Supporting Information

McGowan *et al.* 10.1073/pnas.0807398106

SI Methods

Parasites and Preparation of Parasite Extract. *Plasmodium falciparum* clone D10 was cultured as described (1). For experiments investigating the stage-specific expression of *Pf*A-M1, parasites were synchronized by using 2 rounds of sorbitol treatment (2), and stage-specific parasites were harvested at ring, trophozoite, and schizont stage.

The *P. falciparum* M1 Alanyl Aminopeptidase Gene. Because of the high AT-richness of genes encoding malarial proteins *Pf*A-M1 was chemically synthesized by GeneArt using codons for optimized gene expression in the yeast *Pichia pastoris*. Potential N-linked glycosylation sites were removed during gene synthesis by replacing the asparagine of all Asn-Xaa-(Thr/Ser) motifs with glutamine (Gln, Q). However, expression with the codon optimized *Pf*A-M1 in *P. pastoris* was unsuccessful; therefore, we turned to *Escherichia coli* as the expression host. In this host, expression of the full-length gene did not result in a functionally-active enzyme and, therefore, a truncated form of the *P. falciparum* M1 aminopeptidase (residues 195-1085, r*Pf*A-M1; see Fig. 1), which removed N-terminal low-complexity regions and a putative transmembrane region was prepared by PCR amplification using the synthesized gene as a template followed by directional cloning into the bacterial expression vector pTrcHis2B (Invitrogen). The primers used were M1 forward 5′-CTGCAGAACCAAAGATCCAC-3′, and M1 reverse 5′-GGTACCTCA*ATGATGATGATGATGATG*TGGGCCCA-ACTTGTTTGT-3′. Unique PstI and KpnI sites (underlined) were introduced at the 5′ and 3′ ends of the amplified product. A C-terminal His-tag was introduced into the M1 reverse primer (italics). The construct housing the truncated PfA-M1 gene was verified by DNA sequencing. The verified construct was then transformed into competent BL21 cells. Expression was carried out as per the supplier's instructions (Invitrogen).

Purification and Molecular Analysis of Recombinant M1 Alanine Aminopeptidase (r*Pf*A-M1). The r*Pf*A-M1 was purified from BL21 cells by using a Ni NTA-agarose column as previously described (3). Native *Pf*A-M1 from soluble extracts of malaria parasites (>80% trophozoites) was prepared as described previously (3). Size analysis was performed by (*i*) SDS/12% PAGE analysis, and (*ii*) gel-permeation chromatography using a Amersham Pharmacia Biotech Smart System HPLC equipped with a Superdex-200 gel-filtration column. The proteins (40 μL containing 10–20 μg of protein) were passed through the column at a flow rate of 40 μL/min, and fractions were analyzed for aminopeptidolytic activity toward the fluorogenic substrate H-Arg-7-amino-4-methylcoumarin (NHMec) as described below. Separation of the molecular-size standards apoferritin (440 kDa), β-amylase (232 kDa), BSA (67 kDa), and carbonic anhydrase (29 kDa) was monitored by A_{280}.

Enzymatic Analysis. Aminopeptidase activity was determined by measuring the release of the fluorogenic leaving group, NHMec, from the fluorogenic peptide substrates H-Leu-NHMec, H-Ala-NHMec, H-Arg-NHMec, H-Met-NHMec, H-Phe-NHMec, H-Gly-NHMec, H-Val-NHMec, H-Ile-NHMec H-Glu-NHMec, H-Asp-NHMec, and H-Pro-NHMec. Reactions were carried out in 96-well microtiter plates (200 μL total volume, 30 min, 37 °C) by using a spectrofluorimeter (Bio-Tek KC4), with excitation at 370 nm and emission at 460 nm. Enzyme was first added to 50 mM Tris·HCl, pH 8.0, before the addition of 10 μM H-Leu-NHMec. Initial rates were obtained at 37 °C over a range of substrate concentrations spanning K_m (0.2–500 μM), and at fixed enzyme concentrations in 50 mM Tris·HCl, pH 8.0. Experiments to characterize the pH optimum and metal dependency of the r*Pf*AM1 were carried out by using various buffers over the pH range 3–10, and in the presence or absence of metal chelators EDTA and *o*-phenathroline as described (3).

Northern Blotting. Total RNA was extracted and Northern blotting was performed essentially as described by Kyes *et al.* (4) with the following modifications: 100-μL pellet volumes of infected red blood cells were collected from cultures at ≈5% parasitemia, lysed, and stored in TRIzol (Invitrogen). Samples were separated on a 1% TBE agarose gel containing 10 mM guanidine thiocyanate (Sigma–Aldrich), soaked in 50 mM NaOH for 30 min, and transferred onto a Hybond N+ membrane (Amersham Biosciences).

Blots were probed with a 1,500-bp PCR product amplified from a full-length *Pf*A-M1 pGem clone by using primers *Pf*A-M1IntF (TACAATGGGCTTTAGAATGTC), and *Pf*A-M1IntR (AATTCATCATCTTTTGA). This product was labeled with [α-^{32}P]dCTP by random priming by using a Decaprime II kit (Ambion). The probe was hybridized overnight at 40 °C in a hybridization buffer containing formamide (Northern Max; Ambion). The filter was washed once at low stringency and twice at high stringency (Northern Max; Ambion), then exposed overnight to Super Rx Medical X-Ray film (Fuji), and developed by using a Kodak X-Omat 3000RA processor (Kodak).

Immunoblotting. Parasite protein fractions were extracted by using 0.03% saponin (Sigma–Aldrich) and prepared as described previously (5). SDS/PAGE was performed by using 10% acrylamide gels and run on Miniprotein II rigs (BioRad). Equal loading was estimated by using the Bradford method (6), and by staining gels with Coomassie Brilliant Blue (Bio-Rad) with protein proportions visually estimated.

Protein was transferred onto Hybond C+ membranes (Amersham Biosciences), which were blocked in 5% skim-milk powder for 1 h at 37 °C or overnight at 4 °C. Anti-c-myc (Sigma–Aldrich) were used as primary antibody to label transgenic *Pf*A-M1 protein at a 1/3,000 dilution. The secondary antibody was an anti-mouse IgG (Chemicon) used at a dilution of 1/5,000. Blots were incubated with ECL Detection Reagents (Amersham Biosciences), with exposure times ranging from 5 to 10 min.

In Vitro Sensitivity of *P. falciparum* Malaria Parasites to Aminopeptidase Inhibitors. The in vitro sensitivity of *P. falciparum* (3D7) parasites to bestatin, Co4 (hPheP[CH_2]Phe), and Co2 hPheP[CH_2]Gly was determined by using [^{3}H]hypoxanthine incorporation (7). Briefly, serial dilutions of each inhibitor were prepared in culture media (0.2–200 μM) and added with [^{3}H]hypoxanthine (0.5 μCi per well) to asynchronous cultures. After a 48-h incubation, the amount of [^{3}H]hypoxanthine incorporation was measured and IC_{50} values were determined by linear interpolation of inhibition curves (8). Each assay was performed in triplicate on at least 2 separate occasions.

Crystallization, X-Ray Data Collection, Structure Determination, and Refinement. Before crystallization, the purified r*Pf*A-M1 enzyme was dialyzed against gel filtration buffer [50 mM Hepes (pH 8.5)

300 mM NaCl, 5% (vol/vol) glycerol] before size-exclusion chromatography using a Superdex S200 10/30 column and concentrated to 5 mg/mL. The crystals were grown by using the hanging drop vapor diffusion method, with 1:1 (vol/vol) ratio of protein to mother liquor (0.5 mL well volume). The crystals appeared overnight in 22% (vol/vol) polyethylene glycol 8000, 10% (vol/vol) glycerol, 0.1 M Tris (pH 8.5), and 0.2 M $MgCl_2$ and reached full size in 3 days. Crystals of the rP*f*A-M1-Co4 and rP*f*A-M1-bestatin complex were obtained by cocrystallization under similar conditions in the presence of the ligand at 1 mM. Crystals were dehydrated against reservoir buffer with 15% (vol/vol) glycerol for 16 h. Crystals were equilibrated for 5 min in reservoir buffer in the presence of 20% (vol/vol) glycerol. Cryoannealing was performed 3 times by blocking the cryostream (100 K) for 5 sec. Cryoannealing substantially improved the diffraction quality observed. Crystal quality was variable and a large number had to be screened.

Data were collected at 100 K by using both in-house and synchrotron radiation. The diffraction data for the ligand-free, Co4-bound, and bestatin-bound protease were collected to 2.1-, 2.0-, and 1.65-Å resolution, respectively. Diffraction images were processed by using MOSFLM (9) and SCALA (10) from the CCP4 suite (11); 5% of each dataset was flagged for calculation of R_{free} (12) with neither a σ nor a low-resolution cut-off applied to the data. A summary of statistics is provided in Table S2. Subsequent crystallographic and structural analysis was performed by using the CCP4i interface (13) to the CCP4 suite (11), unless stated otherwise.

Structure solution preceded by using the molecular replacement method and the program PHASER (14). A search model was constructed from the crystal structure of aminopeptidase N from *Neisseria meningitides* (PDB ID 2GTQ), the closest structural homolog identified by using the FFAS server (15). A "mixed" model, consisting of conserved side chains (all other non-alanine/glycine residues truncated at Cγ atom), was then created by using the SCRWL server (15). A single clear peak in both the rotation and translation functions was evident with 1 PfA-M1 monomer placed in the asymmetric unit. Unbiased ($F_o - F_c$) density in initial electron density maps and initial refinement cycles indicated that this solution was correct.

Maximum likelihood refinement using REFMAC (16), incorporating translation, liberation, and screw-rotation displacement (TLS) refinement was carried out, by using a bulk solvent correction (Babinet model with mask). Building was guided by manual inspection of the model and R_{free}. Simulated annealing composite omit maps were generated by using CNS (17) omitting 5% of the model. All model building and structural validation were done by using COOT (18). Water molecules were added to the model by using ARP/wARP (19) when the R_{free} reached 25%. Solvent molecules were retained only if they had acceptable hydrogen-bonding geometry contacts of 2.5–3.5 Å with protein atoms or with existing solvent, and were in good $2F_o - F_c$ and $F_o - F_c$ electron density.

PyMOL was used to produce all structural representations (www.pymol.org). Hydrogen bonds (excluding water-mediated bonds), were calculated by using the CONTACT (11). The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.rcsb.org (PDB ID codes 3EBG.pdb; 3EBH.pdb; and 3EBI.pdb).

1. Trager W, Jensen JB (1976) Human malaria parasites in continuous culture. *Science* 193:673–675.
2. Lambros C, Vanderburg JP (1979) Synchronization of *Plasmodium falciparum* erythrocytic stages in culture. *J Parasitol* 65:418–420.
3. Stack CM, *et al.* (2007) Characterization of the *Plasmodium falciparum* M17 leucyl aminopeptidase. A protease involved in amino acid regulation with potential for antimalarial drug development. *J Biol Chem* 282:2069–2080.
4. Kyes S, Pinches R, Newbold C (2000) A simple RNA analysis method shows var and rif multigene family expression patterns in *Plasmodium falciparum*. *Mol Biochem Parasitol* 105:311–315.
5. Spielmann T, Gardiner DL, Beck HP, Trenholme KR, Kemp DJ (2006) Organization of ETRAMPs and EXP-1 at the parasite-host cell interface of malaria parasites. *Mol Microbiol* 59:779–794.
6. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72:248–254.
7. Geary TG, Delaney EJ, Klotz IM, Jensen JB (1983) Inhibition of the growth of *Plasmodium falciparum* in vitro by covalent modification of hemoglobin. *Mol Biochem Parasitol* 9:59–72.
8. Huber W, Koella JC (1993) A comparison of three methods of estimating EC50 in studies of drug resistance of malaria parasites. *Acta Trop* 55:257–261.
9. Leslie AGW (1992) *Joint CCP4 + ESF-EAMCB Newslett Prot Crystallogr* (Daresbury Laboratory, Warrington, UK), 26.
10. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D* 62:72–82.
11. CCP4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D* 50:760–763.
12. Brunger AT (1993) Assessment of phase accuracy by cross validation: The free R value. Methods and applications. *Acta Crystallogr D* 49:24–36.
13. Potterton E, Briggs P, Turkenburg M, Dodson E (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr D* 59:1131–1137.
14. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr D* 61:458–464.
15. Jaroszewski L, Rychlewski L, Li Z, Li W, Godzik A (2005) FFAS03: A server for profile–profile sequence alignments. *Nucleic Acids Res* 33:W284–288.
16. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* 53:240–255.
17. Brunger AT, *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54:905–921.
18. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D* 60:2126–2132.
19. Cohen SX, *et al.* (2008) ARP/wARP and molecular replacement: The next generation. *Acta Crystallogr D* 64:49–60.
20. Davis IW, *et al.* (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35:W375–W383.

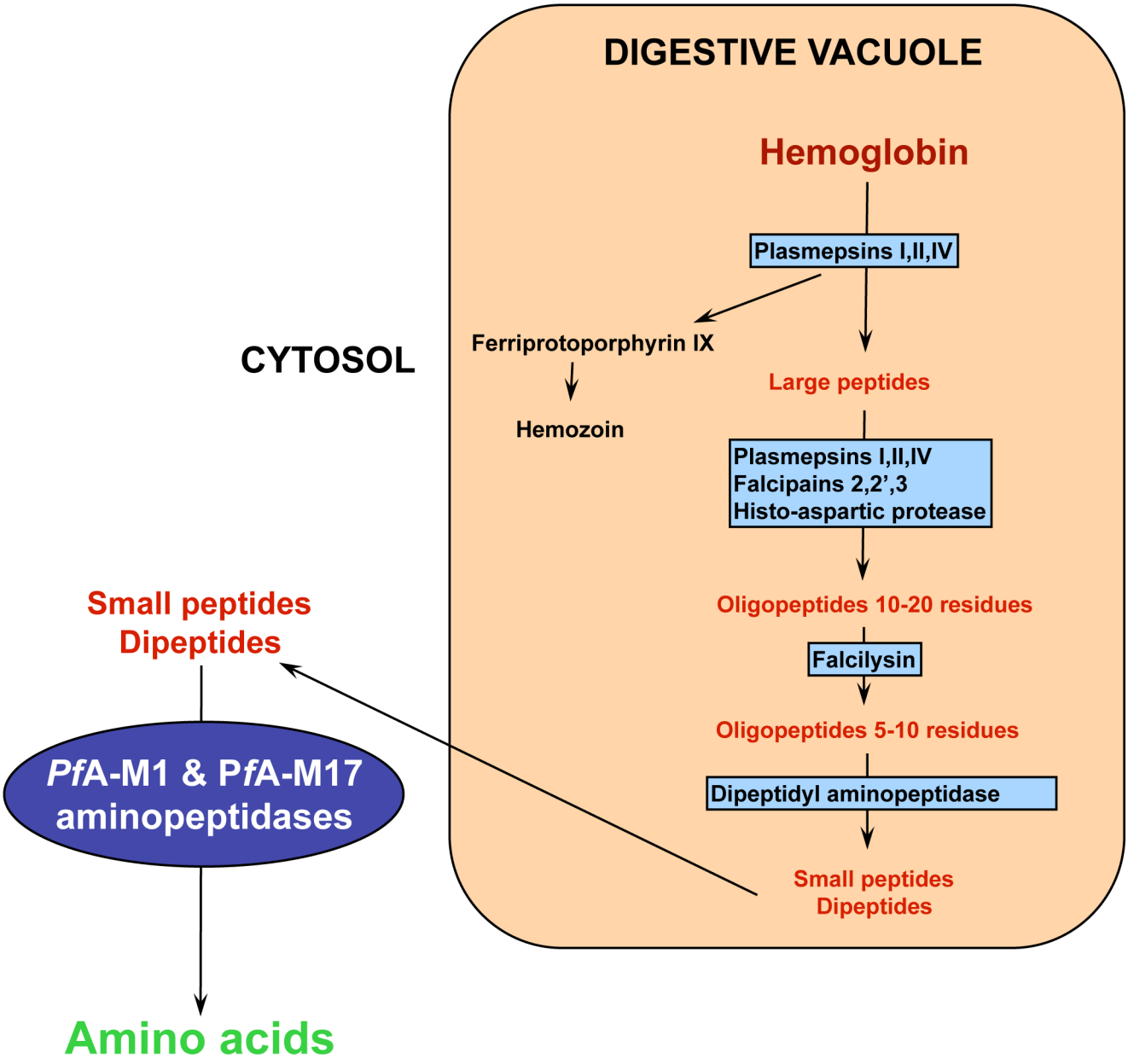

Fig. S1. Aminopeptidases *Pf*A-M1 and *Pf*A-M17 release amino acids from dipeptides that are transported from the digestive vacuole, where hemoglobin digestion is initiated by various proteases.

Fig. S2. Alignment of *Plasmodium* species M1 neutral aminopeptidases. Sequence alignment was prepared by using ClustalW and Espript. Identical residues are highlighted in red and conservatively substituted amino acids are written in red. The *Plasmodium* species are listed on the left. Putative transmembrane domain is boxed in green and asparagine-rich low complexity regions are located within black hashed boxes. The GAMEN substrate recognition motif is boxed in solid black lines and the zinc binding motif is underlined with catalytic residues indicated (asterisk). The first amino acid of the truncated r*Pf*A-M1 that was successfully purified from *E. coli* is circled.

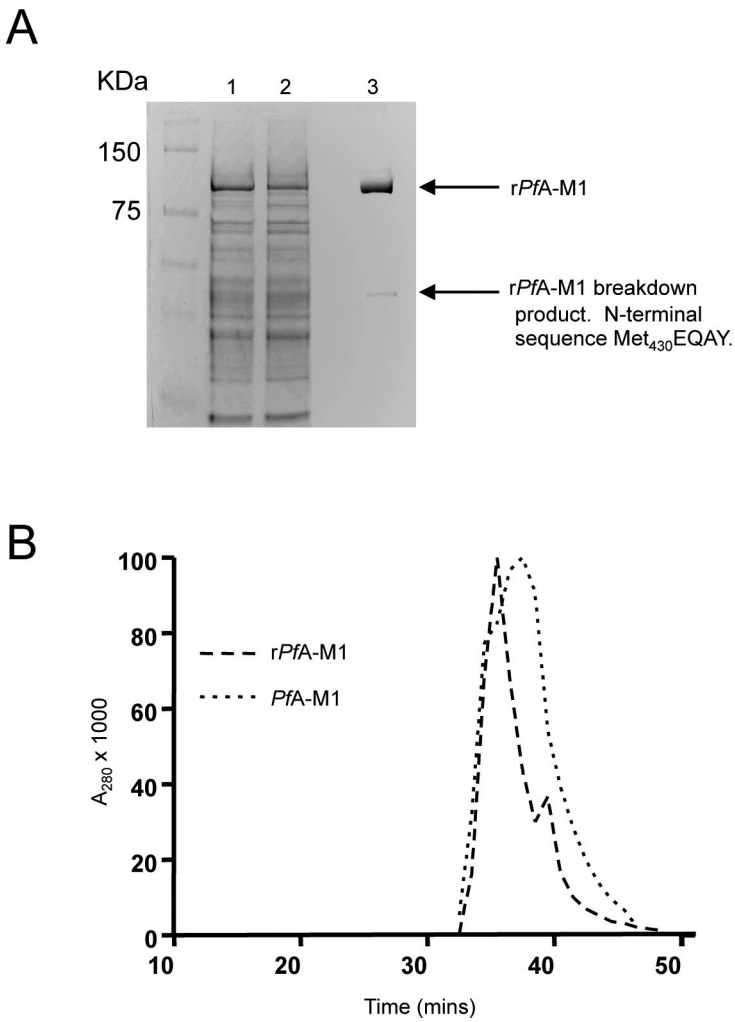

Fig. S3. (*A*) Purification of r*Pf*A-M1 by Ni-NTA-agarose affinity chromatography. Lane 1, bacterial extracts; lane 2, unbound protein; lane 3, purified r*Pf*A-M1. (*B*) Size separation of r*Pf*A-M1 (0.04 mg/mL) and parasite soluble extracts (0.05 mg/mL) on a Superdex-200 column.

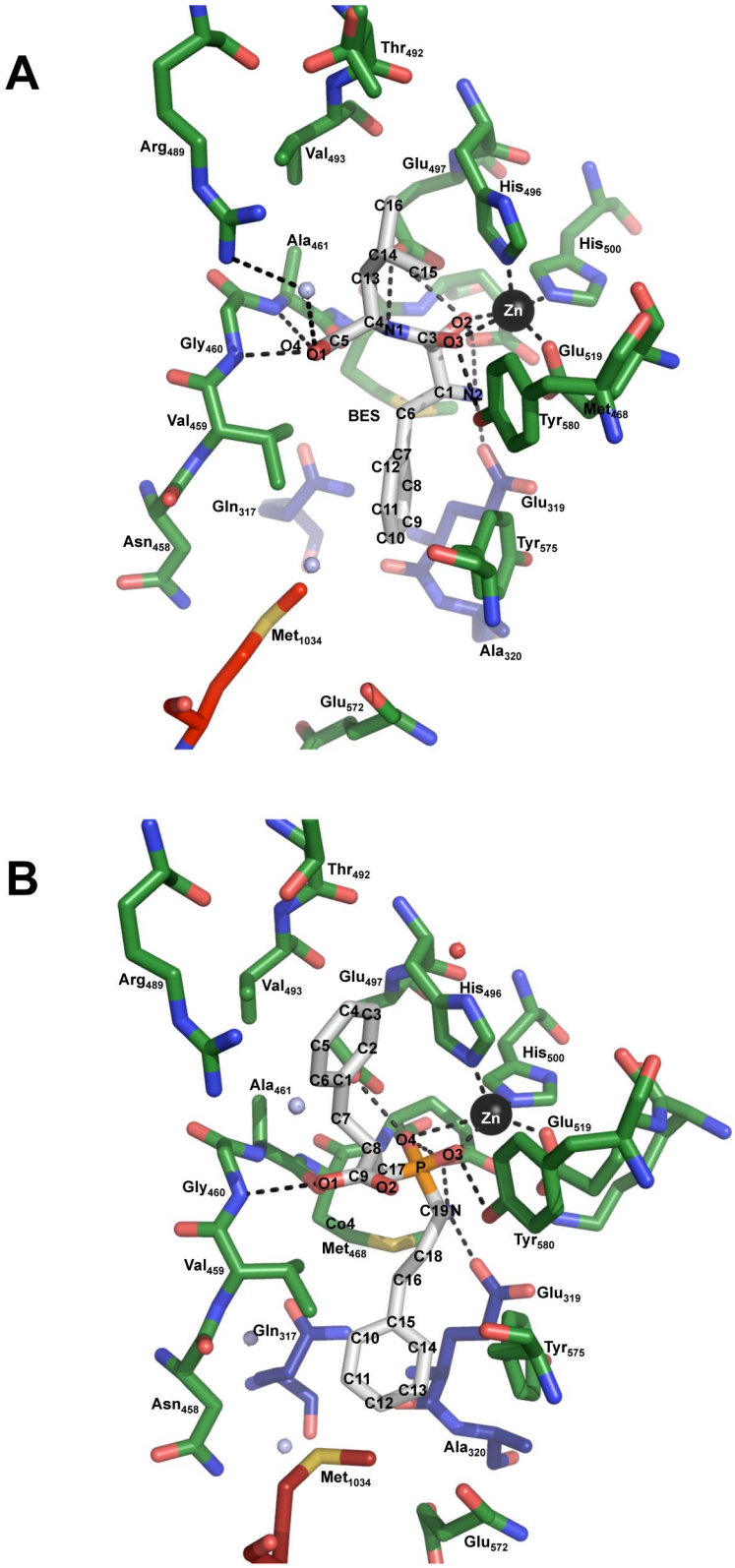

Fig. S4. Binding of inhibitors to active site of r*Pf*A-M1. (*A*) The 1.65-Å rPfA-M1-BES and (*B*) 2.0 Å rPfA-M1-Co4 active site showing inhibitors bound in the active site. Carbon atoms of inhibitors (BES/Co4) are colored in gray and atom numbers are indicated. Carbon atoms of *Pf*A-M1 residues are colored by domain: I (blue), II (green), IV (red), and Zn shown as black sphere. Water molecules are shown as blue spheres. Hydrogen bonds are indicated (dashed lines).

Table S1. Comparison of specificity constants for various N-terminal amino acids for r*Pf*A-M1 at pH 7.5

Substrate	k_{cat}, s^{-1}	K_m, μM	k_{cat}/K_m, $M^{-1}{\cdot}s^{-1}$	Abundance in human Hb, %
H-Leu-NHmec	1.52	329.9	4,607	12.46
H-Ala-NHmec	2.04	888.9	2,295	12.46
H-Arg-NHmec	1.07	717.4	1,491	2.08
H-Phe-NHmec	0.18	194.8	924	5.19
H-Gly-NHmec	0.116	348.6	333	6.92
H-Val-NHmec	0.036	1,068.1	34	10.73
H-Ile-NHmec	0.040	1,706	23	0
H-Pro-NHmec	0.0032	734.4	4	4.84

NHMec, 7-amino-4-methylcoumarin.

Table S2. Data collection and refinement statistics

	r*Pf*A-M1	r*Pf*A-M1-Co4	r*Pf*A-M1-bestatin
Data collection			
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions, Å	a = 75.7, b = 108.7, c = 118.4	a = 75.9, b = 108.6, c = 118.3	a = 75.7, b = 108.6, c = 118.0
Resolution, Å	34.99–2.1 (2.21–2.1)	28.61–2.0 (2.11–2.0)	46.5–1.65 (1.74–1.65)
Total no. of observations	289,352	432,263	440,194
No. of unique observations	56,863	60,523	95,094
Multiplicity	5.1 (3.8)	7.1 (5.5)	4.6 (4.1)
Data completeness, %	98.7 (94.0)	91.0 (77.5)	81.5 (68.1)
$\langle I/\sigma\, I\rangle$	16.7 (2.8)	22.0 (3.0)	18.3 (2.8)
R_{pim}, %*	4.2 (21.3)	3.0 (24.5)	2.9 (22.0)
Structure refinement			
No. of nonhydrogen atoms			
Protein	7,233	7,332	7,337
Solvent	762	739	835
Ligand	—	26	23
R_{free}, %	22.0	22.1	20.2
R_{cryst}, %	17.0	17.5	17.7
rms deviations from ideality			
Bond lengths, Å	0.010	0.009	0.006
Bond angles, °	1.13	1.33	0.97
Ramachandran plot			
Favored, %	98.0	98.2	98.4
Allowed, %	100	100	100
B factors, Å^2			
Mean main chain	21.8	17.4	15.4
Mean side chain	22.6	19.8	18.2
Mean ligand	—	23.5	18.2
Mean water molecule	30.8	33.3	29.7
rmsd bonded *B* values			
Main chain	0.48	0.77	0.57
Side chain	1.1	2.3	1.9
MolProbity score (ref. 20)	1.38 (99th percentile†)	1.45 (98th percentile†)	1.24 (98th percentile‡)

Values in parentheses refer to the highest-resolution shell.

*Agreement between intensities of repeated measurements of the same reflections and can be defined as $\Sigma(I_{h,i} - \langle I_h\rangle)/\Sigma\, I_{h,i}$, where $I_{h,i}$ are individual values and $\langle I_h\rangle$ *is the mean value of the intensity of reflection h.*

†N = 12,522, 1.75 Å (ref. 20).

‡N = 9,033, 1.4–1.9 Å (ref. 20).

Table S3. The r*Pf*A-M1–ligand interactions

Metal core or protein ligand	Bestatin atom	Co4 atom	Type of bond
Metal			
Zn^{2+}		r*Pf*A-M1 His^{496} $N\varepsilon_2$	Metal–protein
Zn^{2+}		r*Pf*A-M1 His^{500} $N\varepsilon_2$	Metal–protein
Zn^{2+}		r*Pf*A-M1 Glu^{519} $O\varepsilon_1$	Metal–protein
Zn^{2+}	O3	O3	Metal–ligand
Zn^{2+}	O2	O4	Metal–ligand
Protein ligand			
Glu^{319} $O\varepsilon_1$	N2	N	H bond
Gly^{460} N	O4	O1	H bond
Ala^{461} N	O4	—	H bond
Glu^{463} $O\varepsilon$	N2/O2	N/O4	H bond
Glu^{497} $O\varepsilon$	O2/N1	O4	H bond
Tyr^{580} OH	O3	O3	H bond
Water HOH	O1	—	H bond
Gln^{317}	P1′ (Phe)	P1	van der Waals
Ala^{320}	—	N	van der Waals
Val^{459}	O4/O1	P1	van der Waals
Ala^{461}	—	C17	van der Waals
Met^{462}	P1′ (Phe)	P1	van der Waals
Thr^{492}	P1 (Leu)	P1′	van der Waals
Val^{493}	—	P1′	van der Waals
Lys^{518}	N2	C19	van der Waals
Tyr^{575}	P1′ (Phe)	C19	van der Waals
Met^{1034}	P1′ (Phe)	P1	van der Waals