ABSTRACT
This project strive for understanding the mechanism with which RNA lariats are debranched by the *P. falciparum* lariat debranching enzyme, in the intron turnover pathway. Expression assays for the full-length protein and its two subunits, respectively, are under development. Structural predictions using the SAM-T02 server shows that the catalytic site, containing histidines and acidic residues, coordinating manganese ions, is probably located on the N-terminal subunit which belong to the metallophos family. The C-terminus predictions are less certain since there is little knowledge about this unit that is common in the DBR enzymes and fewer sequences to align it with. Site-directed mutagenesis of the catalytic site will be made with the purpose of creating an enzyme that binds lariats without debranching them, which would be a useful tool in the study of introns.

INTRODUCTION
More and more information is gained on pre-mRNA splicing mechanisms, but little is yet known about the biological and evolutionary significance of introns. We therefore need a simple way of investigating those, and the enzymes involved in the intron turnover pathway. Since genes are sequenced much faster than structural determinations with X-ray chrystallography or NMR is performed for their corresponding proteins, new bioinformatics techniques for structure predictions made on homology between proteins are used.

The RNA lariat debranching enzyme was first identified in extracts from HeLa cells. The enzyme is a part of the intron turnover pathway enzyme in mammalian cells and yeast cells [1]. The enzyme can use a number of different substrates, as long as they have a 2’-5’-phosphodiester linkage [2]. It cleaves 2’-5’-phosphodiester linkages that exist at intron RNA lariat branchpoints, thus making the intron lariats linear. This seems to be a rate limiting step in the pathway because intron lariats are accumulated in strains with mutations in the *dbr1* gene [3]. Both human and yeast RNA lariat debranching enzymes have high specificity, cleaving only 2’-5’-phosphodiester linkages and no other e.g. 3’-5’-linkages. This specificity is rare in phosphodiesterases in general [1]. After linearization the introns are rapidly degraded by exonucleases [3] which need a free 5’-end to cut from.

The short-term goal of this study is to develop an assay for expressing the target enzyme, the RNA lariat debranching enzyme (DBR) from mosquito-borne parasitic protozoan *Plasmodium falciparum*. After developing such an assay it is desirable to produce a protein that binds lariats without debranching them. This would be a useful laboratory reagent for studying introns, as it would allow easier purification and extraction of intron lariats. The non-debranching enzyme will be created with site-directed mutagenesis, with mutations based on structural predictions using bioinformatics tools, as no experimental structural determination has been performed for the lariat DBR protein.

The DBR protein appears to consist of two domains: an N-terminal domain similar to other phosphatases, that probably contains the catalytic site, and a C-terminal domain that is
unique to debranching enzymes and probably is part of the recognition of branch points [4]. Assays for expressing these two subunits separately will be developed to make it possible to determine whether either of them has catalytic activity by itself, and whether either binds lariats by itself. The current hypothesis is that both domains are essential for both binding and debranching, but it is possible that the N-terminal domain by itself may do some debranching through non-specific interactions, and that the C-terminal domain is to prevent the enzyme from being too promiscuous in vivo.

MATERIALS AND METHODS

Structural predictions
Information on domain structures of the *P. falciparum* DBR protein related DBR proteins were gained with SwissPfam (http://us.expasy.org/sprot/) [4].

Structural predictions were made using the UCSC SAM-T02 server (http://www.soe.ucsc.edu/research/compbio/HMM-apps/T02-query.html), since no experimental determination has been done for the DBR proteins. SAM-T02 uses multitrack Hidden Markov Models for prediction of the structure of a given target amino acid sequence. Alignments are made to similar sequences, found with SAM-T2K, and structural alphabets (such as DSSP, STRIDE, STR etc. [5]) are used for predicting local structure properties with neural nets and fold-recognition approach (SAM-T02). Alignments are made to the template and a target-specific library is built by FRAGFINDER. Finally, a 3D structure prediction is made by UNDER-TAKER that assembles the fragments and the fold-recognition alignments [6][7].

More advanced predictions have been done by Kevin Karplus (Computer Engineering Department, University of California, Santa Cruz) using SAM-T04 software [3][8].

Primer design
Primers, DBR_forward and DBR_reverse, were designed with Primer3 (http://frodo.wi.mit.edu/cgi-bin/prime3/prime3 www.cgi) to amplify the total *P. falciparum* RNA lariat debranching enzyme (1725 bp product). Two additional primers, N_reverse and C_forward, annealing to the linker region between the subunits, were designed to enable amplification of the N-terminus (867 bp product) and the C-terminus (899 bp product), respectively.

The 1728 bp DNA sequence was retrieved from NCBI Genome database (http://www.ncbi.nlm.nih.gov), and by using the Malaria browser (http://exon.ucsc.edu/cgi-bin/hgGateway?db=falciparum) for chr13:1,618,196-1,619,923 in November 2004 assembly, extra bases were added upstream of the gene, enabling the DBR_forward primer to amplify the entire gene. The DBR_reverse primer was chosen to end just before the stop codon, to prepare for adding an expression His-tag. InSilico PCR for *P. falciparum* (http://exon.ucsc.edu/cgi-bin/webPcr) was used to verify that primers did not have multiple matches in the malaria genome. Restriction sites were added to the 5'-ends of the primers, to enable insertion in the cloning vector, pET-24b (Novagen), after control that the enzymes would not cut within the gene (http://rna.lundberg.gu.se/cgi-bin/cutter2/cutter).

The potential primers, DBR_forward 5'-gagagacatATGTTTATTGCTGTTGGTTGTT-3’ (T_m = 58.43 °C, restriction enzyme NdeI), DBR_reverse 5’-ccttccctcgaggccgccCT ATAGTACATTCAATTCTACCTC-3’ (T_m = 58.53 °C, XhoI), N_reverse 5’-agacgctagATCGTACGTTGTGTTGTTGTTGTT-3’ (T_m = 58.68 °C, XhoI), and C_forward 5’-agagacgatcc ggtTTTGATCATGTTCTAGGTTGTTGTTGTTGTTGTTGTTGTT-3’ (T_m = 58.03 °C, BamHI) were ordered from Sigma Genosys.

PCR, cloning and expression
Genomic DNA was isolated from infected erythrocytes kindly provided by Kausik Chakrabarti (Molecular Cell & Developmental Biology, University of California, Santa Cruz)
with a lysis buffer (10 mM Tris HCL, 1 mM EDTA, 0.1 M NaCl, 0.5% SDS, 0.1 mg/ml proteinase K), phenol:chloroform:isoamylalcohol (25:24:1) extraction and precipitation in high salinity ethanol solution. PCR was performed with Platinum Pfx DNA polymerase (Invitrogen), with proofreading activity, according to manufacturer’s recommendations, with 1.5 mM MgSO4. Following PCR program was used to amplify the C-terminus: denaturation 94 °C, 2 min, 1 cycle; denaturation 94 °C, 30 sec, annealing 54 °C, 45 sec and elongation 72 °C, 2 min, 30 cycles; elongation 72 °C, 10 min. Amplification of the full DBR protein and the N-terminus, respectively, required following modifications to the schedule: annealing temperature of 50 °C and elongation time of 2 min 30 sec, during the 30 cycles, as well as primers in excess was successfully performed by Kausik Chakrabarti. Since it took a long time before amplification of the full length protein and the N-terminus was successful, cloning has so far only been performed for the C-terminus.

The C-terminus was purified using QIAquick PCR Purification Kit protocol (Qiagen), and the resulting fragments and the vector pET-24b cut with restriction endonucleases BamHI and XhoI (Fischer Scientific) in 100 µl overnight reactions (1X buffer 2, 1 µg/µl BSA, 0.02 U/µl BamHI, 0.02 U/µl XhoI and 10.8 µg/µl C-term DNA or pET-24b, respectively). Digested C-terminal and pET-24b fragments were ligated in a 10 µl reaction (0.3 U/µl T4 DNA Ligase (Promega), 1X T4 buffer, 2.82 ng/µl C-terminal DNA, 1.56 ng/µl pET-24b). Ligation product and uncut pET-24b as negative control, were transformed into E. coli Top 10 with Qiagen PCR Cloning Plus Kit Transformation protocol (Qiagen), and grown over night on Kan+ LB-plates. Plasmids were retrieved using QIAspin Spin Miniprep Kit protocol (Qiagen) and an additional restriction digest performed for verification of transformation of correct clones. A master plate was made for successful transformants.

Plasmids originating from two colonies on the master plate were grown over night in LB-medium, and isolated with QIAprep Spin Miniprep Kit (Qiagen). E. coli BL21DE3 expression host were made competent, plasmids transformed into them, and grown in a large culture. SDS-page confirms expression of the protein.

RESULTS

Structure predictions

Searching SwissProt for information gained on lariat debranching enzymes show that most of them have two subunits: a metallophos subunit (corresponding to the N-terminus), and a DBR subunit (corresponding to the C-terminus). The P. falciparum DBR protein has the N-terminal subunit, that is supposed to have manganese ion ligands, and probably also the C-terminal [4].

Multiple alignments made with SAM-T02 show high conservation for the N-terminal subunit, while the conservation is lower for the C-terminal unit (see Fig 1). The lack of data due to that the P. falciparum DBR protein is longer than many other debranching proteins generates fewer sequences to compare to, which may decrease the conservation. Secondary structure predictions are also stronger for the N-terminal subunit.

One of the proteins with the most likely similar structure to P. falciparum DBR enzyme is P. Furiosus Mre11 with manganese and damp (PDB id 1II7) (see Fig 2) [8]. It has two Mn2+ ions that are octahedrally coordinated by seven conserved residues of the phosphodiesterase motifs (Asp8, His10, Asp49, Asn84, His173, His206, His208) and by a bridging water molecule [9], and according to predictions made by Kevin Karplus, it is likely that the corresponding residues to these in P. falciparum DBR, in some order are His10, Glu12, Asp39, Asn85, His86, Glu87, Glu139, His175 [8], that probably also coordinate manganese ions.
Expression assay

There were problems with the PCR amplification for the N-term and full-length DBR, but the PCR is now working (see Fig 3a and b). Cloning of the C-terminus has been shown successful (see Fig3c). The expression assay is not yet finished, and expression work is currently performed on the C-terminus, with the plasmids isolated from the transformation host. Expression will also be performed for the N-terminus and the full-length protein.

DISCUSSION

The structural predictions of the N-terminal can be assumed as reasonably correct since there are many templates, and it belongs to a well known family, the metallophos enzymes. Structures have been solved for enzymes with similar structures that also need manganese for
activity. It is more uncertain if the C-terminus is correctly predicted, since there are much fewer relatives to align with, and no structural determinations performed for these types of enzymes, and therefore probably hard to tell whether a predicted structure is accurate or not, without experimental analysis such as X-ray crystallography.

The attempt to create a non-debranching enzyme, for generation of introns, will be performed with site-specific mutagenesis, based on the assumption that the histidines and acidic residues in the predicted catalytic site of the N-terminus coordinate manganese ions. The main problem is how mutations can be performed without changing the structure of the catalytic site, which is probably essential for binding the introns. One approach is to substitute one of the histidines with a smaller, neutral residue like valine that lacks the nitrogens in histidine’s imidazole ring, which will keep the acidic environment, but hopefully change the conformation enough to destroy the essential binding of manganese.

We are close to having developed assays for expressing the *P. falciparum* RNA lariat debranching enzyme and its N-terminal and C-terminal subunits. To confirm correct expression and purification of the lariat debranching enzyme we will need to develop an activity assay. Manny Ares (Molecular Cell & Developmental Biology, University of California, Santa Cruz, USA) has suggested a way of doing this by generating large numbers of lariats using a self-splicing intron, and then mix the lariats with the purified protein. It should be possible to separate the lariats from the debranched lariats on a gel, as the circularization of the RNA should change its rate of movement in the gel. Although this assay should work for determining whether or not we have debranching activity, we will need a different assay for determining whether a mutant protein binds lariats without debranching them. We have not yet figured out a simple assay for testing this. We will also need assays for activity measurement for the two individual subunits, and for examination of whether the two domains fold independently.

It has been shown in *Schizosaccharomyces pombe* [3] that intron accumulation cause severe growth effects, and it has been suggested that lariat debranching is a rate-limiting step in the breakdown of introns [2]. This makes the lariat DBR protein a suitable drug target, and our long-term goal is to find a possible treatment for malaria, a disease that kills 2.7 million people and infects 300 – 500 million people annually, by manipulating the *P. falciparum* lariat DBR enzyme. The *Plasmodium* debranching enzymes differ from the mammalian ones by being longer, with the part after E453 not aligning to the mammalian debranching
enzymes. If this part turns out to be essential for *Plasmodium* (all the *Plasmodium* species have a conserved sequence at the C terminus), then it might make a possible drug target that is specific to *Plasmodium*. This is a key characteristic since interfering with the catalytic domain would most likely also interfere with human DBR, which would not be desirable in a drug.

**ACKNOWLEDGEMENTS**

Thanks to Manny Ares, in whose lab the experimental work has been performed, and to Kausik Chakrabarti who has been supervising the work, both at Molecular Cell & Developmental Biology, Biology Department, University of California, Santa Cruz. Thanks also to Kevin Karplus (Computer Engineering Department, University of California, Santa Cruz), who has done the structural predictions and wet-lab work on the DBR protein.

**REFERENCES**