Supplementary information

A multifunctional serine protease primes the malaria parasite for red blood cell invasion

Konstantinos Koussis, Chrislaine Withers-Martinez, Sharon Yeoh, Matthew Child, Fiona Hackett, Ellen Knuepfer, Luiz Juliano, Ute Woehlbier, Hermann Bujard, and Michael J. Blackman.

Figure S1. Correct cleavage by recombinant PfSUB1 of internally quenched peptide Abz-LVSADNIDIQ-EDDnp

RP-HPLC chromatogram showing fractionation of Abz-LVSADNIDIQ-EDDnp after partial hydrolysis with recombinant PfSUB1 (rPfSUB1). The identities of the three main peaks, including their predicted and measured masses as determined by electrospray mass spectrometry, are indicated.

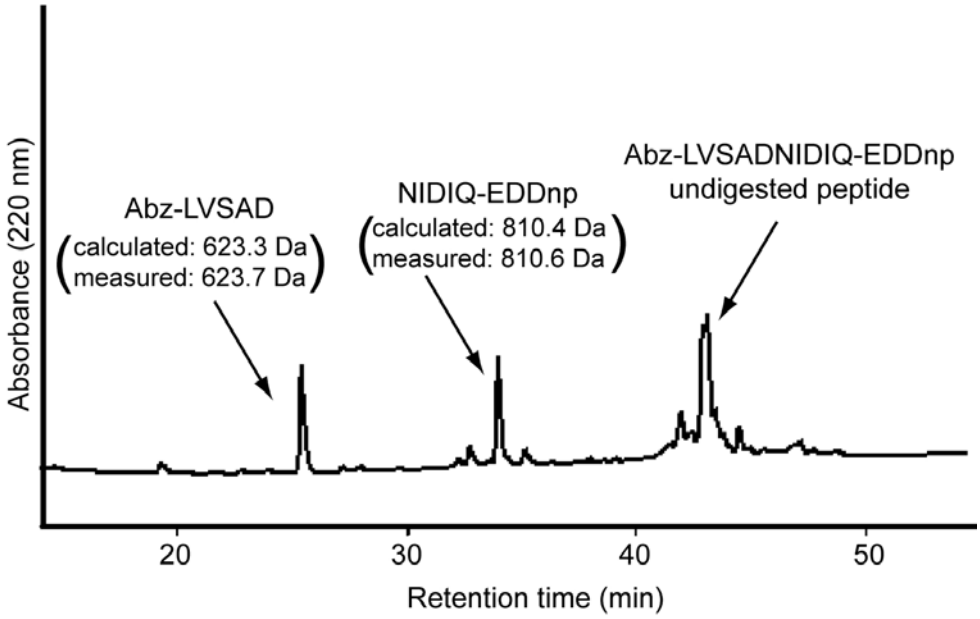

Figure S2. Primary processing sites within MSP1 and the associated peripheral membrane proteins MSP6 and MSP7

For clarity, fragments are not drawn to scale. The $MSP1_{83}$-$MSP1_{30}$ junction in the FCB-1 allelic form of MSP1, and the precise position of an additional processing site within the 3D7 form of $MSP1_{83}$, indicated by question marks, have not been previously determined, but are established in this study (see main manuscript text and below). The $MSP6_{36}$ cleavage site in the FCB-1 form of this protein, also indicated by a question mark, is unknown. For references see Table 1.

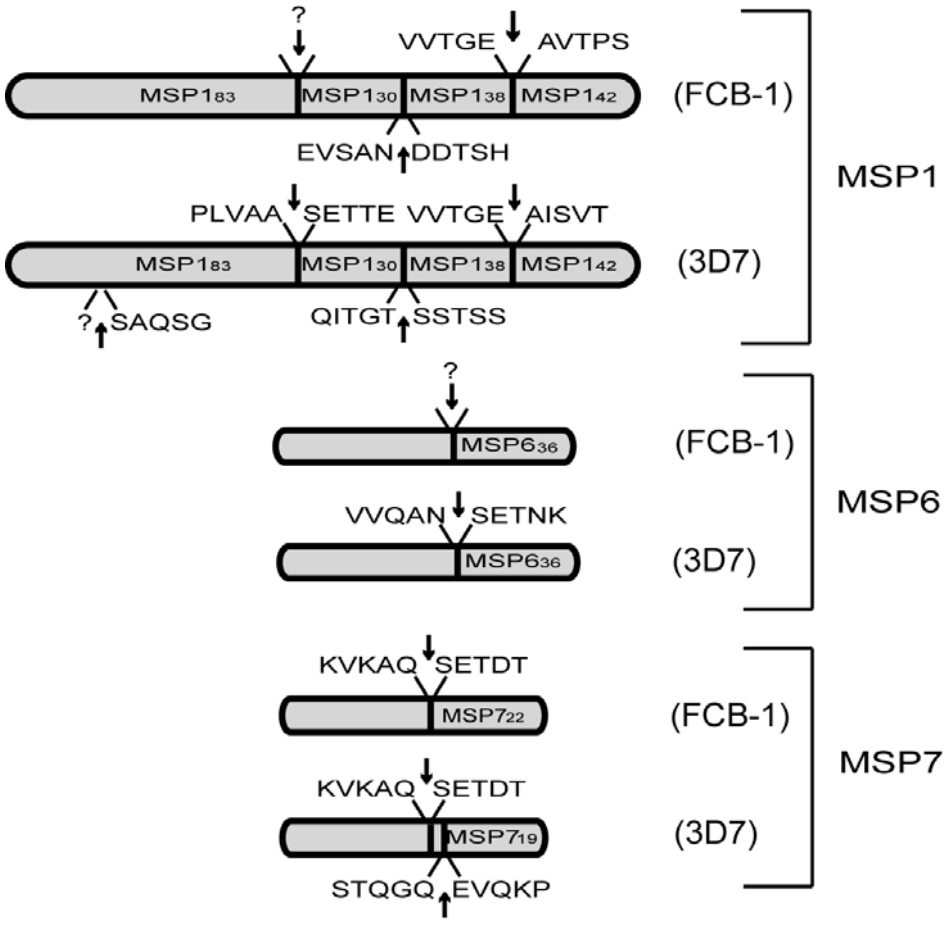

Figure S3. Proteolytic processing of recombinant full-length MSP1 of both allelic forms by rPfSUB1

Coomassie stained gels showing a time course of digestion of recombinant 3D7 and FCB-1 type MSP1 (MSP-1D and MSP-1F respectively) by rPfSUB1. In both cases the proteins were converted to a limited number of stable major fragments.

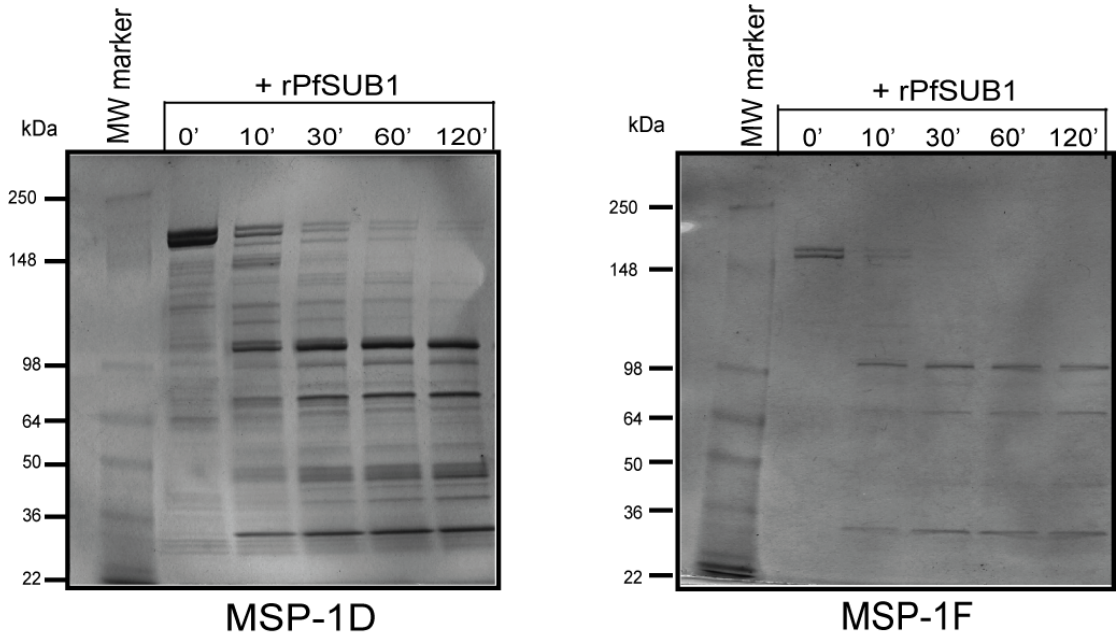

Figure S4. Positional conservation of the $MSP1_{83}$-$MSP1_{30}$ junction in the two major allelic forms of *P. falciparum* MSP1

Clustal alignment of partial predicted *P*. *falciparum* primary MSP1 sequences spanning the junction between $MSP1_{83}$ (green) and $MSP1_{30}$ (yellow). Sequences were obtained from PlasmoDB (3D7 type, PlasmoDB ID PFI1475w) or from NCBI (FCB-1 type, accession no. P04932). Numbering is according to each sequence. The experimentally-determined processing sites are indicated by a downward-pointing arrow, with P1 and P1´ residues shown in red. Gaps introduced to improve the alignment are shown by hyphens. Positions of sequence identity are indicated by asterisks.

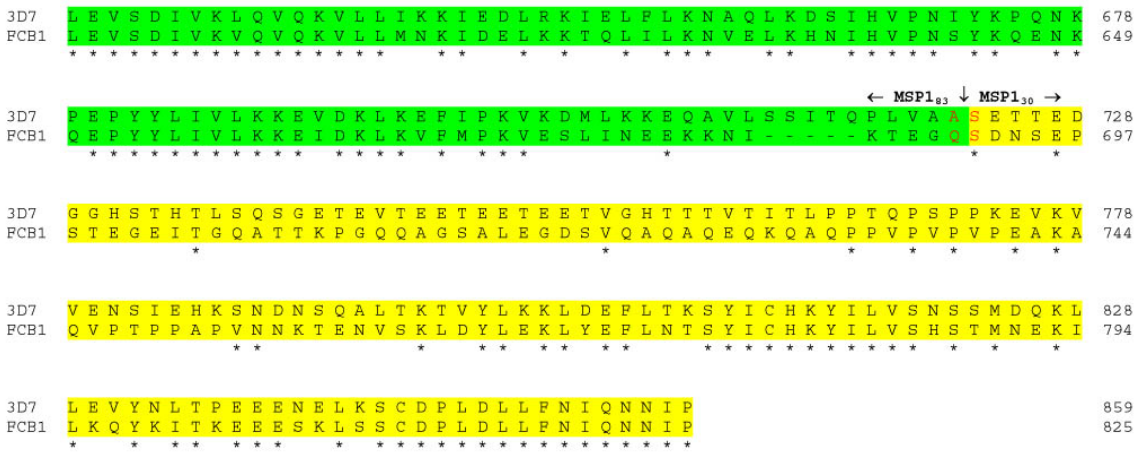

Figure S5. Correct cleavage by rPfSUB1 of synthetic peptides based on MSP processing sites

RP-HPLC chromatograms showing a time course of digestion by rPfSUB1 of a mixture of decapeptides pepMSP1_30/38_3D7 (Ac-QITGTSSTSS, based on sequence flanking the junction between 3D7 $MSP1_{30}$ and $MSP1_{38}$) and SERA6st1 (Ac-KVKAQDDFNP). Both peptides were cleaved at a single internal bond to produce two products of the expected masses, as determined by mass spectrometry (peak identities are indicated). One highly polar cleavage product (NH_2-SSTSS-OH from pepMSP1_30/38_3D7) was not retained by the RP-HPLC column but eluted in the column flow-though. Although the starting concentrations of both peptides were equal, SERA6st1 was cleaved much more rapidly than pepMSP1_30/38_3D7.

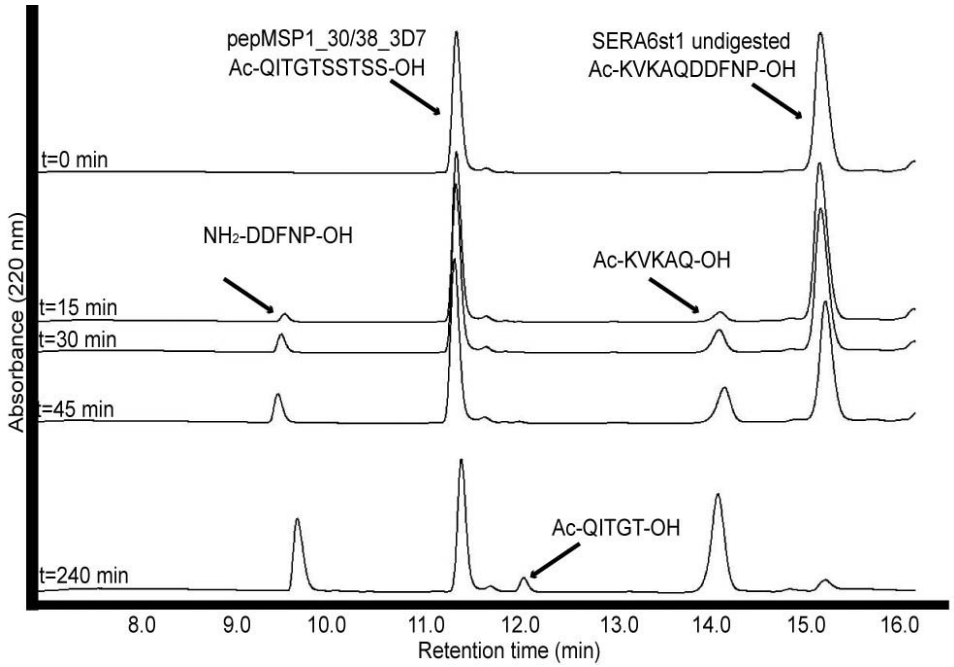

Figure S6. Different cleavage efficiencies of peptides based on SERA4 site 1 and SERA6 site 1 processing sites.

Analytical RP-HPLC chromatograms showing a time course of partial digestion of a mixture of decapeptides SERA4st1 (Ac-KITAQDDEES) and SERA6st1 (Ac-KVKAQDDFNP) with rPfSUB1. Both peptides were cleaved by rPfSUB1 only at a single internal bond to produce products of the expected masses, as determined by mass spectrometry (peak identities are indicated). The relatively polar C-terminal cleavage products (NH_2-DDEES-OH from SERA4st1 and NH_2-DDFNP-OH from SERA6st1) were not retained by the RP-HPLC column but eluted in the column flow-though. Note that, although the starting concentrations of both peptides in the mixture were equal, the initial rate of cleavage of SERA4st1 was much greater than that of SERA6st1.

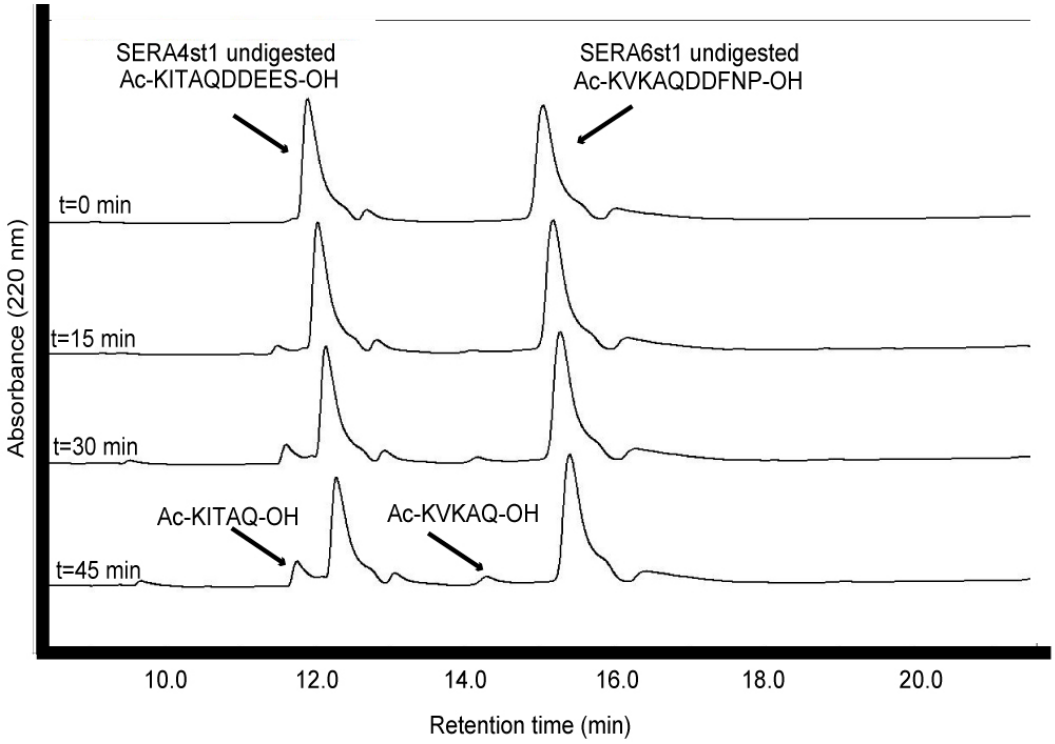

Figure S7. Sequence logo depicting the substrate preference of PfSUB1 based on SERA and MSP processing sites

Graphical representation of a multiple sequence alignment (in single-letter code) of the P5 to P5´ residues flanking known and predicted PfSUB1 cleavage sites within SERA5 (sites 1, 2 and 3), SERA4 and SERA6 (predicted sites 1 and 2), and all the MSP1, MSP6 and MSP7 sequences investigated in this study (a total of 17 different sequences). The overall height of each stack of residues indicates the degree of sequence conservation at that position, while the height of each residue within the stack indicates the relative frequency of that amino acid residue at that position. Residues are colour-coded according to the chemical nature of their side-chains (red, acidic [D,E]; blue, basic [K,R,H]; green, aliphatic [L,V,I]; black, small [G,A]; purple, non-polar, non-aliphatic [W,F,M,P]; orange, uncharged polar [S,T,Y,N,Q]). The scissile bond is indicated by an arrow. The figure was produced using the WebLogo website at http://weblogo.berkeley.edu/logo.cgi and further annotated with Adobe Photoshop.

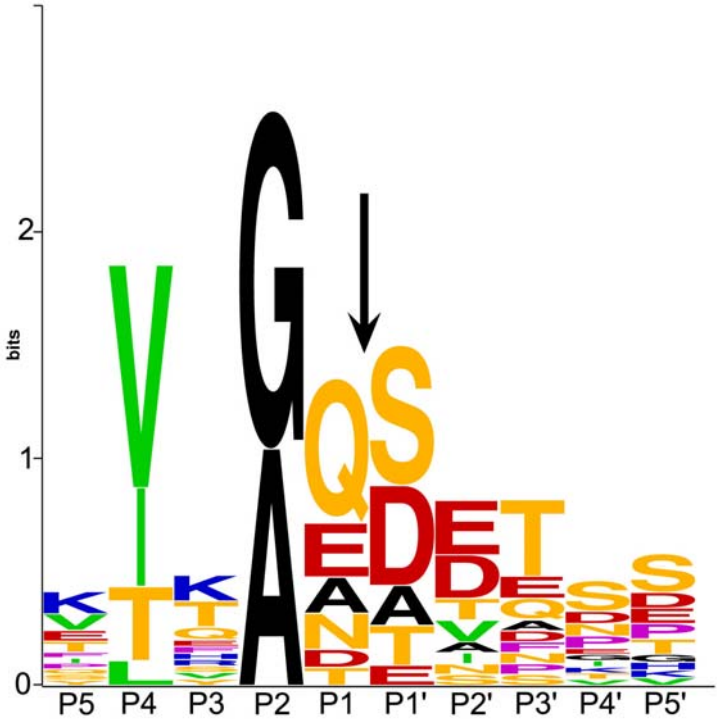

Figure S8. SDS PAGE analysis of purified PfSUB1 and PfSUB2 prodomain.

Coomassie-stained SDS PAGE gel of purified recombinant PfSUB1 and PfSUB2 prodomain (~16 pmol of each), produced and quantified as described in Materials and Methods. The left-hand lane contains molecular mass markers (masses indicated in kDa).

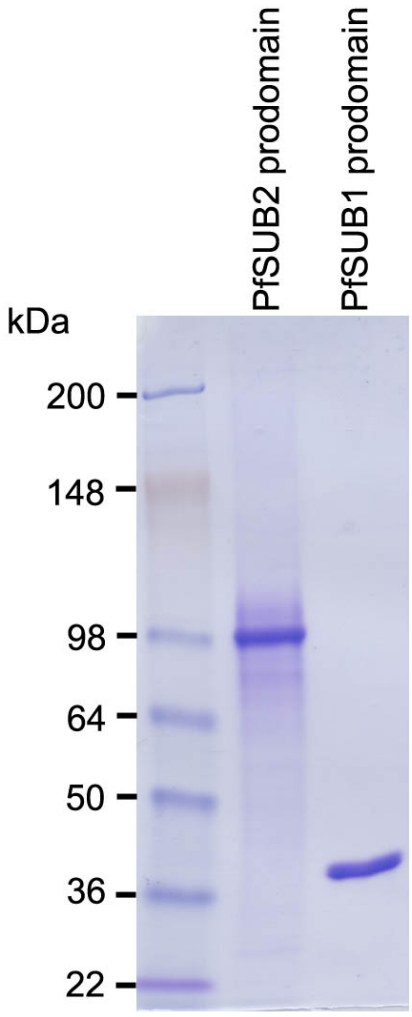

Table S1. Correct cleavage by rPfSUB1 of synthetic decapeptides based on MSP1, MSP6 and MSP7 primary processing sites

Peptide name	Peptide sequence[a]	Digestion product identified[b]	Predicted *m/z*	Measured *m/z*
pepMSP1_83/30_3D7	Ac-PLVAASETTE	Ac-PLVAA	512.30	512.30
pepMSP1_30/38_3D7	Ac-QITGTSSTSS	Ac-QITGT	561.28	560.80
pepMSP1_30/38_FCB	Ac-EVSANDDTSH	Ac-EVSAN	561.25	561.30
pepMSP1_38/42_3D7	Ac-VVTGEAISVT	Ac-VVTGE	546.27	546.24
pepMSP6_36	Ac-VVQANSETNK	Ac-VVQAN	572.30	572.26
pepMSP7_22	Ac-KVKAQSETDT	Ac-KVKAQ	615.38	615.40

[a] See Table 1 (main paper) for the details of predicted MSP processing sites and relevant references.

[b] Digestion products were fractionated by RP-HPLC and identified by electrospray mass spectrometry. Note that in most cases the RP-HPLC column did not retain the highly polar C-terminal cleavage products (which eluted in the column flow-through), allowing identification only of the N-terminal product of cleavage.

Table S2. Protease inhibitors used to pre-treat schizonts used for MSP processing assays

Inhibitor	Inhibitor type, target protease class	Concentration used (final)
4-(2-aminoethyl) benzenesulfonyl fluoride (AEBSF)	Irreversible, serine	1 mM
Antipain	Reversible, cysteine/serine	10 μg ml^{-1}
3,4-dichloroisocoumarin (DCI)[a]	Irreversible, serine	10 μM
E64	Irreversible, cysteine	10 μM
EDTA	Divalent cation chelator, many	5 mM
EGTA	Calcium chelator, many	5 mM
Leupeptin	Reversible, cysteine/serine	10 μg ml^{-1}
Pepstatin A	Reversible, aspartic	1 μM
para-hydroxy mercuribenzoate (pHMB)	Irreversible, cysteine/serine	1 mM
Phenylmethyl sulfonyl fluoride (PMSF)	Irreversible, serine	1 mM
N-tosyl-L-lysine chloromethylketone (TLCK)	Irreversible, serine	10 μM
N-tosyl-L-phenylalanine chloromethylketone (TPCK)	Irreversible, serine	10 μM

[a] Of the inhibitors listed here, PfSUB1 is substantially sensitive only to DCI, EDTA, EGTA and pHMB (Withers-Martinez et al., 2002) and K. Koussis and M. Blackman, unpublished data)

Table S3. Peptides identified by MALDI-TOF analysis of $MSP1_{83}$ tryptic digests[a]

Peptide sequence	Residue numbers	Calculated *m/z*	Observed *m/z*	Error +/- ppm
NYLFTIK	143-149	897.496	898.466	-42.5
LNFYFDLLR	190-198	1199.634	1200.652	8.8
ANELDVLK	220-227	900.492	901.516	18.1
ANELDVLKK	220-228	1028.587	1029.554	-38.9
KLVFGYR	228-234	881.512	882.518	-2.4
LVFGYR	229-234	753.417	754.461	48.2
TTIANINELIEGSKK	256-270	1629.894	1630.910	5.5
LYQAQYDLSIYNK	287-299	1617.804	1618.819	4.5
QLEEAHNLISVLEKR	300-314	1777.969	1779.007	17.1
RIDTLK	314-319	744.449	745.476	25.6
FNIDSLFTDPLELEYYLR	373-390	2247.110	2248.199	36.2
VDVTPKSQDPTK	396-407	1313.683	1314.691	0.5
IITDNKER	458-465	987.535	988.536	-6.9
DVVDKIFSAR	523-532	1148.619	1149.623	-2.9
YTYNVEK	533-539	915.434	916.422	-21.5
ALSYLEDYSLRK	562-573	1456.756	1457.806	28.6
DFNHYYTLKTGLLEADIK	579-595	2027.000	2028.079	35.3
IEDLRK	643-648	772.444	773.467	19.1
KIELFLK	648-654	889.564	890.534	-42.5

[a] All predicted tryptic peptide sequences are based on the 3D7 MSP1 sequence (PlasmoDB ID PFI1475w).