CHAPTER THREE- RESULTS

3.1 Purification of 4.1R

Detergent and salt-based methods were used for purifying 4.1R from human erythrocyte membranes. The protocols relied on the depletion of glyceraldehyde-6-phosphate dehydrogenase (band 6), spectrin and actin from erythrocyte ghosts prior to 4.1R extraction (Figure 9). Band 6 was removed using PBS containing 155mM sodium chloride, which facilitated the disruption of its weak electrostatic interaction with the skeletal framework. Spectrin was depleted by using a strongly alkaline, low ionic strength solution (0.1mM EDTA, pH 9.2). Together with a reducing agent (β-mercaptoethanol), these conditions promoted the dissociation of spectrin oligomers from the membrane (Cole and Ralston, 1992;

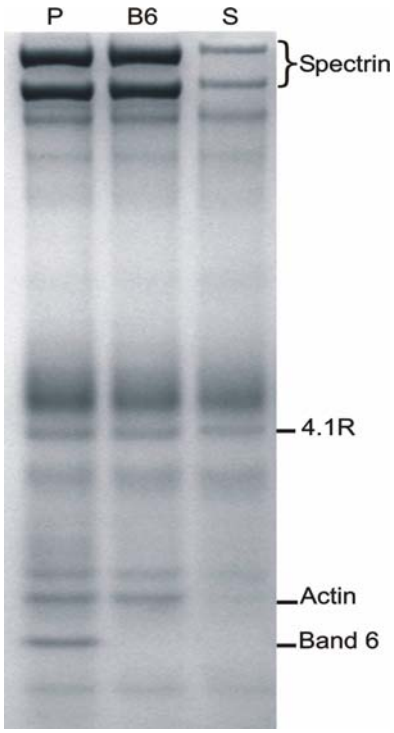

Figure 9- Band 6, spectrin and actin extraction from erythrocyte membranes

Erythrocyte membranes were prepared from venous blood. Approximately 25μg protein samples from successive steps in the 4.1R purification procedure were analysed by SDS-PAGE on a 3.5-17% exponential gradient Fairbanks gel. Depletion of spectrin, actin and band 6 is clearly evident. **P**, erythrocyte membrane preparation; **B6**, band 6 extraction and **S**, spectrin and actin extraction

Fujita et al., 1998). As a result, actin and a small amount of 4.1R were also depleted from the vesicles due to their association with spectrin.

3.1.1 Detergent-based purification of 4.1R

The treatment of vesicles with Tween 20 (Becker et al., 1983) resulted in 4.1R extracts depleted of major erythrocyte membrane proteins. However, analysis revealed that residual spectrin and actin were extracted simultaneously with 4.1R (Figure 10A and B). Scanning densitometry of a

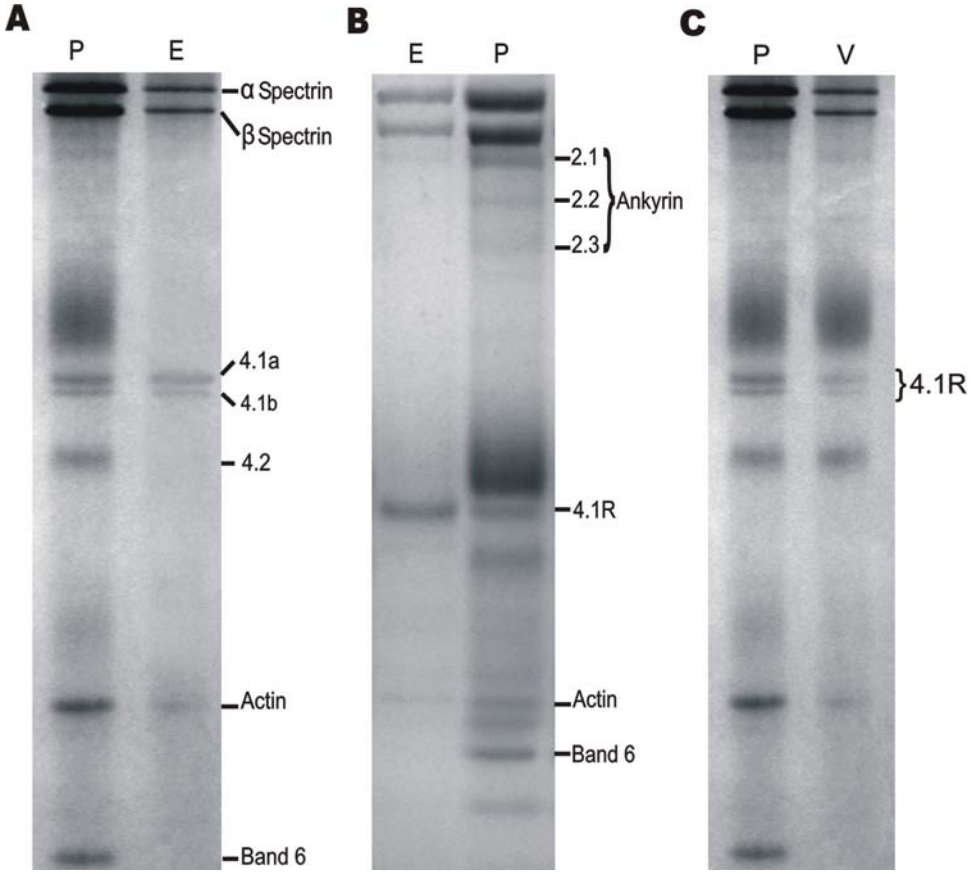

Figure 10- Detergent-based extraction of 4.1R from erythrocyte membranes

An erythrocyte membrane preparation (**P**) and Tween 20-4.1R extract (**E**) were electrophoresed on a 12% Laemmli gel (**A** and **C**) and exponential gradient Fairbanks system (**B**). The Laemmli gel resolves all major erythrocyte proteins including 4.1R into its isomers, 4.1a and 4.1b, however it does not separate β spectrin and ankyrin. 4.1R resolves as one band on a Fairbanks gel, whereas ankyrin separates out into three isoforms (2.1, 2.2 and 2.3). 4.1R was not completely removed from erythrocyte vesicles following Tween 20 treatment (**V**).

12% Laemmli gel showed that up to 30% of 4.1R remained attached to the membrane vesicles, probably due to interaction with the residual spectrin (Conboy, 1993; Takakuwa, 2000; Walensky et al., 2003).

4.1R was purified from the extracts by DEAE-anion exchange chromatography. Purified protein eluted from the column between a sodium chloride concentration of approximately 140-200mM (Figure 11A), which was confirmed by SDS-PAGE analysis of the relevant fractions (Figure 11B). Spectrin, with residual amounts of 4.1R, eluted at 200mM sodium chloride. A major peak registering an absorbance of 0.45 at 80mM sodium chloride was identified in the chromatogram. The peak was free of major proteins (data not shown), but the high A_{280} reading suggested the presence of contaminants which absorb efficiently at 280nm UV light. The heterocyclic ring system of haem groups (Voet and Voet, 1995), which dissociated from haemoglobin subunits during Tween 20 treatment, was probably responsible for this finding. The inclusion of PMSF throughout the purification and elution of 4.1R was important for preventing proteolysis of the protein. Despite this, 4.1R degraded on the anion exchange matrix. PMSF is a reversible inhibitor and only provides protection against proteolysis by serine proteases. In addition, the inhibitor has a short half-life (30 minutes) in aqueous solutions at alkaline pH. An immunoblot was performed (Figure 11C) not only to verify the identity of 4.1R, but to confirm the presence of smaller 4.1R fragments.

The protocol for detergent-based purification generated a relatively large amount of protein. Fractions 23-28 were pooled and the purity of native 4.1R was judged to be in excess of 85%, as measured by scanning densitometry (Figure 12). Approximately 4mg of pure intact 4.1R was obtained from 60ml whole blood, representing 72% of the maximum expected yield.

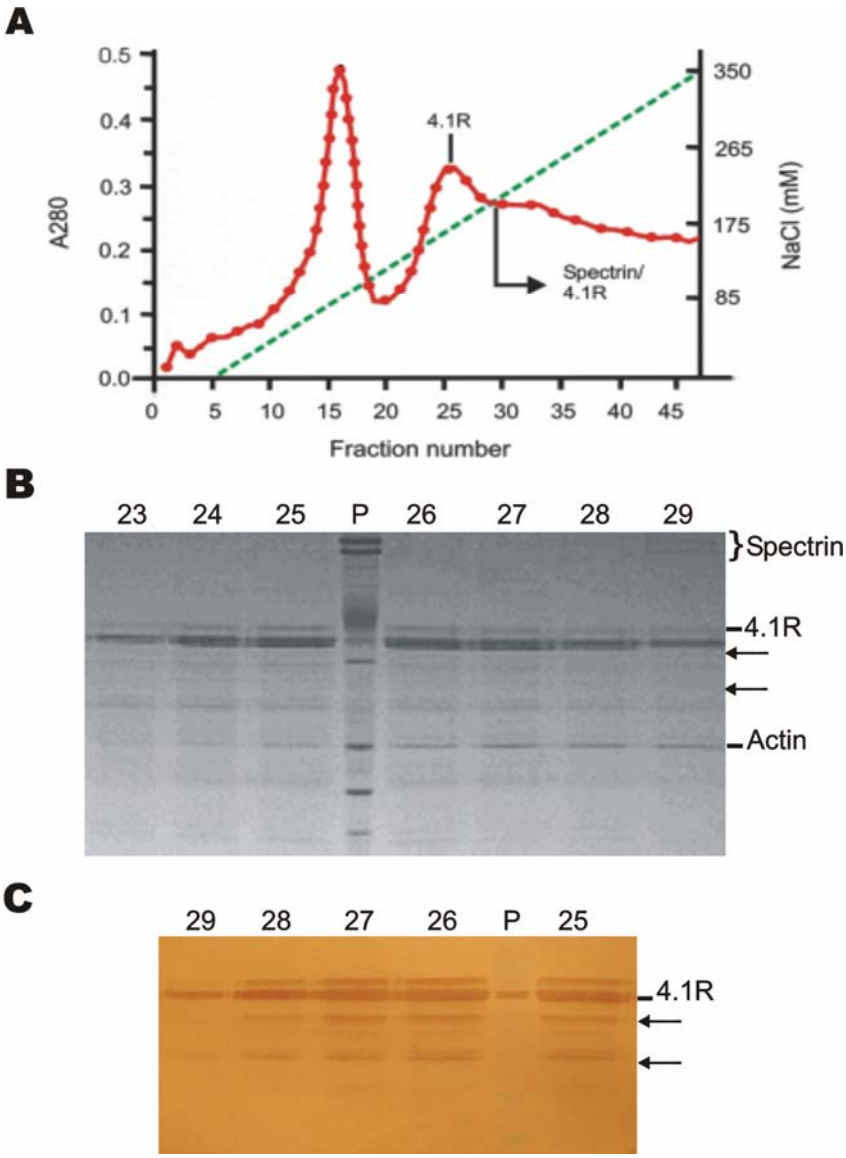

Figure 11- Purification of 4.1R from crude extracts

A) Chromatogram of 4.1R extracts on a DEAE-anion exchange column. Elution was performed with a continuous gradient of 0 to 350mM NaCl in equilibration buffer (2.1.3). The gradient was initiated at fraction 5, which corresponded to the bed volume of the DEAE matrix. Fractions were collected and monitored for absorbance at 280nm UV light. Pure 4.1R eluted from fraction 23 (140mM NaCl), whereas spectrin and 4.1R complexes eluted at fraction 29 (200mM NaCl). **B**) SDS-PAGE analysis of samples 23-29 collected from the column. Purified 4.1R appeared in samples 23-25 with 4.1R and actin eluting in fractions 26-29. Spectrin eluted from fraction 29 but is not visible against the dark background of the photograph. Arrows indicate proteolysed 4.1R fragments. **P**, erythrocyte membrane preparation. **C**) Immunoblot to confirm the identity of purified 4.1R using rabbit anti-4.1R antibody and 1, 3, diaminobenzidine. Arrows indicate smaller molecular weight fragments of 4.1R.

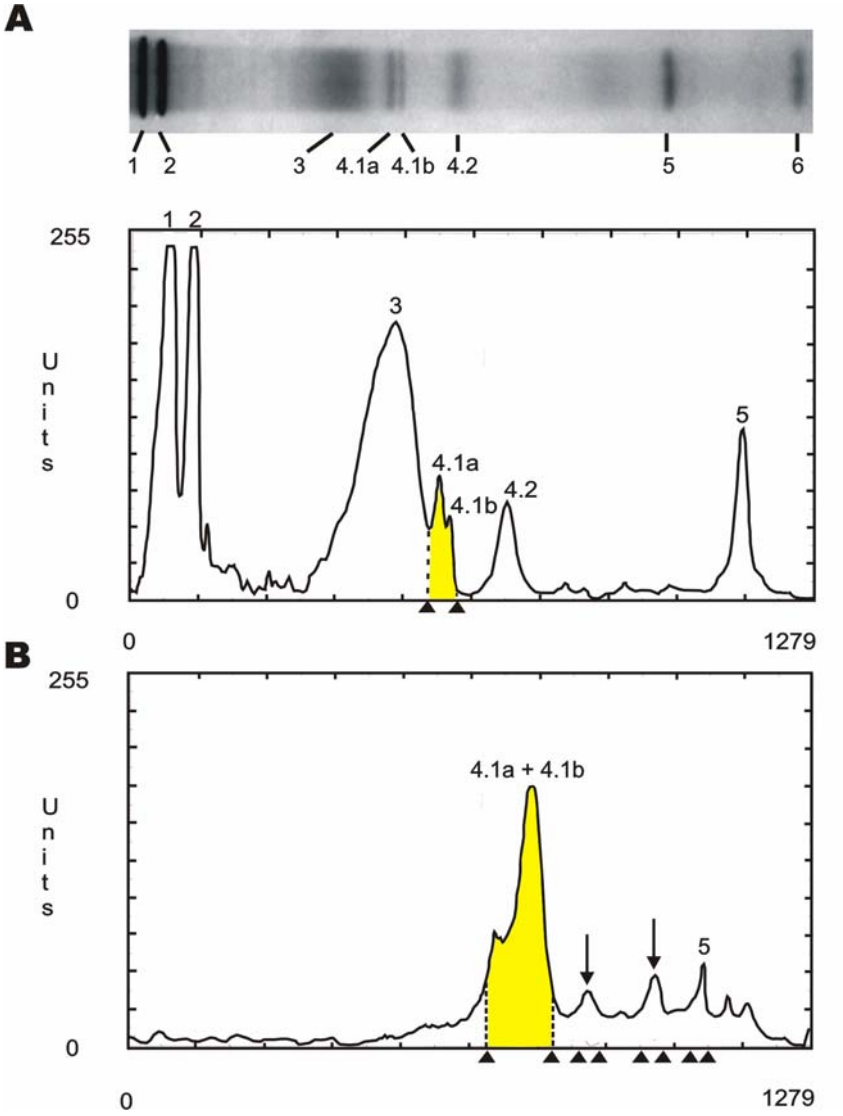

Figure 12- Densitometric scanning of erythrocyte membrane proteins

A) A 12% Laemmli gel was stained with Coomassie blue and analysed by scanning densitometry. The major protein bands from a membrane preparation are indicated: 1, α spectrin; 2, β spectrin; 3, band 3; 4.1R isoforms (4.1a and 4.1b), band 4.2, band 5 (actin) and band 6. The protein content was calculated from the area under the peaks. The yellow region demarcates 4.1R, which is approximately 5% of total erythrocyte membrane proteins. **B**) Densitometric scan of purified 4.1R. The yellow area represents intact 4.1R and was calculated to be 85% of the total protein. The arrows indicate areas corresponding to proteolysed 4.1R fragments (figure 11B/C). Band 5 or actin comprised approximately 4% of the purified 4.1R preparation.

Although this method produced a high yield of 4.1R, it was inconsistent and generally unreliable. Several attempts to purify 4.1R were unsuccessful due to the presence of contaminating spectrin and actin (Figure 13). Fifteen ml of crude 4.1R preparation (±8.5mg protein) was loaded onto a DEAE ion-exchange matrix. On application of a linear salt gradient, both proteins co-migrated through the resin, whereas stepwise elution with different sodium chloride concentrations proved just as problematic.

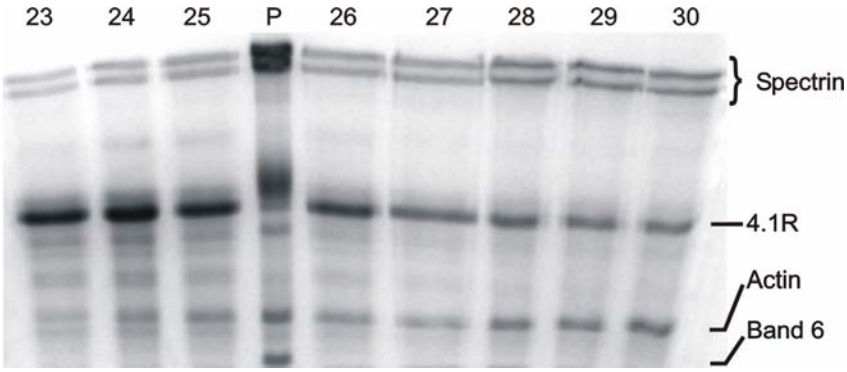

Figure 13- Elution of 4.1R and contaminating proteins from a DEAE ion exchange column

Proteins were eluted using a continuous salt gradient and aliquots of eluate collected for analysis on a 12% Laemmli gel. Fractions 23-30 contain 4.1R, spectrin and actin as indicated. Spectrin and actin should typically elute at later salt concentrations, however the proteins dissociated from the DEAE matrix together with 4.1R for the majority of preparations. **P**, erythrocyte membrane preparation.

3.1.2 Salt-based purification of 4.1R

To improve the purity of 4.1R, the method described by Tyler and colleagues (1980) was used. The protein was extracted from erythrocyte membranes with a high concentration of salt (1M KCl). 4.1R extracts contained a number of contaminants, predominantly ankyrin and actin (Figure 14). The crude extract was loaded onto a DEAE column and after removal of minor contaminants, 100mM KCl was successfully used to elute intact 4.1R. The use of protease inhibitors (PMSF and Pefablock® *SC) and less time (±20 minutes) for DEAE chromatography prevented any notable degree of 4.1R degradation. Pefablock is an irreversible

serine protease inhibitor and stable at alkaline pH. The inhibitor was not used in the Tween 20-extraction method and was probably the main reason for reduced 4.1R proteolysis during salt extraction and purification. 4.1R was found to be pure as determined by silver staining of a 12% Laemmli gel (Figure 14). It is possible though that the 4.1R preparation may have been contaminated with protein kinases, which were not visible by gel electrophoretic analysis (Husain and Branton, 1986).

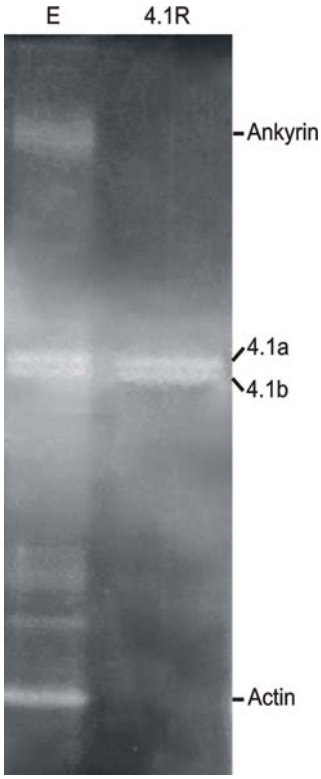

Figure 14- Salt-based extraction of 4.1R from erythrocyte membranes

4.1R was isolated from vesicles depleted of band 6, spectrin and actin using 1M KCl. Ankyrin and actin were major contaminants which were removed by chromatography. The 4.1R extract (**E**) and purified protein (**4.1R**) were analysed by Laemmli SDS-PAGE and silver staining. No contaminants or proteolysed 4.1R fragments were visible.

Salt treatment was not as efficient as Tween 20 for extracting 4.1R from erythrocyte membranes depleted of spectrin and band 6. The method produced significantly lower yields of 4.1R and only 0.1-0.2mg of intact protein was obtained from 60ml venous blood, which corresponded to less than 4% of the maximum expected yield. Despite this drawback, the

method described by Tyler and colleagues was consistent in generating pure and intact protein. Due to the low yield, 4.1R was extracted from 200ml whole blood or approximately 80ml packed erythrocytes, after which the purified protein was concentrated to approximately 2μg/μl for biotinylation.

3.1.3 Protein biotinylation

Purified 4.1R (Figure 14) and 4.1R/spectrin samples (Figure 10A/B) were biotinylated for biopanning against *P. falciparum* phage display libraries. The reaction was verified using a streptavidin-peroxidase test (2.2.1) (Figure 15).

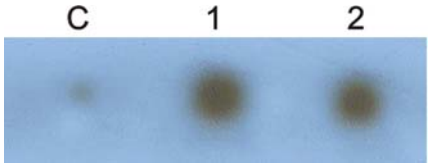

Figure 15- Assay to confirm protein biotinylation

Approximately 2μg of unbiotinylated and biotinylated protein samples were spotted on a nitrocellulose membrane. Biotin label was detected with a streptavidin-peroxidase conjugate using Supersignal® West Pico Chemiluminescence substrate. **C**, unbiotinylated 4.1R; **1**, biotinylated spectrin/4.1R and **2**, biotinylated 4.1R.

3.2 Construction of *P. falciparum* phage display libraries

3.2.1 RNA extraction

The method described by Chomczynski and Sacchi (1987) was used with modifications for isolating total RNA from *P. falciparum* parasites (strain FCR3). Maximum RNA yield was obtained from cultures containing a high parasitaemia (12-15%) and predominately late asexual stages. 20-30μg total RNA was extracted per 20ml culture volume, and 2-4.5μg mRNA purified using Dynal® Oligo$(dT)_{25}$ bead technology. Spectrophotometric analyses revealed A_{260}: A_{280} ratios between 1.8 and 2.0, indicating pure

RNA samples. RNA was evaluated qualitatively by estimating the intensities of the different ribosomal RNA subunits (Figure 16). Intact, high quality RNA has a 28S: 18S ratio of approximately 2:1 (Ambion TechNotes 11-1: http://www.ambion.com/techlib/tn). Some preparations were contaminated with parasite genomic DNA, which was removed by digesting with RNase free DNase. mRNA was visible as a smear on the agarose gel. Due to the AT rich genome of the *P. falciparum* parasite, a small proportion of ribosomal RNA was captured on the Dynabeads, along with poly $(A)^+$ mRNA. As a result, all mRNA preparations were contaminated with 28S and 18S transcripts.

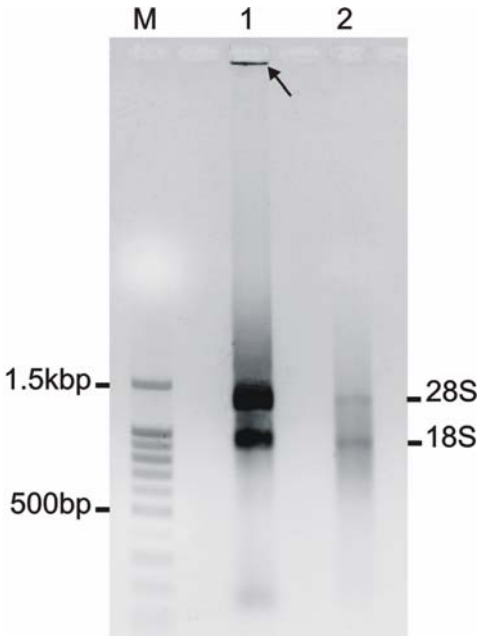

Figure 16- Analysis of *P. falciparum* RNA preparations

Total RNA (**1**) and mRNA (**2**) from asexual parasites were electrophoresed alongside a double-stranded DNA size marker (**M**) on a 1% agarose gel. The 28S and 18S ribosomal subunits migrate at approximately 1.4kbp and 900bp respectively. The arrow indicates contamination with high molecular weight parasite DNA. bp, base pairs and kbp, kilobase pairs.

3.2.2 Preparation of parasite cDNA

P. falciparum cDNA was synthesised from purified mRNA (2.5.1.1) and the ends modified (2.5.1.2) for cloning into the T7Select 10-3b vector. End-

modified cDNA was fractionated on Sepharose resin to remove cDNA synthesis products of less than 300bp (2.5.1.2). An aliquot of the void volume was electrophoresed on a denaturing polyacrylamide gel alongside two different size markers (2.5.1.3) (Figure 17). The average size of synthesised parasite cDNA was greater than 722bp in length.

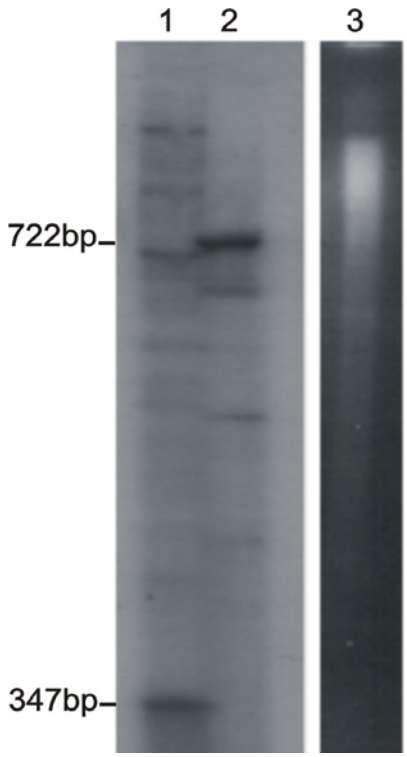

Figure 17- Gel filtration of *P. falciparum* end-modified cDNA

Size markers (lanes **1** and **2**) were synthesised by PCR utilising primers specific for human erythrocyte α-spectrin and band 3 (A-2.2). Lane 3 represents parasite double stranded DNA, which appears as a smear of an average size greater than 722bp. cDNA and PCR products were labeled with [α^{32}P] dATP and evaluated using an 8% denaturing polyacrylamide gel and autoradiography.

3.2.3 *In vitro* packaging and plate lysate amplification

Parasite cDNA was cloned in T7Select 10-3b vector arms (2.5.2.1) and packaged into bacteriophage particles using T7 extracts (2.5.2.2). The number of phage recombinants generated was determined with a plaque assay (Figure 18). The titre of the *P. falciparum* phage library and control library was 10^6pfu/ml and 10^9pfu/ml respectively. Efficiency of packaging

and/or ligation between parasite cDNA and T7Select vector was 10^5pfu/μg vector. This was significantly lower than the efficiency of the positive control, which was 10^8pfu/μg vector. A favourable insert: vector ratio had been selected for cloning fragments of 1.5kbp. However, the actual fragment sizes were smaller and may have accounted for the difference in efficiency. To express and display cloned sequences on the surface of phage particles, a single round of amplification was necessary. The *P. falciparum* phage display library was amplified by plate lysate amplification and this generated a titre of 10^{10}pfu/ml as determined by plaque assay. To enrich the library in phage specific for spectrin/4.1R and purified 4.1R, four rounds of biopanning were performed against each sample (2.6).

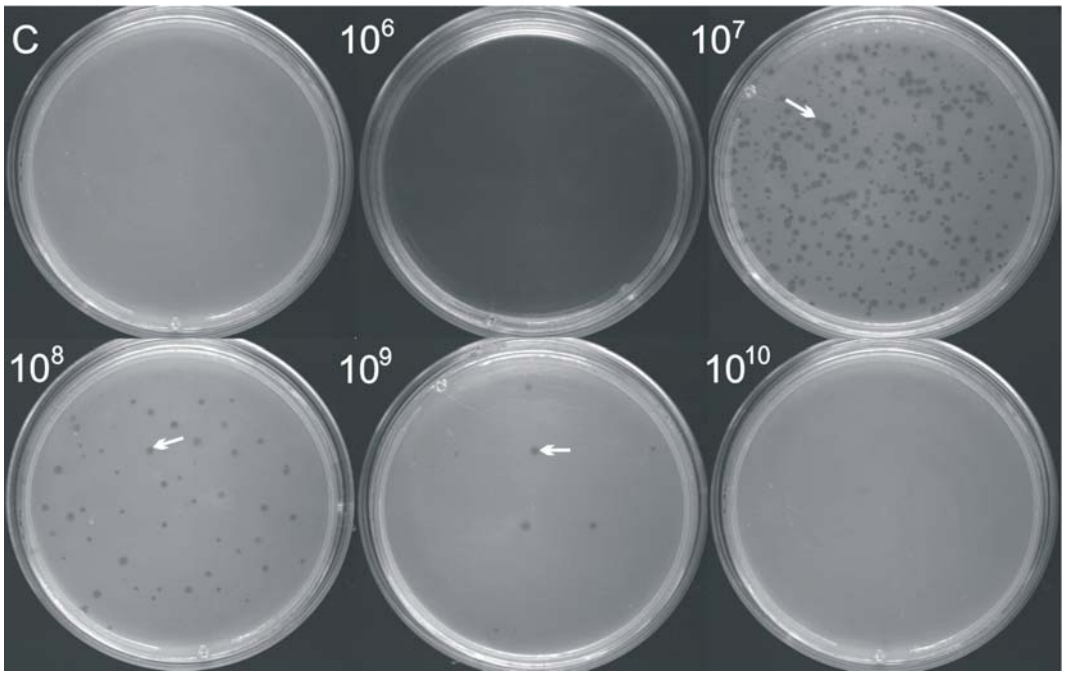

Figure 18- Plaque assay

LB/agar plates were overlaid with top agarose containing *E. coli* (strain BLT5403) and serial dilutions (**10^6-10^{10}** times) of recombinant phage, prepared from the *P. falciparum* library. Arrows pointing to the darker spots on the plates represent plaques or areas of phage growth. Phage diluted one million times completely lysed the E. coli host cells. A control plate (**C**) contains a lawn of bacterial cells only. The phage titre described in plaque forming units per ml is calculated by multiplying the number of plaques by the serial and plating dilutions.

3.3 Identification of interacting parasite peptides

3.3.1 Size determination of cDNA inserts

Multiple plaques from fourth round biopanning were selected for verification of cDNA inserts. DNA was extracted from a total of 200 different phage clones and the *P. falciparum* inserts amplified using the T7Select primers. The majority of inserts were 100-200bp in length, and the largest insert was ±300bp (Figure 19). All PCR fragments included 107bp of vector sequence. The presence of small inserts was unexpected, since gel filtration of the modified cDNA should have excluded fragments smaller than 300bp. Electrophoretic analysis did not reveal smaller DNA fragments (Figure 17), but since these would have incorporated a very low amount of radiolabel, they were not visible on the gel. Inefficient size exclusion during column chromatography presumably resulted in a higher molar ratio of smaller to larger cDNA fragments, which would therefore have facilitated a favourable cloning efficiency for the smaller fragments.

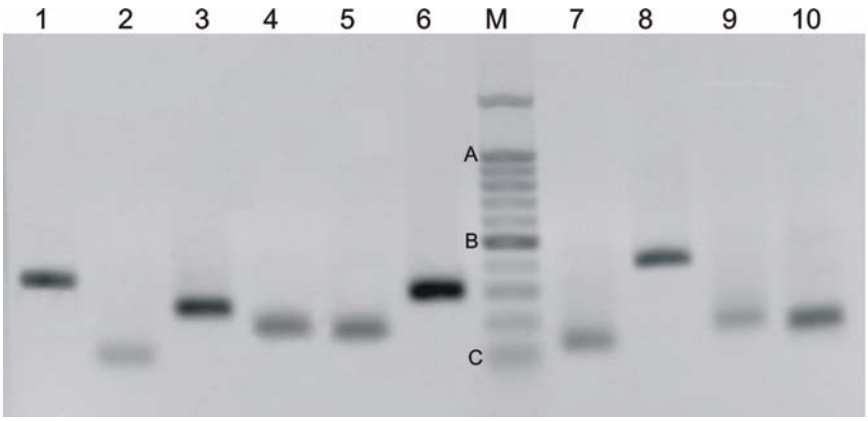

Figure 19- Analysis of parasite cDNA inserts cloned into the T7Select 10-3b vector

Lanes **1-10** depict amplified DNA from selected phage recombinants. All PCR fragments include 107bp of vector sequence. The PCR product in lane 2 contains no parasite insert. **M** represents a 100bp DNA ladder: **A**, 1000bp; **B**, 500bp and **C**, 100bp. PCR products were evaluated using 1% agarose gel electrophoresis.

3.3.2 Sequencing and bioinformatic analyses

The largest cDNA inserts were selected for sequencing by using the T7SelectUP primer. To determine the identity and correct reading frame, a portion of the insert sequences were submitted to the *P. falciparum* genome database (PlasmoDB version 4.1) (A-2.7). The BLAST algorithm was used to search for homologous sequences in the parasite's genome. Sequence identity with the database was not 100% due to unclear sequencing data, but revealed significant homology (98-100%) to matching sequences. The letter N, which represents all four nucleotides, was inserted into the sequence at positions on the gel where discrete bands could not be discerned. Annotated genes were chosen based on the best match or E value between query and database sequences. The E value is a measure of statistical significance and indicates the probability of a query sequence being present in the genome. The best match from nucleotide (BLAST-N) and protein (BLAST-X) searches provided further confirmation with which the query sequence could be allocated to an annotated gene. E values for BLASTN searches ranged from 9e-7 to 1.4e-15, while BLASTX values ranged from 1.6e-6 to 3e-13. The correct reading frame of the query sequence was determined by comparing the initial codon to the translational frame of the predicted gene.

Sixty *P. falciparum* cDNA inserts were analysed, of which 23 contained the correct translational frame. As expected, this represented approximately one third of the total number of fragments sequenced, and was based on a one in three chance of cDNA synthesis ending at the correct nucleotide during reverse transcription. The remaining two thirds of the nucleotide sequences were out-of-frame and encoded spectrin/4.1R and 4.1R interacting peptides that did not exist in the parasite. For this reason, they were not selected for further study. Several inserts with the correct open reading frame were found more than once and only ten represented unique sequences. This was attributed to the significant degree of enrichment of these clones during biopanning.

Spectrin/ 4.1R specific clones

Of the ten unique clones, four clones interacted specifically with spectrin and/or 4.1R (Lauterbach et al., 2003). Genes of known or predicted function included MAL13P1.278 (serine/threonine kinase), PFI1570c (putative aminopeptidase) and PFA0125c (EBA-181/JESEBL). PF14_0201 was identified as a predicted gene, as it encodes a hypothetical protein that displays no known protein homology. These annotated proteins contain an amino acid sequence that interacts with either spectrin or 4.1R, but does not rule out the possibility of binding to both proteins. To refine the interaction and identify clones that bind exclusively to either of the two major skeletal proteins, the smaller globular protein, 4.1R, was selected for biopanning against the initial *P. falciparum* phage display library.

4.1R specific clones

Eight clones contained *P. falciparum* sequences that encoded peptides specific for purified 4.1R (Lanzillotti and Coetzer, 2004). Two of these clones were previously identified in the library that was biopanned against spectrin/4.1R. The open reading frames of the eight inserts are shown in figure 20. The letter N (blue) was inserted into the sequence at positions on the gel where discrete bands could not be discerned. Sequences in clones 2 and 4-7 contained a reading frame beginning with three codons supplied by the *Eco*R I linker (red nucleotides in figure: **I**, AAT; **II**, TCA and **III**, AGC). Nucleotides in orange depict T7Select vector sequence. Clones 1, 3 and 8 contained cDNA inserts with an internal *Eco*RI restriction site, which had been digested by the enzyme during cDNA end-modification. Blunt-ending of parasite cDNA had been performed with unmethylated dNTPs and probably accounted for this finding. As a result, codons **I**-**III** were not present in the inserts and were thus excluded from the vector constructs.

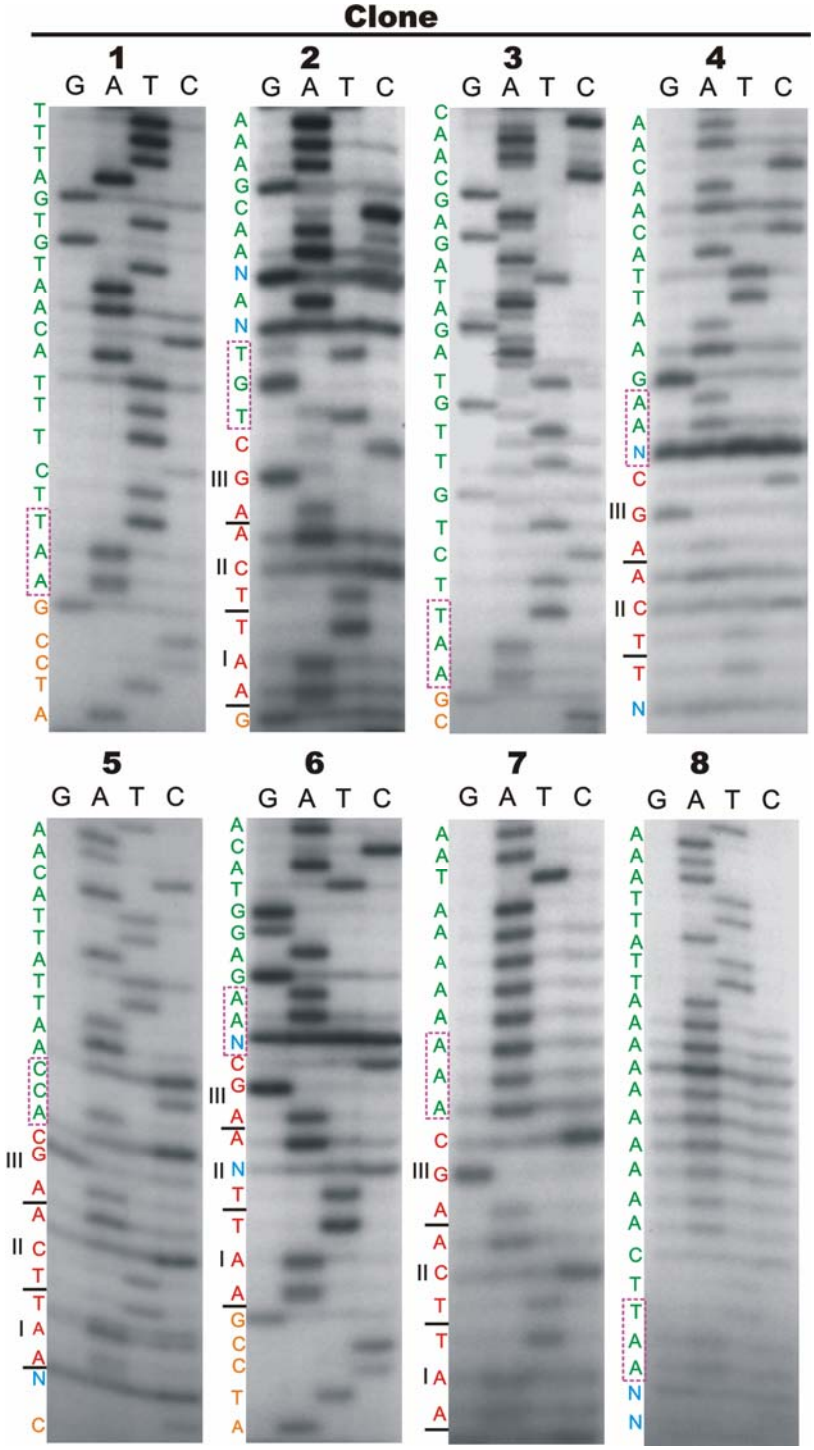

Figure 20- Sequence analysis of virions biopanned against 4.1R

Eight unique *P. falciparum* open reading frames were identified by sequencing the parasite cDNA inserts. Nucleotides in red indicate *Eco*R I linker sequence and comprise three codons (**I-III**), while orange bases represent the T7Select vector. Nucleotides in green represent *P. falciparum* sequences and the boxed (purple) region the first parasite-derived codon. A, adenine; G, guanine; C, cytosine; T, thymine and N, A/G/C or T.

Transcripts with predicted and known functions included PFB0150c (clone 1: putative protein kinase) and two members of the erythrocyte protein binding family, PF07_0128 (clone 2: EBA-175) and PFA0125c (clone 3: EBA-181/JESEBL). BLAST searches also revealed genes encoding five hypothetical proteins including PF14_0201 (clone 4), PFE0570w (clone 5), PFA0420w (clone 6), PF11_0191 (clone 7), and MAL6P1.48 (clone 8). The fact that clones 3 and 4 were also enriched in the library that was biopanned against spectrin/4.1R indicates that EBA-181 and PF14_0201 have high binding specificity toward 4.1R.

3.3.3 Annotated *P. falciparum* proteins

Gene identities and corresponding annotated proteins of the ten unique parasite sequences are indicated in table 7. The position of the 4.1R/spectrin binding peptide and predicted motif in each protein is also presented in the table.

Erythrocyte binding proteins: EBA-175 and EBA-181

EBA-175 (PF07_0128) and EBA-181 (PFA0125c), as described in the PlasmoDB database version 4.1, consist of 1462 and 1567 amino acids respectively. The genes are predominantly expressed in the late schizonts and merozoite stages of the asexual parasite (Figure 21). Profiles

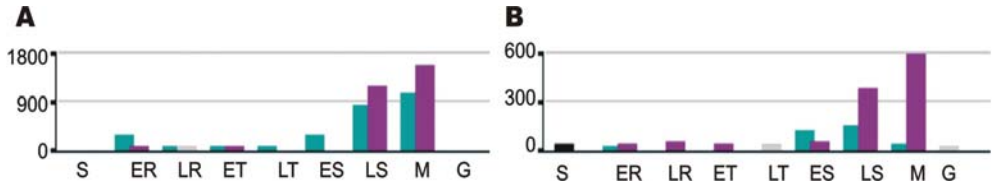

Figure 21- mRNA expression profiles of EBA-175 and EBA-181

Absolute expression data for EBA-175 (**A**) and EBA-181 mRNA (**B**) were obtained from the PlasmoDB database (version 4.1) (http://www.plasmodb.org). Expression is maximal during the late schizont and merozoite stages of the asexual parasite. The green and purple bars indicate *P. falciparum* 3D7 cultures synchronised by sorbitol and temperature respectively. Abbreviations: **S**, sporozoites; **ER**, early rings; **LR**, late rings; **ET**, early trophozoites; **LT**, late trophozoites; **ES**, early schizonts; **LS**, late schizonts; **M**, merozoites and **G**, gametocytes (Le Roch et al., 2003).

Table 7- *P. falciparum* annotated proteins containing binding sequences specific for 4.1R and spectrin

Clone	Gene locus	Gene product	E value: BLASTN/X	Binding sequence[1]	Predicted motif
1	PFB0150c	Putative protein kinase	1.9e-14/ 5.3e-10	26-55	Unknown
2	PF07_0128	EBA-175	3.1e-13/ 1.8e-10	1349-78	Duffy antigen
3	PFA0125c	EBA-181	3.2e-14/ 1.2e-10	971-1000	Myosin-like
4	PF14_0201	Hypothetical protein	7.7e-11/ 1.6e-6	801-60	Myosin-like
5	PFE0570w	Hypothetical protein	1.9e-14/ 1.3e-10	8611-710	Unknown
6	PFA0420w	Hypothetical protein	3.7e-10/ 1.4e-8	100-32	Neuro-filament-like
7	PF11_0191	Hypothetical protein	1.4e-15/ 3e-13	459-88	Myosin-like
8	MAL6P1.48	Hypothetical protein	1.4e-14/ 3.4e-11	2675-704	Unknown
9	MAL13P1.278	Ser/Thr protein kinase	7.6e-10/ 6e-6	786-815	Unknown
10	PFI1570c	Putative aminopeptidase	9e-7/ 9.6e-6	216-45	Ligase I Ubiquitin-protein

Ten unique gene transcripts from asexual parasite stages were identified by BLASTN sequence similarity searches in PlasmoDB, version 4.1. The identities of the genes represent the best match, which correlated with BLASTX analyses. Clones 1-8 are specific for 4.1R, whereas clones 9 and 10 interact with 4.1R and/or spectrin. The 4.1R binding sequence in clones 3 and 4 may also contain parasite amino acids specific for spectrin (A-2.7). [1] Amino acid residues of spectrin/4.1R and 4.1R binding sequences. The C-terminal amino acid position of the interacting region was estimated from the size of the parasite cDNA insert. Abbreviations: EBA, erythrocyte binding antigen; Ser, serine and Thr, threonine.

revealed that EBA-175 RNA is expressed in significantly larger quantities than EBA-181 RNA. No expression has been recorded in either gametocytes or sporozoites. Mass spectrometry data for EBA-175 shows evidence for peptide fragments in trophozoites and merozoites. The data for EBA-181 indicate fragments in merozoites as well as sporozoites. Both proteins have a predicted signal sequence and putative transmembrane

domains at their N and C-terminal ends respectively. Sequences encoding the DBL erythrocyte binding domains and C-terminal cys-rich domain are located in a large exon, whereas the C-terminal transmembrane region and cytoplasmic tail are encoded by at least two significantly smaller exons. The 4.1R binding sequence in EBA-175 is located in the C-terminal cys-rich region, whereas the spectrin/4.1R sequence in EBA-181 is located in a highly charged interdomain region. The interacting sequence in both proteins has a net negative charge. Each peptide contains six acidic amino acids (glutamic acid and aspartic acid), as well as five and two basic amino acids (arginine and lysine) respectively (A-2.7). Secondary structure predictions in PlasmoDB indicate that the cys-rich regions contain a number of α-helical motifs. The hydrophilic interdomain regions contain low complexity sequences and assume a coiled structure.

Annotated enzymes

PFB0150c has a two-exon gene structure, which encodes a putative protein kinase of 2 987 amino acids. It consists of a protein kinase domain (amino acids 2 624-854), two predicted transmembrane regions and short stretches of low complexity sequence. The protein kinase domain is predicted to contain an abundance of α-helical structures. α-Helix, β-sheet and coiled structures are interspersed throughout the entire protein. mRNA expression is maximal in sporozoites and gametocytes (Figure 22A), and correlates with mass spectrometry data. Lower levels of expression are also evident in late asexual development stages. The 4.1R binding sequence is located 26 amino acids from the N-terminal methionine, and is found adjacent to a low complexity region containing tandem repeats of DNKNVIKS. The 4.1R specific sequence has no net charge, but comprises 47% charged amino acids (A-2.7).

MAL13P1.278 is a single exon in excess of 12kbp. The annotated serine/threonine kinase contains six predicted amidation sites, a protein kinase active site and ATP binding site, as well as an abundance of low complexity sequence that spans the entire protein. The region (1 282-409) encompassing the functional domain contains amino acids which conform to α-helical and β-sheet motifs, and is relatively hydrophobic compared to the full-length protein. mRNA encoding this kinase is expressed in late schizonts, merozoites and gametocytes (Figure 22B). However, mass spectrometry data on PlasmoDB showed evidence for a MAL13P1.278 peptide fragment in sporozoites and trophozoites. The protein kinase contains 4 044 amino acids. The spectrin/4.1R binding sequence is situated 500 amino acids N-terminally of the ATP binding site, which is located at amino acid position 1282-306. The interacting sequence has a net positive charge and incorporates a low complexity region which is rich in lysine.

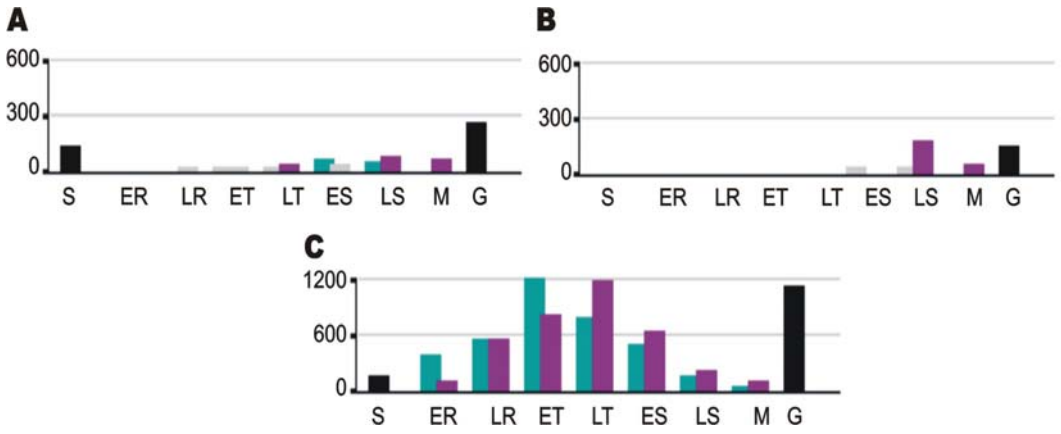

Figure 22- mRNA expression profiles of annotated *P. falciparum* enzymes

Absolute expression data for PFB0150c (**A**), MAL13P1.278 (**B**) and PFI1570c (**C**) were obtained from the PlasmoDB database version 4.1 (http://www.plasmodb.org). Expression of PFB0150c mRNA is evident at very low levels in gametocytes (**G**), merozoites (**M**) late schizonts (**LS**), early schizonts (**ES**), late trophozoites (**LT**) and sporozoites (**S**). Similarly, low levels of MAL13P1.278 transcript are found in **G**, **M** and **LS**. Expression of PFI1570c mRNA is maximal during the **G**, **LT** and early trophozoite (**ET**) stage, with varying mRNA levels also present in other asexual parasite stages (Le Roch et al., 2003).

PFI1570c is a putative zinc-metallo aminopeptidase which is encoded by a single exon approximately 1.7kbp in length. The protein bears sequence similarity to human aspartyl aminopeptidase and contains two large active domains at the C terminal end as well as two predicted transmembrane regions. The spectrin/4.1R interacting sequence has a net negative charge and is located adjacent to an active site at amino acid position 216-245. Secondary structure predictions indicate that the functional domains are rich in α-helix and β-sheet motifs, and are interlinked by a low complexity coiled sequence of approximately 30 amino acids. mRNA expression predominates in trophozoites and gametocytes (Figure 22C), and is higher in all stages when compared to the putative protein kinases. Mass spectrometry data indicate that the aminopeptidase is present in gametocytes and asexual parasite stages. Interestingly, four PFI1570c fragments of the protein have been localised at the infected erythrocyte membrane. Together with a spectrin/4.1R recognition sequence, this finding strongly suggests that PFI1570c associates with the erythrocyte skeleton during asexual parasitic development.

Hypothetical proteins

Specific features of the annotated hypothetical proteins are summarised in table 8. The proteins contain either large areas or short stretches of low complexity sequences. According to the respective hydropathy plots, these sequences are characterised by hydrophilic regions and contain a predominance of negatively (glutamic acid and aspartic acid) and positively (lysine and arginine) charged amino acids. In addition, many contain stretches of homopolymeric residues, most commonly asparagine. Secondary structure predictions indicate that low complexity sequences assume a coiled structure, with low probability of helical and β-sheet conformations (A-2.8). Generally, the amino acid composition of the 4.1R binding sequences consists mainly of charged residues at physiological pH. Sequences encoded by PF11_0191 and MAL6P1.48 consist of 30% and 18% lysine, respectively, and have net positive charges. PF14_0201

and PFA0420w consist largely of acidic amino acids and have net negative charges, whereas the 4.1R binding sequence in PFE0570w has no net charge (A-2.7).

Table 8- Hypothetical *P. falciparum* proteins binding to 4.1R

Gene locus	Number of exons	Protein size[1]	Expression profile[2]	Proteomics data[3]	Domains[4]
PF14_0201	2	969	LR;M (♦;G;S)	S; T	SS;TM
PFE0570w	1	10 061	♦;S	S; T; M	DED; MATH; TM
PFA0420w	1	179	M (♦;S;G)	M	ELM2
PF11_0191	1	594	S (♦;G)	G; T; M; I	snoRNA
MAL6P1.48	4	2 874	LT;M (♦;G;S)	S; G; T; M	TM

Annotated protein data were obtained from the *P. falciparum* genome database, PlasmoDB version 4.1 (http://www.plasmodb.org). [1] Protein size refers to number of amino acids. [2] Parasite developmental stages that feature maximal levels of mRNA expression are shown. The letters in parenthesis indicate the *P. falciparum* stages in which expression levels are substantially lower. [3] Mass spectrometry data have evidenced peptide fragments from the annotated proteins in different parasite developmental stages. Correlation between the proteomics data and mRNA expression profile is exhibited in at least one life-cycle stage. [4] Predicted domains present in each hypothetical protein. Abbreviations: **bp**, base pairs; **DED**, death effector domain; **ELM2**, domain of unknown function; **G**, gametocytes; **I**, infected erythrocyte membrane; **kbp**, kilobase pairs; **LR**, late rings; **LT**, late trophozoites; **M**, merozoites; **MATH**, Meprin And TRAF-Homology domain; **S**, sporozoites; **snoRNA**, small nucleolar ribonucleic acid; **SS**, signal sequence; **T**, trophozoites; **TM**, transmembrane domain and ♦, early rings, LR, early trophozoites, LT, early schizonts and late schizonts.

PF14_0201 is a protein of 969 amino acids, which comprises a predicted signal sequence and two transmembrane domains located at the C and N-terminal ends. The spectrin/4.1R binding peptide is positioned in a low complexity sequence at amino acid position 801-60.

MAL6P1.48 and PFE0570w are much larger molecules and also contain predicted transmembrane domains. The 4.1R binding peptide is located near the C-terminal end of both proteins. PFE0570w has additional motifs including a zinc binding region (zinc-finger), death effector domain (DED),

Meprin And TRAF-homology (MATH) domain and two RNA pseudouridylate synthase regions.

PFA0420w contains a domain of unknown function (ELM2) at the N-terminal end. The full-length protein is composed of 70% low complexity sequence and reveals a strong coiled secondary structure prediction. The 4.1 binding peptide contains four tandem repeats of EEV/AQG (A-2.7). This is followed by a series of PANDE and PANEE repeats that make up the rest of the protein (A-2.8).

PF11_0191 has a putative small nucleolar RNA binding domain of 212 amino acids (position: 265-413) and contains a 4.1R binding peptide situated near the C-terminal end of the protein. Mass spectrometry data revealed a PF11_0191 peptide fragment of 33 amino acids (position: 248-81) at the infected erythrocyte membrane.

3.4 Identification of a common filament repeat motif

Bioinformatic analyses in the PlasmoDB database (version 4.1) revealed a common, negatively charged binding motif in four of the annotated *P. falciparum* proteins (Table 9). The 4.1 binding sequences were identified in regions that showed homology to coiled filament heptad repeat domains; PF14_0201, PF11_0191 and PFA0125c (EBA-181) mapped to the coiled-coil domain of eukaryotic myosin, whereas the binding sequence in PFA0420w revealed homology to the coiled repeat pattern in eukaryotic neurofilament sequences. Filament-like proteins such as keratins and myosin represent well-characterised groups of coiled-coil proteins. Coiled-coil domains consist of two or more alpha helices that intertwine around each other to form a supercoil, and are often characterised by amino acid sequences which form a heptad repeat pattern (Burkhard et al., 2001; Lupas, 1996). Different computational methods were used to predict the secondary structure of the parasite motifs, including the presence of

Table 9- Common 4.1R binding motifs

Gene locus	Predicted motif	E value	Amino acid position	Negative charge (%)[1]	Coil (%)[2]
PFA0125c (EBA-181)	Myosin-like	1.0e-23	716-1235	25	67
PF14_0201	Myosin-like	1.5e-28	405-944	26	49
PFA0420w	Neurofilament-like	4.9e-10	38-165	35	87
PF11_0191	Myosin-like	1.4e-14	400-594	33	64

4.1R binding sequences were localised to a common motif in four parasite proteins (PlasmoDB version 4.1: http://www.plasmodb.org). [1] The EXPASY proteomics server (http://www. expasy.org) was used to determine the percentage of negatively charged amino acids (glutamic acid and aspartic acid). [2] SSPro software was used for predicting secondary structure of the 4.1R binding domains (A-2.8). The algorithm revealed that a large percentage the *P. falciparum* myosin and neurofilament-like motifs was organized into stretches of coiled regions.

coiled-coil domains. Secondary structure predictions were obtained using SSPro software available from the Institute of Genomics and Bioinformatics (http://www.igb.uci.edu/servers/psss.html). Predictions indicated that at least 50% of the amino acids in these sequences contributed to coil formation (A-2.8). The parasite motifs that showed similarity to myosin eukaryotic sequences comprised two coiled-coil domains. The existence of these domains was confirmed (figure 23) using the Simple Modular Architecture Research Tool (SMART: http://smart.embl-heidelberg.de/), and verified in all cases by the MultiCoil algorithm (http://multicoil.lcs.mit.edu/cgi-in/multicoil).

The position of the 4.1R binding sequences was similar in each myosin-like motif, based on their location relative to a predicted coiled-coil domain. The sequence in PFA0125c (EBA-181) was positioned in close proximity to a coiled-coil motif. The 4.1R specific peptide in PF14_0201 was located within a coiled-coil region, while the relevant amino acids in PF11_0191

extended partially into a coiled-coil domain. Despite the common architectural features and charge properties, ClustalW analysis did not reveal any significant alignment between the coiled-coil motifs nor the 4.1R binding sequences.

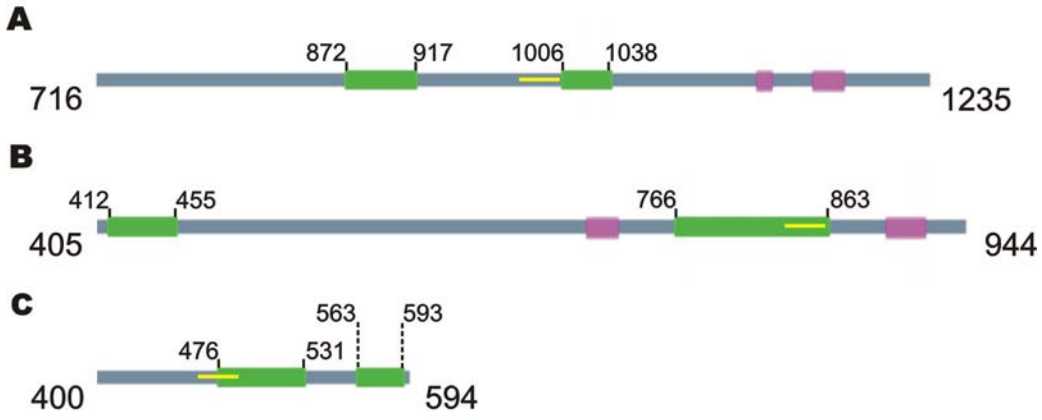

Figure 23- Coiled-coil domains in 4.1R specific *P. falciparum* myosin-like motifs

SMART and MultiCoil analyses software were used for predicting the locations of coiled-coil domains (green boxes) in PFA0125c (**A**), PF14_0201 (**B**) and PF11_0191 (**C**). The purple boxes depict low complexity regions, whereas the yellow line shows the position of the 4.1R specific amino acid sequence.

3.5 Cloning of *P. falciparum* sequences binding to 4.1R

PFA0420w and the parasite myosin-like motifs were selected for cloning into the pET15b expression vector, whereas EBA-175 was subcloned in the pGEX-4T-2 vector due to the presence of an *Nde* I site. Regions encompassing the different 4.1R-binding sequences were amplified from purified *P. falciparum* FCR3 genomic DNA using appropriately designed primers (A-2.9). The amplified products are illustrated in figure 24.

After ligation with the vector (2.7.4), DH5α cells were transformed (2.7.5) with the different vector constructs and the presence of inserts screened via PCR (2.7.6) (Figure 25). Following insert verification, plasmid DNA was submitted for automated sequencing to ensure the correct reading frame

and that no errors had been introduced during PCR amplification. ChromasPro (version 1.2) software was used to analyze the results, and indicated that all sequences were correct and in frame.

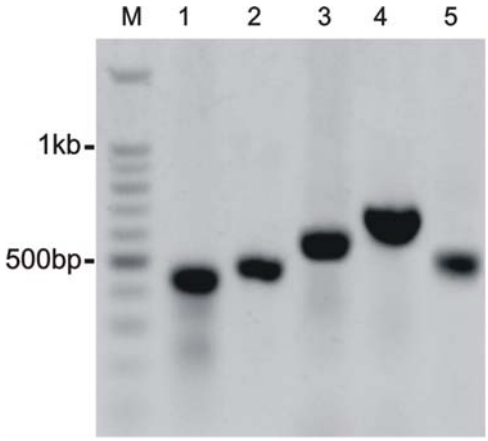

Figure 24- PCR amplification of parasite 4.1R binding sequences

P. falciparum sequences (**1-5**) were amplified and electrophoresed alongside a DNA size marker (**M**). 10μl PCR products were resolved on a 1% agarose gel. **1**, PF07_0128 (EBA-175) (449bp); **2**, PFA0125c (EBA-181) (486bp); **3**, PF11_0191 (570bp); **4**, PF14_0201 (696bp) and **5**, PFA0420w (492bp)

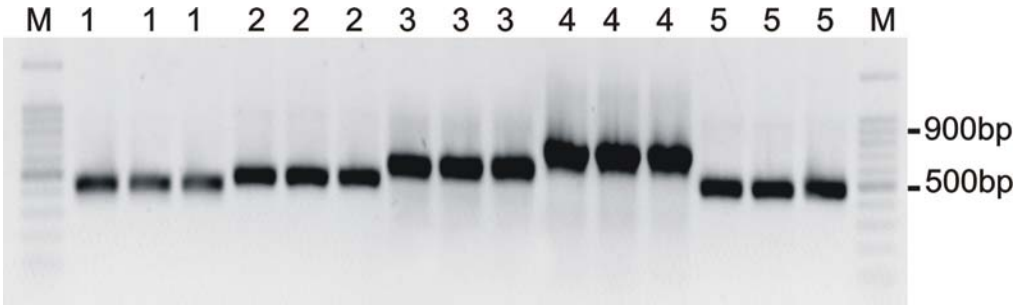

Figure 25- Verification of the presence of *P. falciparum* inserts

Plasmid DNA was purified from different colonies of transformed DH5α cells and used in a PCR with gene specific primers. The reaction generated products of the expected insert sizes. **M**, 100bp DNA ladder; **1**, PF07_0128 (EBA-175); **2**, PFA0125c (EBA-181); **3**, PF11_0191; **4**, PF14_0201 and **5**, PFA0420w.

Unexpectedly, six nucleotides were included at the 5' end of the PFA0125c (EBA-181) sequence at positions 135-140 of the chromatogram (Figure 26). The addition of TGC (cysteine) and ATG (methionine) did not

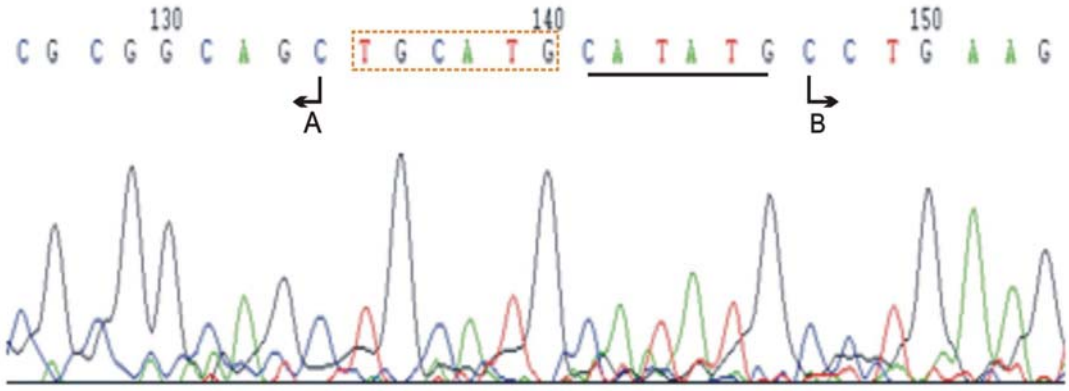

Figure 26- Partial DNA sequence of 6His-EBA-181

The chromatogram depicts the partial sequence of 6His-EBA-181 in pET15b vector. Arrow **A** indicates vector sequence, whereas arrow **B** shows *P. falciparum* sequence with CCT as the initial codon. The underlined nucleotides represent an *Nde* I recognition sequence which was regenerated after ligation of insert and vector. TGC and ATG (boxed region) are additional codons that should have been excluded from the sequence.

result in a frame-shift. These nucleotides were provided by additional bases (CATGCATG) that were appended at the 5' end of the forward primer used for amplifying the EBA-181 region. The similarity of this octet to the *Nde* I endonuclease recognition sequence (CATATG), presumably resulted in enzyme digestion within the additional bases rather than the intended sequence. However, star activity has not been reported for *Nde* I. The extra cysteine residue represented the only thiol-containing amino acid in the EBA-181 sequence, thus foreclosing any chance of internal disulphide bond formation. The construct was therefore used for further study.

3.6 Expression and purification of recombinant parasite proteins

Two parasite proteins were selected for recombinant protein expression and purification. The selection was based largely on the anticipated significance of the corresponding peptide's interaction with 4.1R and to a lesser extent on the predicted solubility of the recombinant protein (Table 10). Due to the high percentage of charged amino acids in their protein sequences, PFA0125c (EBA-181), PF11_0191, PF14_0201 and

PFA0420w exhibited significant predicted solubility when over-expressed in *E. coli*. Upon induction, PF07_0128 (EBA-175) had a 73% chance of being expressed as insoluble aggregates in *E. coli* inclusion bodies. However, the predicted solubility increased to 36% when expressed in conjunction with GST. Based on the potential importance of the interaction between *P. falciparum* invasion proteins and 4.1R, EBA-181 (PFA0125c) and EBA-175 (PF07_0128) were selected for further study. Despite their high chance of solubility, the 4.1R binding properties of the hypothetical proteins were not further investigated in this study.

Table 10- Solubility profiles of recombinant *P. falciparum* proteins

Parasite protein (gene locus)	Vector tag	Annotated protein function	% solubility (-tag)	% solubility (+ tag)
EBA-175	GST	Merozoite invasion	27	36
EBA-181	His	Merozoite invasion	93	81
PF11_0191	His	Hypothetical	96	89
PF14_0201	His	Hypothetical	97	97
PFA0420w	His	Hypothetical	97	97

A region including the 4.1R binding sequence in EBA-175 (PF07_0128) was cloned in frame with glutathione S-transferase (**GST**) encoding sequences in the pGEX-4T-2 vector. The other proteins were cloned downstream of a histidine (**His**) tag provided by the pET15b vector. Protein solubility was determined using a statistical model available from the School of Chemical Engineering and Materials Science at the University of Oklahoma (http://www. biotech.ou.edu/)

E. coli BL21 CodonPlus® competent cells were transformed with EBA-181 and EBA-175 vector constructs (2.8.1). Optimal expression of the 4.1R specific EBA-181 region (denoted *Pf*J; Figure 27A) was obtained by induction with 1mM IPTG at 37^{o}C for 2-3hrs (2.8.2). The different bacterial fractions indicated that *Pf*J was expressed predominantly as a soluble protein, as evidenced by SDS-PAGE analysis (Figure 27B). The gel also indicated that the recombinant protein migrated as an apparent dimer (~50kDa) under reducing conditions. This was most likely attributed to the

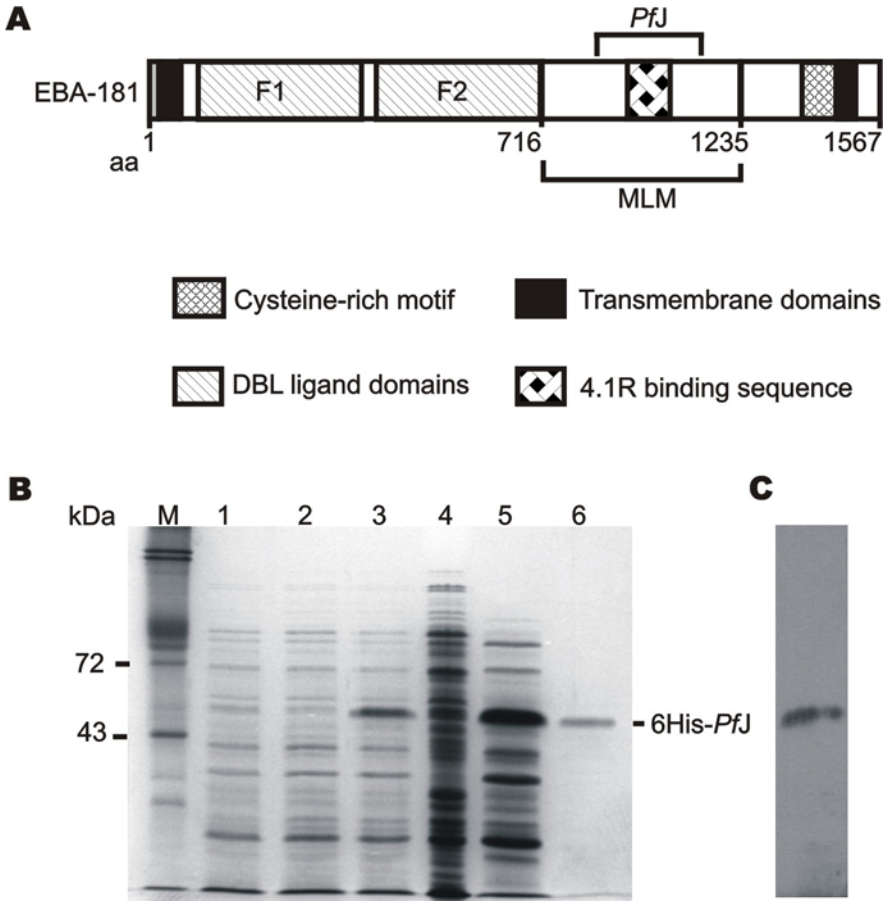

Figure 27- Expression and purification of 6His-*Pf*J

A) Schematic representation of the full length *P. falciparum* invasion protein EBA-181/JESEBL. Amino acid positions underneath the schematic demarcate the various domains in the protein. The molecule comprises two Duffy binding-like (DBL) domains (denoted F1 and F2), as well as two transmembrane regions and a C-terminal cysteine-rich motif. The 4.1R binding sequence is located in a myosin-like motif (MLM), and the relative position of the EBA-181 fragment (*Pf*J) used for cloning and expression is shown. **B**) 6His-*Pf*J was expressed and purified from BL21 CodonPlus® *E. coli* cells. Aliquots from different stages of the purification procedure were resolved on a 12% Laemmli gel: erythrocyte membrane marker (**M**); induced vector control (**1**); uninduced vector construct (**2**); induced vector construct (**3**); insoluble *E. coli* protein fraction from 3 (**4**); soluble *E. coli* protein fraction from 3 (**5**) and purified *Pf*J (**6**). **C**) Immunoblot to confirm the expression of 6His-*Pf*J using mouse anti-His antibody and chemiluminescent substrate.

high proportion of charged amino acids in *Pf*J, which presumably resulted in atypical binding of SDS molecules leading to aberrant and unpredictable migration patterns on denaturing polyacrylamide gels (Cooke et al., 2004b). 6His-*Pf*J was purified from soluble *E. coli* extracts using nickel coated magnetic beads (2.8.3), and approximately 0.8μg of protein was

obtained from 1ml culture volume. An immunoblot was performed (2.8.4) to verify the presence of the His tag (Figure 27C).

EBA-175 transformed cells were initially induced with 1mM IPTG for 2-3hrs at 37^{o}C. The GST-tagged *P. falciparum* protein of approximately 42kDa was expressed in *E. coli* inclusion bodies, as determined by SDS-PAGE analyses of insoluble bacterial fractions. This correlated with its solubility profile presented in Table 10, which indicated that the recombinant protein had a 64% chance of being insoluble when over-expressed in *E. coli* (http://www.biotech.ou.edu/). In an effort to increase the protein's solubility, culture conditions were varied. Induction was carried out with lower concentrations of IPTG (0.1-0.4mM). 250ml cultures were also induced at room temperature for a period of 16hrs. These changes aid in reducing the rate of recombinant protein production, thus allowing more time for the newly synthesised proteins to fold correctly. However, the altered conditions had no effect on solubility as most of the parasite protein was sequestered in the insoluble bacterial fraction (Figure 28). No further attempts were made to increase the expression of soluble GST-tagged protein.

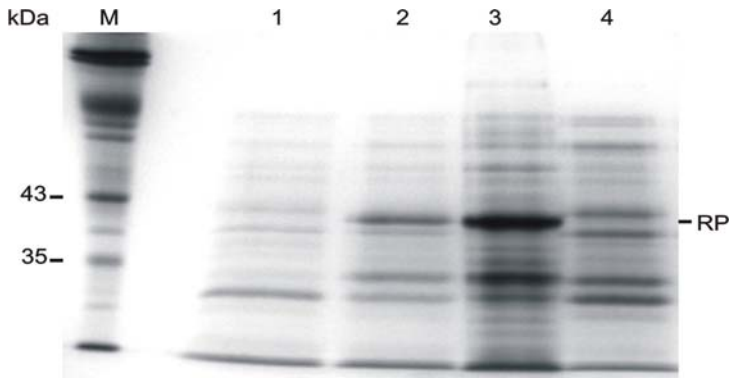

Figure 28- Expression of a GST-tagged 4.1R binding peptide from EBA-175

Cultures were induced with 0.1mM IPTG at room temperature for 16hrs. Bacterial fractions were analysed on a 12% Laemmli gel. The GST-tagged recombinant protein (**RP**) is approximately 42kDa in size and is mostly evident in the insoluble fraction. **M**, erythrocyte membrane marker; lane **1**, uninduced vector construct; lane **2**, induced vector construct; lane **3**, insoluble *E. coli* protein fraction from 2 and lane **4**, soluble *E. coli* protein fraction from 2.

3.7 Cloning of 4.1R domain sequences

Human reticulocyte RNA was purified from venous blood (2.9.1). 0.5-1μg total RNA was obtained per 20ml preparation. The quality and purity of the RNA were assessed by agarose gel electrophoresis (Figure 29) and spectrophotometry respectively. Relative to a MassRuler^{TM} double strand DNA ladder, the 28S and 18S ribosomal subunits migrated close to 2000bp and 700bp. The intensity of the 28S band was approximately twice that of the 18S band, indicating good quality RNA. mRNA was evident as a faint smear between the 300bp and 2kbp DNA marker (not visible in the photograph). Spectrophotometric analyses of the total RNA revealed A_{260}: A_{280} ratios ranging from 1.8-2.0, which corresponds to pure preparations.

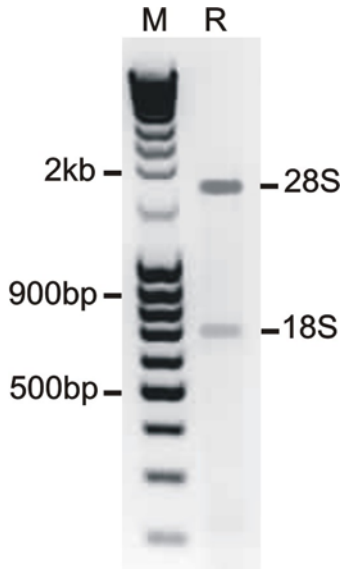

Figure 29- Analysis of human reticulocyte RNA

Purified RNA from human reticulocytes (**R**) was electrophoresed alongside a MassRuler^{TM} DNA ladder (**M**) on a 1% agarose gel. The 28S and 18S ribosomal RNA subunits are indicated.

The 30, 16, 10 and 22kDa domain sequences of 4.1R were reverse transcribed from reticulocyte RNA using gene specific primers (2.9.2; A-2.6) (Figure 30). The 16, 22 and 30kDa PCR products contained low molecular weight products, presumably primer dimers which would have

generated false positives after subcloning in the pGEX-4T-2 vector. For this reason, the relevant fragments were purified from agarose gels (2.9.3). After subcloning, DH5α cells were transformed with the different domain constructs (2.10.1). Single colonies were selected and plasmid DNA screened for the presence of inserts (Figure 31).

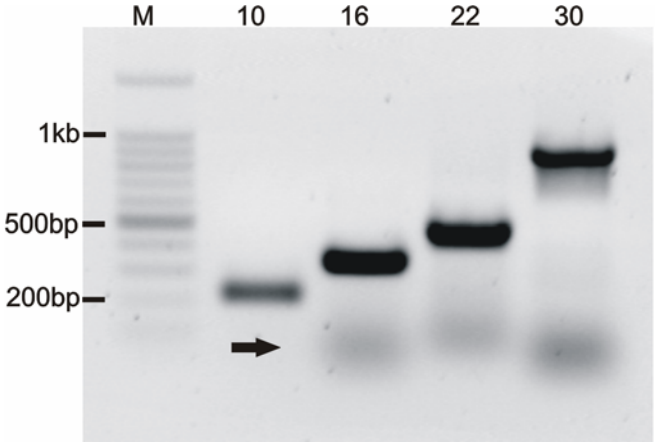

Figure 30- RT-PCR of 4.1R domains

A 100bp DNA ladder (**M**) and 10μl PCR products of 10, 16, 22 and 30kDa 4.1R domains were electrophoresed on a 1% agarose gel. The arrow indicates the presence of low molecular weight products, presumably primer dimers. The sizes of the different 4.1R DNA fragments are: **10**, 224bp; **16**, 344bp; **22**, 479bp and **30**, 914bp.

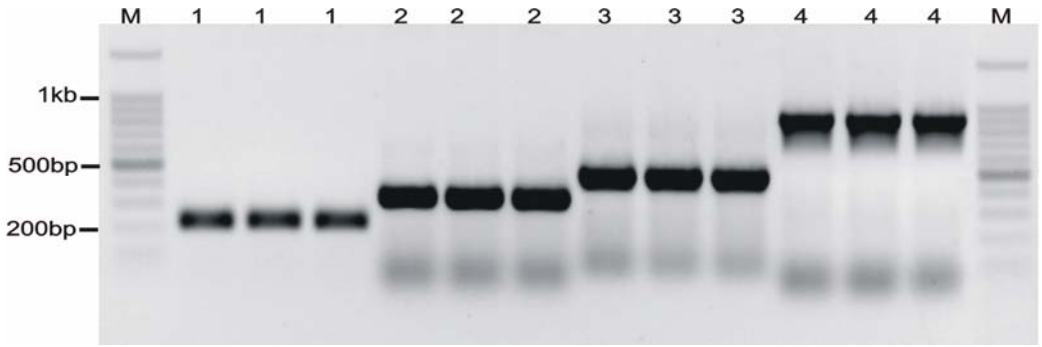

Figure 31- Confirmation of 4.1R domain inserts

Plasmid DNA was extracted from colonies of transformed DH5α cells and then used in a PCR with 4.1R domain primers. The reactions generated products of the expected domain sizes. **M**, 100 base pair DNA ladder; **1**, 10kDa; **2**, 16kDa; **3**, 22kDa and **4**, 30kDa encoding cDNA fragment.

Automated DNA sequencing indicated that the pGEX-4T-2/4.1R domain constructs were in-frame and error-free. The exon map for the corresponding 4.1R structural domain is presented in figure 32.

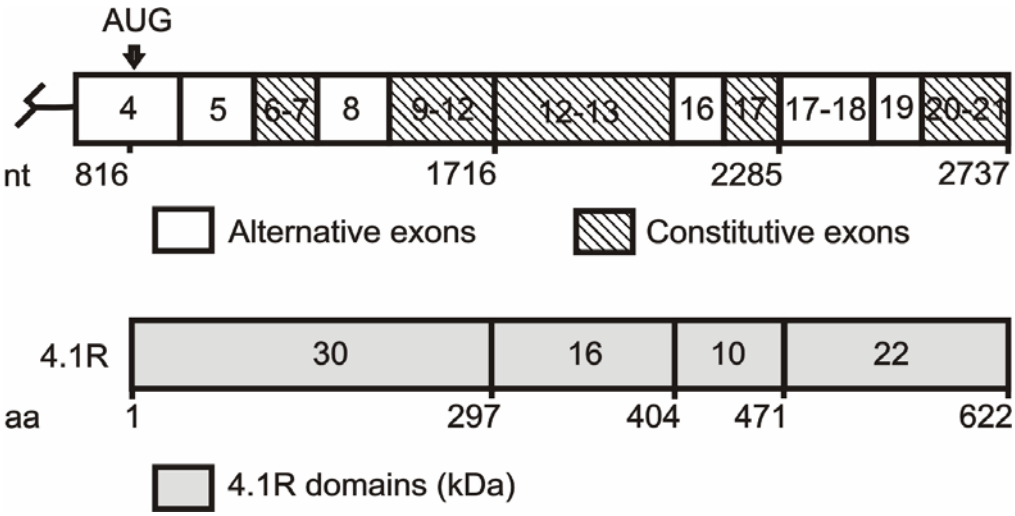

Figure 32- Exon map of 4.1R cDNA

Schematic representation of 4.1R cDNA used in this study. The erythrocyte translation initiation site is indicated at nucleotide (**nt**) position 816 in exon 4 with the coding sequence extending to position 2737. Positions of alternative and constitutive exons are shown. The 30/16 and 10/22 exon boundaries overlap in exons 12 and 17 respectively, and 4.1R primers were designed accordingly. The corresponding 80kDa protein is depicted below the cDNA. Amino acid (**aa**) residue numbers depict the relative locations of the four 4.1R structural domains. The 30kDa domain (aa: 1-297) corresponds to nt 816-1716, whereas the 16kDa and 10kDa domains (aa: 298-471) corresponds to nt 1717-2285. The 22kDa domain (aa: 472-622) is encoded by nt 2286-2737 (Conboy et al., 1991; Conboy, 1993; Walensky et al., 2003; Waller et al., 2003).

3.8 Expression and purification of GST-4.1R domains

Rosetta^{TM} 2 (DE3) cells were transformed (2.10.1) with the different domain constructs and 1ml culture volumes induced with 1mM IPTG at 37^{o}C for 2-3hrs. The cells were lysed and the bacterial fractions analysed with SDS-PAGE (Figure 33). Under these conditions, the expressed 10kDa, 16kDa and 22kDa recombinant proteins were soluble, while the GST-tagged 30kDa protein was completely insoluble. In an attempt to increase the yield of soluble GST-30kDa, parameters relating to IPTG concentration, induction temperature and cellular densities were varied.

However, these changes had no marked effect on the expression profile. Some protein was expressed in the soluble fraction, but the concentration was too low for purification and binding studies.

The use of the Overnight Express™ Autoinduction system significantly improved the yield of soluble GST-30kDa. Twenty five ml cultures were induced at 37°C for 16hrs and the cells lysed in 500µl BugBuster®.

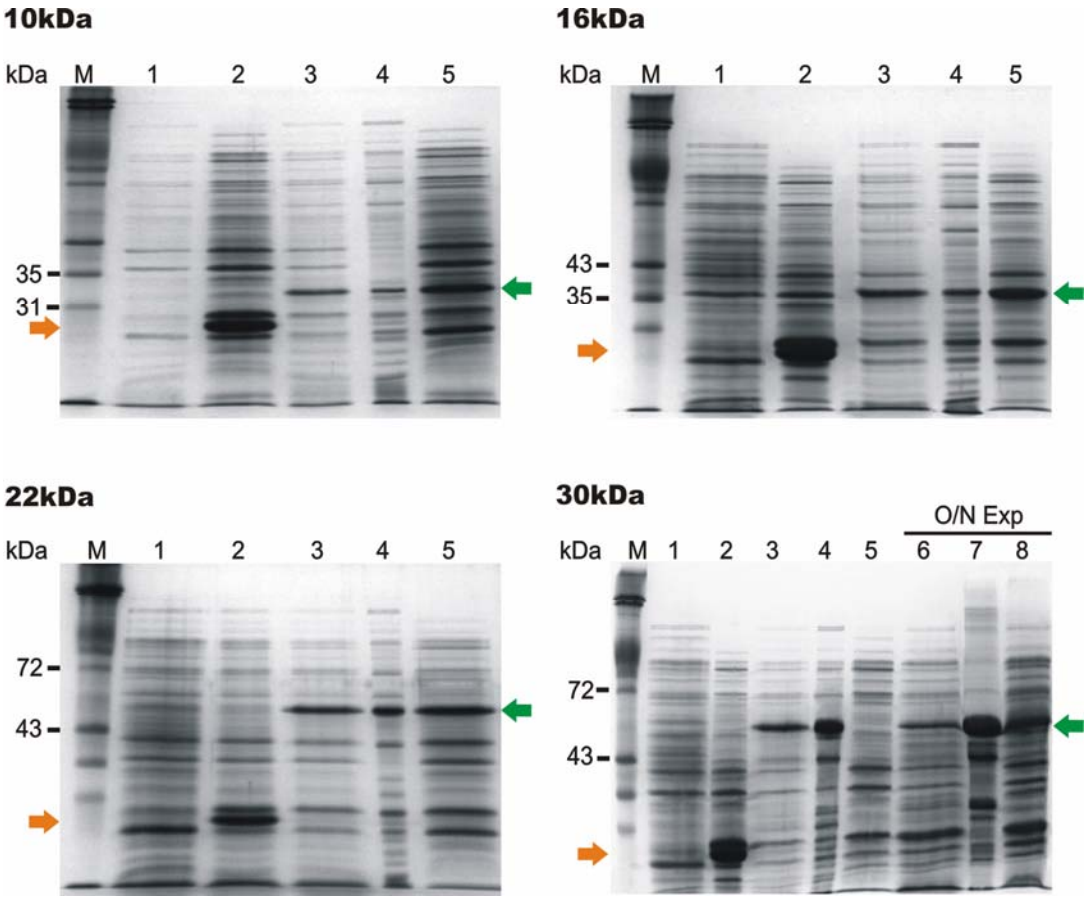

Figure 33- Expression of GST-4.1R domains

Aliquots of solubilised *E. coli* culture from uninduced vector constructs (**1**), induced vector control (**2**) and induced vector constructs (**3**) were electrophoresed on a 12% Laemmli gel alongside an erythrocyte membrane marker (**M**), as well as insoluble (**4**) and soluble (**5**) bacterial fractions from induced vector constructs. Soluble GST-30kDa protein was obtained by overnight expression in Overnight Express™ Autoinduction medium (**O/N Exp**): **6**, induced vector construct; **7**, insoluble fraction and **8**, soluble *E. coli* lysate. GST (orange arrows) is 25kDa in size. GST-tagged 10kDa, 16kDa, 22kDa and 30kDa domains (green arrows) resolve at approximately 33kDa, 36kDa, 48kDa and 60kDa respectively, as determined by molecular weight prediction software available at http://us.expasy.org.

GST-4.1R domains were purified from soluble bacterial fractions using MagneGST^{TM} beads (2.10.3) (Figure 34A). 0.5-1μg of purified protein was obtained from 1ml LB culture volumes or 25ml Overnight Express cultures. Scanning densitometry of a 12% Laemmli gel revealed that the samples were more than 85% pure. An immunoblot was performed to confirm the identity of GST-tagged 4.1R domains (Figure 34B).

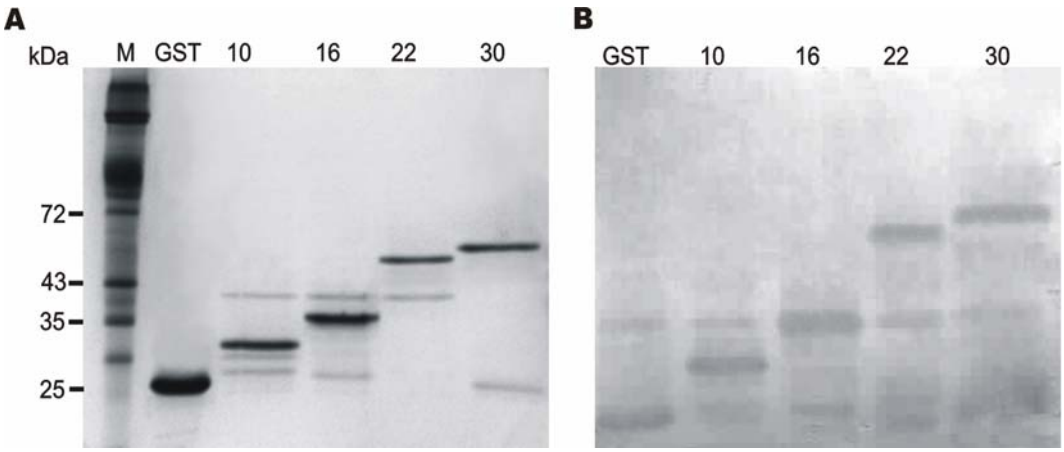

Figure 34- Purification and immuno-detection of GST-4.1R domains

A) GST-4.1R fusion domains were expressed and purified using glutathione magnetic affinity beads. 20μl or approximately 1-2μg of total protein was resolved by 12% SDS-PAGE: **M**, erythrocyte membrane marker. **B**) Immunoblot to confirm the expression of GST-4.1R domains using an anti-GST antibody.

3.9 Binding studies between 6His-*Pf*J and GST-4.1R domains

Recombinant proteins were used in protein-protein interaction assays to determine the binding site in 4.1R. GST-4.1R domains were purified and examined for their ability to bind 6His-*Pf*J using blot overlay (2.11.1) and histidine pull down assays (2.11.2).

Approximately 5μg of 6His-*Pf*J and GST were spotted on nitrocellulose membrane strips and probed with 0.5-1μg of the different 4.1R recombinant proteins. Under the binding conditions used in these experiments native unbiotinylated 4.1R and the GST-10kDa domain were

able to interact with histidine-tagged parasite protein (Figure 35). 6His-*Pf*J did not bind to purified GST nor GST-tagged 16, 22 and 30kDa fragments.

To confirm the interaction between GST-4.1R fragments and 6His-*Pf*J, histidine pull-down assays were performed. Seven hundred ng of the parasite protein was immobilised on His-Select™ magnetic beads and incubated with 150, 400 and 800ng of GST-10kDa. In agreement with the blot overlay data, GST-10kDa bound specifically to 6His-*Pf*J and was found to be dose dependent (Figure 36A). The interaction was abolished by heat denaturation of 5μg 10kDa domain. Histidine pull-down assays were also performed with 6μg of GST-16, 22 and 30kDa protein and ~800ng 6His-*Pf*J. As shown in figure 36B, no binding was detectable in the 30 and 22kDa reactions. The GST-16kDa pull-down assay revealed some binding with 6His-*Pf*J, but this was non-specific as the interaction was comparable to heat-denatured 4.1R domain.

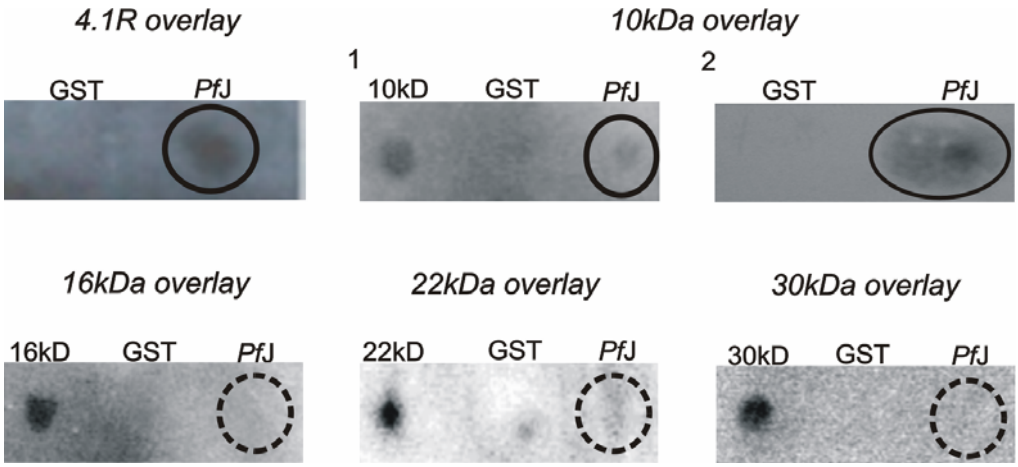

Figure 35- Blot overlay of 6His-*Pf*J with native 4.1R and GST-tagged 4.1R domains

The headings above each Western blot represent the proteins spotted on the nitrocellulose membrane. GST and the relevant 4.1R domain served as negative and positive controls respectively. The membranes were overlaid with native 4.1R (*4.1R overlay*) and GST-tagged domains (*10kDa*, *16kDa* and *30kDa overlays*) and binding detected with polyclonal rabbit anti-4.1R antibody using a chemiluminescent substrate. All the 4.1R fragments were detectable using the primary polyclonal antibody as demonstrated by the positive controls on the overlays. The *4.1R overlay* indicates that 6His-*Pf*J binds native 4.1R. Two independent experiments (**1** and **2**) revealed specific interaction between 6His-*Pf*J and GST-10kDa. **GST**, glutathione S-transferase.

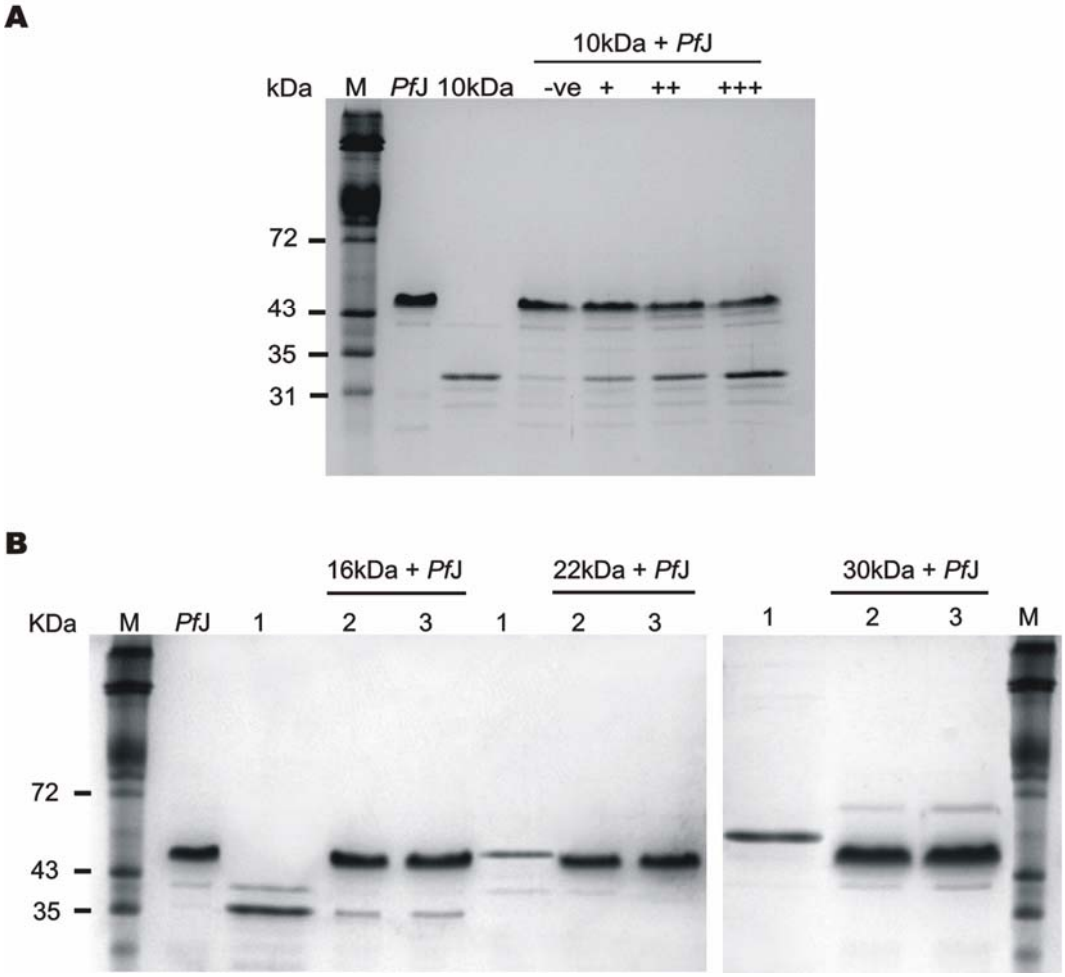

Figure 36- Histidine pull-down of GST-4.1R recombinant proteins

A) 6His-*Pf*J bound specifically to GST-10kDa in a histidine pull-down assay. The interaction was dose dependent and reliant on native GST-10kDa protein conformation. **M**, erythrocyte membrane marker; **-ve**, negative control pull-down using 5μg heat denatured GST-10kDa and +, assays indicating the dose-dependent nature of the interaction (+/++/+++ ≈ 150ng/ 400ng/ 800ng GST-10kDa respectively). **B**) 6His-*Pf*J did not bind specifically to GST-16, 22 and 30kDa proteins. **1**, purified GST-4.1R domain; **2**, histidine pull-down using heat denatured recombinant protein and **3**, pull-down using non-denatured protein.

To determine whether the interaction between 6His-*Pf*J and GST-10kDa was saturable, the parasite protein was incubated with a wider concentration range of 10kDa domain. Protein complexes were eluted and resolved by 12% SDS-PAGE. 54ng heat denatured GST-10kDa per μg 6His-*Pf*J was subtracted from the amount of bound protein. This value was determined by averaging the non-specific binding displayed by 0.5μg and 10μg of heat denatured 4.1R domain, which was 51 and 59ng/μg *Pf*J respectively. Scanning densitometry of the polyacrylamide gels revealed that the interaction was concentration dependent and saturable in three independent experiments using different protein preparations (Figure 37A-C). This result demonstrated that 6His-*Pf*J bound specifically to GST-10kDa. A rough estimate of the dissociation constant (Kd) was obtained by linear transformations of the three curves and ranged from 163nM to 1 500nM with an average Kd of 745nM.

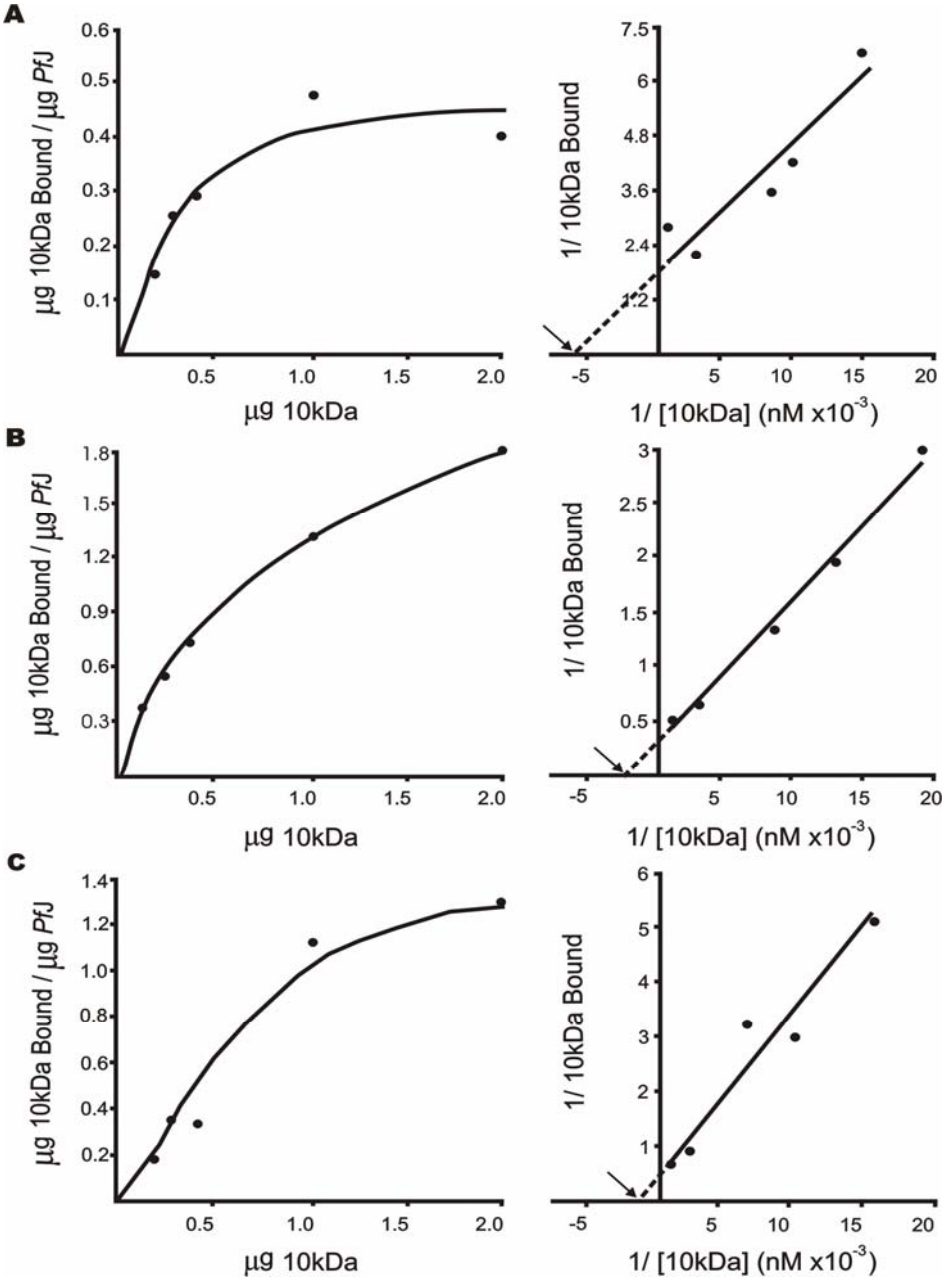

Figure 37- GST-10kDa/6His-*Pf*J solution binding assays

GST-10kDa (200, 300, 400, 1000 and 2000ng) bound to ~700ng 6His-*Pf*J on His-Select™ magnetic beads in a concentration dependent manner (panels **A**-**C**: left hand side). Protein complexes were resolved on 12% Laemmli gels and analysed with scanning densitometry. Non-specific binding by heat denatured GST-10kDa (54ng/µg *Pf*J) was subtracted from the amount of bound protein. Graphs of 1/bound (bound represented as µg 10kDa bound/ µg *Pf*J) versus 1/[10kDa] (nM) revealed an average dissociation constant (**Kd**) of 745nM, which was obtained from the x intercepts of the linear curves (indicated by the arrows) in panels **A**-**C** on the right hand side (Kd: **A**, 163nM; **B**, 573nM and **C**, 1 500nM). Binding data for curve A were obtained from different GST-10kDa and 6His-*Pf*J protein preparations, than for curves B and C.