CHEMBIOCHEM

Supporting Information

CHEMBIOCHEM

Supporting Information

for

Synthesis, Solution Structure and Virosomal Display of an Epidermal Growth Factor Domain from *Plasmodiun falciparum* Merozoite Surface Protein-1

Sonya James, Kerstin Moehle, Annabelle Renard, Markus S. Müller, Denise Vogel, Rinaldo Zurbriggen, Gerd Pluschke* and John A. Robinson*

Peptide Synthesis: Peptides were synthesised on an *Applied Biosystems ABI433A* peptide synthesiser, using standard Fmoc chemistry and Rink Amide MBHA resin (390 mg, 0.64 mmol/g), or for disulfide mapping studies on 2-chlorotrityl chloride resin (loading 0.25-0.3 mmol/g). The following side-chain protected amino acids were used for the synthesis: Fmoc-Asn(Trt)-OH, Fmoc-Ile-OH, Fmoc-Ser(tBu)-OH, Fmoc-Gln(Trt)-OH, Fmoc-His(Trt)-OH, Fmoc-Cys(Trt)-OH, Fmoc-Val-OH, Fmoc-Lys(Boc)-OH, Fmoc-Pro-OH, Fmoc-Gly-OH, Fmoc-Phe-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Leu-OH, Fmoc-Asp(tBu)-OH, Fmoc-Tyr(tBu)-OH, Fmoc-Glu(tBu)-OH and Fmoc-Thr(tBu)-OH. Fmoc deprotection was performed with 20% piperidine in NMP. Coupling reactions were carried out using HBTU/HOBt (4 eq.) in N-methylpyrrolidone (NMP) with diisopropylethylamine (8 eq.). The peptide was cleaved from the resin with trifluoroacetic acid/triisopropylsilane/water/ethanedithiol (94:1:2.5:2.5) for 3 h. The resin was removed by filtration, and after concentration *in vacuo* the crude linear peptide was precipitated by addition of cold diethyl ether, washed twice with diethyl ether, and dried to give approx. 1g of white solid. The product was purified by preparative reversed-phase HPLC (Vydac C18 218TP1022 column, 250mm x 22mm, 10µm, 300Å

pore size) and a gradient of 5 to 35% acetonitrile in water (+0.1% TFA) over 25 minutes. ESI-MS: *m/z* 5591.72 (±0.01%) (calc. 5592.36)

Linear peptide (17 mg) in Tris buffer (3.5 mL, 100 mM, pH 8.4) containing EDTA (10 mM), cysteine (3 mM) and cystine (0.3 mM) was stirred for 48 h at rt. The folded product was purified by semi-preparative reversed-phase HPLC (Vydac C18 218TP1010 column) and a gradient of 5 to 30% acetonitrile in water (+ 0.1% TFA) over 25 min to give product **1** (4 mg). ESI-MS: *m/z* 5586.3 (±0.01%) (calc. 5586.31).

For the synthesis of **2**, N',N',N-tris-Boc-hydrazinoglycine (3 eq.) was coupled to resin bound linear peptide using HATU (3 eq.) and HOAt (3 eq.) in DMF and DIEA (12 equiv). Cleavage from the resin and deprotection, purification and folding was as for **1**. The phospholipid **4** (2.8 mg, 1.8 eq.) was coupled to folded peptide (10 mg) in distilled water (1.5 mL; pH 5.2) and tert-butanol (1.5 mL). The reaction was followed by LC-MS and was complete after 24 h. The product was purified by HPLC on an Interchrom UP10WC4/25M (21.2 x 250mm) C4 HPLC column using a gradient of 10 to 100% acetonitrile in water (+0.1% TFA) over 17 min (yield ca. 2 mg). ESI-MS: *m/z* 6432.5 (±0.01%) (calc. 6433.4).

Disulfide mapping: Although the NMR structure of the oxidized EGF domain provided strong evidence for the disulfide bridge between Cys^{30} and Cys^{41}, the pairings between Cys^{7}, Cys^{12}, Cys^{18} and Cys^{28} could not be deduced unambiguously from the NMR data, due to the close spatial proximity of the two disulfide bridges in the folded protein. For this, a combination of proteolytic digestion, mass spectrometry (MS) and Edman sequencing was used. Digestions of the synthetic folded peptide were performed with trypsin and thermolysin at a pH that minimized the possibility of disulfide scrambling. The cleavage products were analyzed by reverse phase HPLC-MS; the corresponding chromatograms are shown in Figures S1 and S2, and the fragments identified are presented in Tables S1 (trypsin digestion) and S2 (thermolysin digestion).

Upon trypsin digestion, four main fractions were identified by RP-HPLC, numbered 13, 19, 20 and 21 in Figure S1. The fragment in peak 13 (mass 2827.2) consists of peptides NISQHQCVK, KQQ**C**PQNSG**C**FR and EREE**C**K, linked by two disulfide bridges involving Cys^{7}(1), Cys^{12}(2), Cys^{18}(3) and Cys^{28}(4). These assignments were confirmed upon analysis of products following reductive cleavage of the disulfide bridges (Table S3). The connectivity of the four Cys residues, however,

could be either 1-2 and 3-4, or 1-3 and 2-4. The fragments identified in peaks 19 and 20, shown in Table S1, also do not resolve the connectivity amongst Cys 1-4.

The fragments identified after digestion with thermolysin are shown in Figure S2 and Table S2. Peak 6 contained two fragments (mass 1092.2 and 1005.1) whose masses are consistent with the peptide NISQHQC linked *via* one disulfide bond to either the tripeptide SGC or the dipeptide GC, thus confirming the connection between Cys7 and Cys18. This conclusion is supported by the three structurally related fragments identified in fractions 11 and 12, which were found to be constituted from three peptides linked by two disulfide bridges involving the four cysteine residues Cys12, Cys28, Cys30 and Cys41. Since the link between Cys30 and Cys41 has been established by the trypsin digest and by NMR, the bridge between the two remaining cysteine residues Cys12 and Cys28 was implied. To confirm this conclusion, the thermolysin fragment 12 was treated with trypsin and the resulting digestion products were isolated by RP-HPLC and analysed by MALDI-TOF-MS (Table S4). Fragment 1, consisting of peptides KQ**C**PQ and EE**C**K connected by a disulfide bond establish firmly the disulfide bridge between Cys12 and Cys28.

Methods

Trypsin digestion. EGF domain (1 mg) was digested with TPCK-treated trypsin (50 µg) in phosphate buffer (2 mL, 50 mM, pH 6.8) for 24 h at 37°C. The digest was then analysed by RP-HPLC (Vydac 218TP54 analytical column, 4 x 100 mm, flow rate 1 mL/min) using a gradient of 5 to 23% acetonitrile in water (+0.1 % TFA) over 43 min. The digestion products were then separated by RP-HPLC (same conditions) into the fractions identified in Figure S1. The fractions were analysed by LC-MS. The peak masses found and fragments identified are given in Table S1.

Thermolysin digestion. EGF domain (1 mg) was digested with thermolysin (100 µg) in *N*-ethyl morpholine/acetate buffer (2 mL, 25 mM, pH 6.4) for 2 h at 40°C. The digest was analysed by RP-HPLC (Zorbax Eclipse analytical C18, 4.6 x 250mm, 5 µm, flow rate 1 mL/min), gradient 3 to 30% acetonitrile in water (+0.1% TFA) over 43 min (Figure S2). The HPLC fractions were analysed by LC-MS. The peak masses found and fragments identified are given in Table S2.

Trypsin digestion of thermolysin fragment 11. The peptide in fraction 11 collected from the thermolysin digest (estimated to be approx. 70 µg) was digested with TPCK-treated trypsin (3.5 µg) in phosphate buffer (200 µl, 50 mM, pH 6.8) for 24 h at 37°C.

The digest was then analysed by RP-HPLC (Zorbax Eclipse analytical C18, 4.6 x 250mm, 5 μm 80 Å, flow rate 1 mL/min), gradient 3 to 30% acetonitrile in water (+0.1% TFA) over 43 min. The HPLC fractions were analysed by LC-MS. The peak masses found and fragments identified are given in Table S3.

Reduction of digestion products. Five fragments were selected for further investigation (see Table S4). Samples of each digestion product were treated with DTT (5 equiv per Cys residue) in ammonium carbonate buffer (0.1 M, pH 8.0) for 6 h at RT. To stop the reaction, 10% (*v/v*) acetic acid was added to each sample to give a final pH of 2-3. The samples were then frozen and lyophilised. After clean-up by RP-HPLC, the samples were analysed by MALDI-MS.

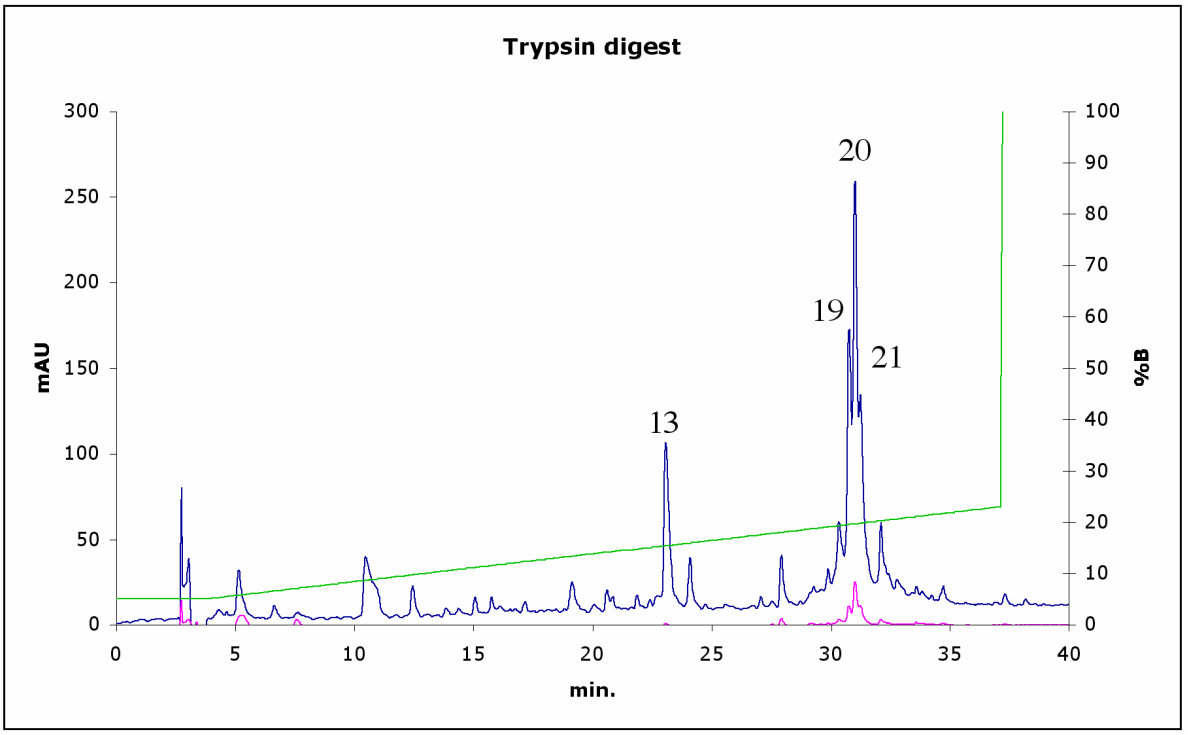

Figure S1: RP-HPLC chromatogram of the trypsin digestion products of the EGF domain.

Table S1: Fragments identified by LC-ESI-MS upon trypsin digestion of the EGF domain. The HPLC peak numbers are shown in Figure S1.

HPLC Peak	Fragment	Mass calc.	*m/z* calc.	*m/z* exp.
13	NISQHQ**C**VK KQ**C**PQNSG**C**FR EE**C**K	2827.2	707.8(+4) 566.4(+5)	707.7(+4) 566.4(+5)
19	NISQHQ**C**VK KQ**C**PQNSG**C**FR EE**C**K**C**LLNYKQEGDK**C**VENPNPT	4972.6	1244.2(+4) 995.5(+5) 829.8(+6) 711.4(+7)	1244.4(+4), 995.6(+5), 830.4(+6), 711.5(+7)
20	**C**LLNYK **C**VENPNPT	1623.9	813.0(+2)	813.1 (+2)
21	**C**LLNYKQEGDK**C**VENPNPT	2163.4	1082.7(+2) 722.1(+3)	1082.9(+2), 722.3(+3)

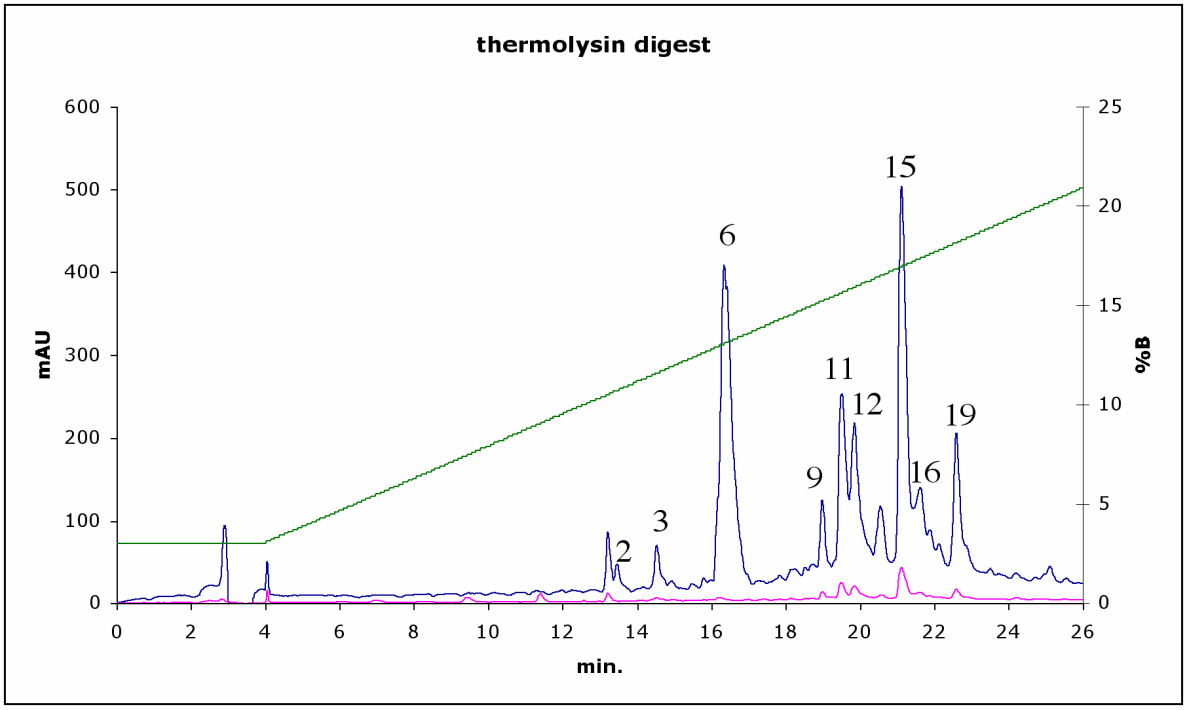

Figure S2. RP-HPLC chromatogram of the thermolysin digestion products of the EGF domain.

Table S2: Fragments identified by LC-ESI-MS from the thermolysin digestion of the EGF domain.

HPLC Peak	Fragment	Mass calc.	M/z calc.	M/z exp.
1	YKQE	566.6	567.6(+1)	567.3(+1)
	C YKQEGDK**C**	1089.2	1090.2(+1) 545.6(+3)	1089.7(+1) 545.3(+2)
2	NISQH	597.6	598.6(+1)	598.4(+1)
3	**C**PQ E**C**K**C** DK**C**	1188.4	1189.4(+1) 595.2(+2)	1189.7(+1) 595.7(+2)
6	NISQHQ**C** G**C**	1005.1	1006.1(+1)	1005.4(+1)
	NISQHQ**C** SG**C**	1092.2	1093.2(+1)	1093.1(+1)
	VENPNPT	769.8	770.8(+1)	770.5(+1)
	FRH	458.5	459.5(+1)	459.2(+1)
9	Q**C**PQN HLDEREE**C**K**C** DK**C**	2210.5	1106.3(+2) 737.8(+3) 553.6(+4)	1107.7(+2) 738.6(+3) 554.4(+4)
11	VKKQ**C**PQN LDEREE**C**K**C** YKQEGDK**C**	3034.4	1012.5(+3) 759.6(+4) 607.9(+5)	1012.4(+3) 759.4(+4) 607.9(+5)
12	VKKQ**C**PQ LDEREE**C**K**C** YKQEGDK**C**	2920.3	974.4(+3) 731.1(+4) 585.1(+5)	974.3(+3) 730.9(+4) 585.0(+5)
14	NISQHQ**C** VKKQ**C**PQNSG**C** DEREE**C**K**C** YKQEGDK**C**	3995.5	1332.8(+3) 999.9(+4)	1332.6(+3) 1000.2(+4)
15	NISQHQ**C** VKKQ**C**PQNSG**C** LDEREE**C**K**C** YKQEGDK**C**	4108.6	1028.2(+4) 822.7(+5) 685.8(+6) 587.9(+7)	1028.8(+4) 823.5(+5) 686.3(+6) 588.6(+7)
16	LLN	358.4	359.4(+1)	359.8(+1)
19	NISQHQ**C** VKKQ**C**PQNSG**C** LDEREE**C**K**C** YKQEGDK**C**VENPNPT	4860.4	1621.1(+3) 1216.1(+4) 973.1(+5)	1621.2(+3) 1216.1(+4) 973.3(+5)

Table S3. Fragments found after reduction of selected fragments isolated from trypsin and thermolysin digestions. The fragments are those shown in Tables S1 and S-2. The calculated masses of all peptides expected after reductive cleavage of intermolecular disulfide bonds are shown together with the mass fragments observed by MALDI-MS.

	Fragment	Mass calc. of fragment	Reduced peptides mass calc.	Observed mass (*m/z*)
13	NISQHQ**C**VK KQ**C**PQNSG**C**FR EE**C**K	2827.2	1056.2 1267.5 507.6	1056.5 1267.5 n.d.
	or			
	NISQHQ**C**VKK Q**C**PQNSG**C**FR EE**C**K		1184.4 1139.3 507.6	n.d. n.d. n.d.
19	NISQHQ**C**VK KQ**C**PQNSG**C**FR EE**C**K**C**LLNYKQEGDK**C**VENPNPT	4972.6	1056.2 1267.5 2655.0	1056.6 1267.6 2653.1 (Cys5-Cys6 not reduced)
	or			
	NISQHQ**C**VKK Q**C**PQNSG**C**FR EE**C**K**C**LLNYKQEGDK**C**VENPNPT		1184.4 1139.3 2655.0	n.d. 1139.5 n.d.
20	**C**LLNYK **C**VENPNPT	1623.9	752.9 873.0	n.d. 873.5
12	VKKQ**C**PQ LDEREE**C**K**C** YKQEGDK**C**	2920.3	830.0 1124.3 970.1	830.5 1124.5 970.5

n.d. = not detected

Table S4: Fragments identified by LC-ESI-MS from the trypsin digest of the thermolytic fragment 12.

VKKQCPQ LDEREECKC YKQEGDKC

HPLC Peak	Fragment	Mass calc.	M/z calc.	M/z exp.
1*	KQ**C**PQ EE**C**K	1107.5	1108.5(+1)	1108.5(+1)
2	YKQEGDK	866.9	867.9(+1) 434.5(+2)	867.8(+1) 434.2(+2)
3	**C** YKQEGDK**C**	1089.2	1090.2(+1) 545.6(+2)	1089.6(+1) 545.5(+2)

* the sequence of this fragment was confirmed by Edman degradation.

NMR spectra

NOESY plot of the peptide **1** in 90% H_2O/10% D_2O, pH 5, 298 K is shown below.