biotinylation was used in the discovery of ISG-60, ISG-65 and ISG-75.[9,10] ISG-64, ISG-70 and ISG-100 were found using enzyme catalyzed surface radioiodination techniques.[11,12] Several other surface proteins have also been characterized in detail including the heterodimeric transferrin receptor,[13-15] a Ca^{2+}-regulated adenylate cyclase,[16,17] and the glucose transporter.[18] These proteins represent additional vaccine candidates because they do not undergo antigenic variation and deliberate immunization against such antigens may provide protection against the disease.

ISG-75 is unique and unrelated to any other ISG. The genes encoding ISG-75 are present at two loci, A and B, in tandem arrays containing five and two copies, respectively. In the A locus, a single gene encodes a structural isoform of ISG-75. At least one copy of the gene

(SW478) encoding this unique isoform of ISG-75 is present in every strain tested. ISG-75 contains a large 440 amino acid N-terminal folding domain, a short hydrophobic trans-membrane domain, and a short 34 amino acid C-terminal folding domain. It has been postulated that the N-terminal domain is external and C-terminal domain is internal to the plasma membrane.[19] However, there is little experimental evidence to validate this orientation claim for ISG-75. In ISG-75, the N-terminal domain consists of 16 tyrosines and the C-terminal contains no tyrosine. Both domains contain lysine. Given that radioiodination only detects tyrosines, this would account for the failure of this method to detect the protein in intact cells. Biotinylation modifies lysines, and detects lysine regardless of the orientation. Therefore, because tyrosine was not detected using radioiodination and tyrosine is only on the N-terminal domain, it is likely that the N-terminal is internal and the C-terminal external to the plasma membrane.

MATERIALS AND METHODS

Preparation of the N-terminal domain

Oligonucleotide PCR Primers:

The genomic DNA encoding ISG-75 was amplified by PCR with forward and reverse primers. The forward primer,
5'-ggaattccatatggaggagctctctgttgcg-3',
and the reverse primer,
5'-actagagtggcggccgctcacttcgttgtcccaactca-3',
were synthesized by the MWG-Biotechnology Oligonucleotide Laboratory.

PCR:

The specific primers were used to perform PCR thermal cycles (1 denaturing cycle of 94°C for 1 minute, 35 annealing cycles of 94°C for 30 seconds, 62°C for 1 min 20 s, 72°C for 3 min and 1 polishing cycle of 72°C for 6 min). The reaction was performed in a 50µl reaction volume containing: 10mM Tris-HCL, pH 8.3, 50mM KCl, 1.5mM $MgCl_2$, 0.2mM dNTPs, 50ng DNA template, 0.2 µmol L^{-1} of each primer, 2.5U Taq DNA polymerase (Sigma). The amplification was performed in a thermal cycler (PCR Sprint).

Restriction Digest:

The amplified partial gene product encoding the N-terminal domain of ISG-75 was double digested with the restriction enzymes Nde1 and Not1 at 37°C overnight. The digested product was electrophoresed on a 1% agarose gel using Tris-acetate-EDTA (TAE) buffer. The gel was visualized using ethidium bromide and a long wavelength ultraviolet light source. The desired band, 1354 basepairs, was cut from the gel and extracted using the Qaiger QIAquick gel extraction kit protocol.

Preparation of the vector

The expression vector, pET-21a, was used. The multicloning site contains Nde1 and Not1 restriction sites. This vector also contains an ampicillin resistant gene, which facilitates growth on ampicillin Luria-Bertani (LB) agar plates. A sample of the purified vector was digested with Nde1 and Not1. The digested product was electrophoresed on a 1% agarose gel, the band containing the cut vector excised, and extracted from the gel using the methodology stated above.

Quantification of the products

The concentrations of both the N-terminal domain and pET-21a vector were determined by comparison with a standard (Invitrogen Ready-Load 1Kb DNA ladder).

Ligation

A ligation of the N-terminal domain insert and pET-21a vector was performed. Equimolar quantities of the insert and vector were heated to 70°C for 10 minutes and then cooled. To this mixture 4.3mM Tris-HCl, pH 7.5, 0.87mM $MgCl_2$, 0.87mM dithiothreitol, 87µmol L^{-1}, 2.17µg ml^{-1} bovine serum albumin and 87U ml^{-1} of T4 DNA ligase (New England BioLabs) were added and incubated at 16°C overnight. The ligated material was then transformed into *E. coli*.

Transformation into *E. coli*

Making Competent Cells:

Several colonies of *E. coli* in TG1 were inoculated into 10 ml of SOB medium. Growth at 37°C with rapid shaking took place for four hours. The culture was placed on ice for ten minutes. The cells were recovered by centrifuging at 3000 rpm for 10 minutes, at 4°C. The cell pellet was resuspended in 3ml of TFB and kept on ice for 15 minutes. This was centrifuged for 10 minutes at 3000 rpm at room temperature. The cell pellet was resuspended in 0.8ml of TFB. 28µl of DND was added and the mixture placed on ice for 10 minutes. After this another 28µl of DND was added and the mixture kept on ice for 10 minutes.

Transformation into Competent Cells:

The ligation mixture was pipetted into 1.5ml Eppendorf tubes and placed on ice. Competent cells (210µl) were added to each of the two Eppendorf tubes, one for the ligation mixture and one for the control (cut vector with no insert). The mixtures were left on ice for 30 minutes. After this time they were heat shocked for 115 seconds at 42°C and then returned to ice. The

final product (50µl) was spread onto ampicillin resistant LB agar plates and incubated at 37°C overnight.

Analysis of Colonies:

Following overnight incubation, colony counts were performed on both control and ligation mixture LB plates.

RESULTS

Analysis of the ISG-75 Sequence

Studying the basepair sequence of variant 1 was necessary to design appropriate primers in order to obtain the N-terminal domain. ISG-75 has three genetic sequence variations, for this project the variant 1 isoform was the desired product. Complimentary primers were designed based on the flanking sequences and used in the PCR. Appropriate restriction enzymes, Nde1 and Not1, were chosen based on this information. The primers designed were

_ Nde1 cuts here

5'-ggaattcca tatggaggagctctctgttgcg-3' and

_ Not1 cuts here

5'-actagagtggc ggccgctcacttcgttgtcccaactca-3'. The underlined sections correspond to the underline non-bold/italicized sections in Figure-1 respectively.

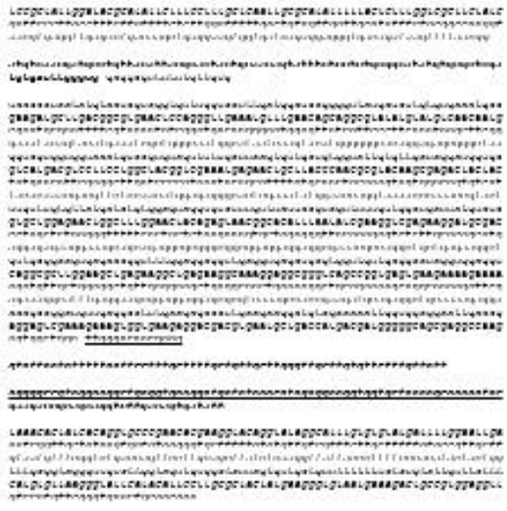

Figure-1: >ISG-75 variant-1 (genomic clone SW331), M86711,2162 BP[19]

In figure 1, each section is coded and separated for convenience. The italicized sections at each end of the sequence are noncoding regions. Underlined sections encode the leader sequence. Normal font text encodes the desired N-terminal domain. Bold and italicized section encodes the hydrophobic transmembrane domain. Bold, italicized and underlined section encodes the C-terminal domain. Underlined regions are the sequences that were used to construct the primers listed above.

PCR of the N-terminal from Genomic DNA of *T. brucei*

The expected band size of the N-terminal domain was 1354 basepairs. By comparing the band on the gel to a set of standards of known basepair size, it was found that the product of the PCR reaction was the correct size. The PCR product was cleaned, gel purified and the isolated product rerun on an agarose gel. Again, the DNA fragment encoding the N-terminal fragment was located in the expected region of the gel.

Restriction Digest of the N-terminal and pET-21a

After incubating the insert and vector with the restriction enzymes overnight, they were viewed using a long wavelength ultraviolet light source to prevent mutation. The N-terminal domain was visible in the appropriate region based on comparison with the standard. It should be noted that two smaller band fragments are also visible after this digestion. These fragments are the cleaved product of the ISG-75 variant 3 isoform. The restriction enzyme Nde1 cleaves this isoform into fragments 600 and 754 basepairs in length (see Figure 2). The pET-21a vector was also found in the appropriate region and measured 5443 basepairs in length.

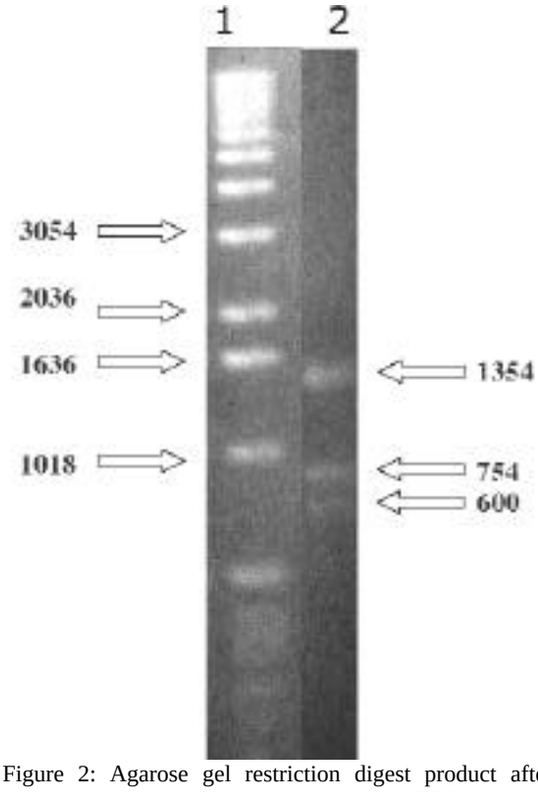

Figure 2: Agarose gel restriction digest product after incubating the N-terminal domain and pET-21a vector with restriction enzymes Nde1 and Not1. Column 1 is the control. Column 2 shows the digest products. Numbers listed are base pairs.

Transformation of Ligation Mixture into *E. coli*

E. coli colony counts where performed on both the ligation mixture and control LB agar plates. Thirteen colonies were found on the ligation mixture plate. No colonies were found on the control plate.

DISCUSSION

This experiment has provided an opportunity to learn and master many techniques in molecular biology. These include designing and running PCR, gel electrophoresis, gel extractions, ligations and transformations. A sound understanding and respect for these techniques will benefit a future medical career. These methods provide one of the modern foundations on which new drugs are developed and other therapeutic measures developed.

Other researchers will continue this project by subcloning the C-terminal domain. This 34 amino acid domain will be inserted into pGEX to produce a recombinant gene encoding a GST-fusion protein with a spacer arm linking the C-terminal domain to the GST. The purification of the C-terminal domain will be accomplished with a glutathione affinity resin.

After both domains have been subcloned the *E. coli* cell colonies of transformed material will be minipreped. The miniprep will be analyzed to ensure that the desired product has been obtained. Then the N-terminal domain will be purified using two ion exchange resins in tandem, DEAE-cellulose and CM-cellulose. Recent bioinformatic work has revealed that the 440 amino acid N-terminal domain has an isoelectric point of 5.17, using the protein characterization programme at the University of Marseille.[20] Consequently, at pH 5.17, this peptide should bind to neither DEAE nor CM cellulose, while essentially all other proteins will bind to one or the other of these exchangers, due to their charge at this pH. If necessary, a final gel filtration step can be included on either Biogel P-100 or Sephacryl S-300 to ensure complete purification.

SUMMARY

The goal of this project was to subclone the portion of the gene encoding the 440 amino acid N-terminal domain of *T. brucei* and transform it into *E. coli*. PCR was performed using genomic DNA to obtain the N-terminal domain of ISG-75. The gene fragment encoding the N-terminal domain of ISG-75 and the pET-21a vector were successfully double digested with the restriction enzymes, Nde1 and Not1. The insert and vector were cut from 1% agarose gels and extracted from the gel material. They were then ligated together and transformed into *E. coli*. The next part of this project will be to perform a similar procedure to obtain the gene fragment encoding the C-terminal domain. After this, both the N-terminal domain and the C-terminal domain will be expressed in *E. coli* and purified. These purified proteins will be used to raise antibodies for use in determining the orientation of ISG-75 in the plasma membrane of *T. brucei*. The orientation of this protein is important as a possible target for parasite control by chemotherapy or immune control via a single chain. This involves the production of a single domain recombinant antibody that would be small enough to pass the VSG barrier and reach its ISG target. Understanding which part of the protein is exposed externally on the cellular surface is critical in selecting the portion of the protein to target.

REFERENCES

1. Cabern A, Biround C, Baltz T, Maltern P. Trypasoma equiperdum: etude des variations antigeniques au cours de la trypanosome experimentle du lapin. *Exp Parasitol.* 1977; 42: 6-13.

2. Van der Ploeg LHT, Bernards A, Rijswijk FAM, Borst P. Characterization of the DNA duplication-transposition that controls the expression of two genes for variant surface glycoproteins in Trypanosoma brucei. *Nucleic Acids Res.* 1982:10: 593-609.

3. Kuzoa FA. Current situation of African trypanosomiasis. *Acta Tropica*. 1993; 10: 153-62.

4. Hursey BS, Slingenbergh J. The tsetse fly and its effects on agriculture in sub-Saharan Africa. *World Animal Review*, 1995; 3-4: 67-73.

5. Zeibig E.A. Clinical Parasitology: A Practical Approach. Philadelphia, PA: W.B. Saunders Company; 1997.

6. Cross GA. Cellular and genetic aspects of antigenic variation in trypanosomes. *Annu Rev Immunol*. 1990; 8: 83-770.

7. Borst P. Transferrin receptor, antigenic variation and the prospect of a trypanosome vaccine. *Trends Genet*. 1991a; 7: 307-9.

8. Bringaud F, Baltz T. African trypanosome glucose transporter genes: organization and evolution of a multigene family. *Mol Cell Biol*. 1993;13:1146-54.

9. Ziegelbauer K, Overath P. Identification of invariant surface glycoproteins in the bloodstream stage of Trypanosoma brucei. *J Biol Chem*. 1992; 267: 10791-96.

10. Ziegelbauer K, Multhaup G, Overath P. Molecular characterization of two invariant surface glycoproteins specific for the bloodstream stage of Trypanosoma brucei. *J Biol Chem*. 1992; 267: 10797-803.

11. Jackson DG, Windle H, Voorheis H. The identification, purification, and characterization of two

invariant surface glycoproteins located beneath the surface coat barrier of bloodstream forms of Trypanosoma brucei. *J Biol Chem*. 1993; 268: 8085-95.

12. Borst P. Molecular genetics of antigenic variation. *Immunol Today* 1991b; 12: A29-A33.

13. Salmon D, Geuskens M, Hanocq F, Hanocq-Quertier J, et al. A novel heterodimeric transferrin receptor encoded by a pair of VSG expression site-associated genes in T. brucei. *Cell* 1994; 78: 75-86.

14. Ligtenberg MJL, Bitter W, Kieft R, Steverding D, et al. Reconstitution of a surface transferrin binding complex in insect form Trypanosoma brucei. *EMBO J*. 1994; 13: 2565-73.

15. Chaudhri M, Steverding D, Kittelberger D, Overath P. Expression of a glycosylphosphatidylinositol-anchored Trypanosoma brucei transferrin-binding protein complex in insect cells *Proc Natl Acad Sci*. 1994; 91: 6443-47.

16. Rolin S, Halleux St, Van Sande J, Dumont JE, et al. Stage-specific adenylate cyclase activity in Trypanosoma brucei. *Exp Parasitol*.1990; 71: 350-2.

17. Paindavoine P, Rolin S, Van Assel S, Geuskens M, et al. A gene from the variant surface glycoprotein expression site encodes one of several transmembrane adenylate cyclases located on the flagellum of Trypanosoma brucei. *Mol Cell Biol*. 1992 Mar; 12(3):1218-25.

18. Nolan DP, Jackson DG, Windle HJ, Pays A, et al. Characterization of a novel, stage-specific, invariant surface protein in Trypanosoma brucei containing an internal, serine-rich, repetitive motif. *J Biol Chem*. 1997 Nov 14; 272(46):29212-21.

19. Ziegelbauer K, Rudenko G, Kieft R, Overath P. Genomic organization of an invariant surface glycoprotein gene family of Trypanosoma brucei. *Mol Biochem Parasitol*.1995; 69: 53-63.

20. Gleeson, M. Protein Composition. 2001. Available from : URL : http://www.up.univ-mrs.fr/~wabim/d_abim/compo-p.html.