Amino Acid Substitution and Modification Resulting from *Escherichia coli* Expression of Recombinant *Plasmodium falciparum* Histidine-Rich Protein II†

Eric L. Schneider,‡ David S. King,§,|| and Michael A. Marletta*,‡,§,⊥

Department of Chemistry, Department of Molecular and Cell Biology, Division of Physical Biosciences, Lawrence Berkeley National Laboratory, and Howard Hughes Medical Institute, University of California, Berkeley, Berkeley, California 94720-1460

Received July 6, 2004; Revised Manuscript Received November 1, 2004

ABSTRACT: The histidine-rich protein II (HRP II) from *Plasmodium falciparum* is an unusual protein composed of 40% alanine, 36% histidine, and 11% aspartate residues. Expression of HRP II in *Escherichia coli* results in the isolation of a heterogeneous protein. Mass spectrometry reveals a reduction in mass by multiples of 9 Da from the expected molecular mass that can be attributed to the substitution of glutamine for some histidine residues in the sequence. The extent of the glutamine for histidine substitution can be reduced by slowing the expression rate. Mass spectral analysis of HRP II also revealed α-amino methylation of the N-terminal alanine residue of HRP II.

Malaria continues to have a major impact on humans, infecting 300−500 million people and resulting in over 1 million deaths every year (*1*). Unfortunately, because of the spread of resistance to the common antimalarial chloroquine, these numbers are increasing (*2*). Antimalarial resistance has forced the search for new drug targets within *Plasmodium*, the protozoan parasite responsible for malaria infections. During the erythrocytic stage of infection, the parasite degrades massive amounts of hemoglobin as a nutrient source (*3*, *4*) and as a result is exposed to high concentrations of toxic, free heme (*5*). Consequently, heme metabolism has been targeted as a strategy to effect selective parasite killing. The histidine-rich protein II (HRP II)[1] from *Plasmodium falciparum*, the most virulent *Plasmodium* species, has been reported to bind between 15 and 18 heme molecules per protein (*6*, *7*, *24*). It has also been shown to initiate in vitro formation of hemozoin, a nontoxic, crystalline form of heme (*6*). Because of the toxic effects of free heme, hemozoin formation is crucial to the survival of the parasite (*8*, *9*).

HRP II is an unusual protein with a highly repetitive primary sequence. The 30 kDa, 277 amino acid protein is composed of 40% alanyl, 36% histidyl, and 11% aspartyl residues contained within the repeats His-His-Ala-His-His-Ala-Ala-Asp-Ala and His-His-Ala-Ala-Asp-Ala throughout the sequence with minor variations in those repeats (*10*, *11*). Expression of HRP II in *Escherichia coli* typically yields 10−20 mg of purified protein/L of culture. The purified protein appears to be homogeneous by SDS−PAGE and reversed-phase HPLC. However, we show here that mass spectral analysis reveals extensive heterogeneity. Furthermore, mass spectral studies show that the unusual amino acid composition of HRP II combined with overexpression in *E. coli* leads to sporadic substitution of glutamine for histidine in the sequence. We have also found the protein to have an unexpected methylation site at the N-terminal alanine residue, adding to a small number of proteins, both endogenous and recombinant, that have been found to be targets for N-methylation by *E. coli* methyltransferases.

MATERIALS AND METHODS

Materials. The perfectly blunt cloning kit and all *E. coli* cell lines were from Novagen. The rapid ligation kit was from Roche. IPTG was purchased from Promega. Restriction enzymes were from New England Biolabs. Sequencing grade chymotrypsin was from Roche. Disposable Sep-Pak C-18 cartridges and C-18 ZipTips were from Millipore. Precast Novex 10−20% Tris−glycine gels were from Invitrogen. FMOC amino acids were from NovaBiochem. All other chemicals were purchased from Sigma-Aldrich.

General Methods. The biuret assay was used according to the method of Gornall et al. (*12*) for the quantitation of HRP II and HRP II(C274S). MS data for HRP II, HRP II(C274S), and the chymotryptic digest fragments were collected with an ESI Bruker Esquire 3000 ion trap. FTICR-MS data for the chymotryptic digest fragments were collected

† This work was supported by the Burroughs-Wellcome Fund, New Initiatives in Malaria Research.

* To whom correspondence should be addressed at the Department of Chemistry, University of California, Berkeley, 211 Lewis Hall, Berkeley, CA 94720-1460. Phone: (510) 643-9325. Fax: (510) 643-9388. E-mail: marletta@berkeley.edu.

‡ Department of Chemistry, UCB.

§ Department of Molecular and Cell Biology, UCB.

|| Howard Hughes Medical Institute, UCB.

⊥ Division of Physical Biosciences, LBNL.

[1] Abbreviations: HRP II, histidine-rich protein II; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; HPLC, high-pressure liquid chromatography; IPTG, isopropyl β-D-thiogalactopyranoside; FMOC, *N*-α-(9-fluorenylmethyloxycarbonyl); MS, mass spectrometry; MS/MS, tandem mass spectrometry; ESI, electrospray ionization; FTICR, Fourier transform ion cyclotron resonance; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; TFA, trifluoroacetic acid; MALDI, matrix-assisted laser desorption/ionization; CID, collision-induced dissociation; Glx, glutamine or glutamic acid; Asx, asparagine or aspartic acid.

10.1021/bi048571h CCC: $30.25
Published on Web 12/24/2004

on a 9.4 T ESI Bruker Apex III instrument. MS/MS data were acquired on an AB4700 tandem TOF instrument.

Construction of the HRP II(C274S) Expression Vector. The HRP II construct in pET3d (*6*) was a gift from Dr. Daniel Goldberg (Washington University). The protein sequence matches the NCBI database entry U69551, except for the deleted hydrophobic leader residues (*6*) and two amino acid differences, a His at position 202 (reported as Arg) and an Arg at position 263 (reported as His, numbering according to NCBI database entry). The HRP II(C274S) construct was made by PCR using a sense primer which included the existing *Nco*I site and ATG start codon (5′-AAGAAG-GAGATATACCATGGCAAAAA-3′). The mutagenic antisense primer deleted the second existing *Nco*I site and introduced a unique *Bam*HI site along with the single underlined base change for the C274S mutation (5′-GG-GATCCTTAATGGCGTAGGCTATGTG-3′). PCR of pET3d-HRP II yielded an 837-bp product, which was inserted into pETBlue-1 blunt vector using the perfectly blunt cloning kit and transformed into NovaBlue competent cells. A pETBlue-1 plasmid containing the HRP II(C274S) gene was digested with *Nco*I and *Bam*HI and ligated into a similarly digested pET21d vector with a rapid ligation kit to make pET21d-SerHRP II. The plasmid DNA was sequenced by the UC Berkeley DNA sequencing facility to verify the presence of the mutation.

Construction of the pET21d-HRP II Expression Vector. The HRP II gene was excised from the pET3d vector using *Xba*I and *Bam*HI. This fragment was ligated into a similarly treated pET21d vector using a rapid ligation kit. The plasmid DNA was sequenced by the UC Berkeley DNA sequencing facility.

Expression and Purification of HRP II and HRP II(C274S). Recombinant HRP II was expressed using the appropriate plasmid [pET3d-HRP II (*6*) or pET21d-HRP II for wild-type HRP II and pET21d-SerHRP II for HRP II(C274S)] in *E. coli* BL21(DE3) or BL21(DE3) pLysS competent cells as noted in Results. Expression with BL21(DE3) competent cells was achieved either without induction or by induction with addition of up to 1 mM IPTG, as specified in Results, at a cell OD_{600} of 0.6−0.8. Expression in BL21(DE3) pLysS competent cells was induced by addition of 5 μM IPTG at a cell OD_{600} of 0.6−0.8. All expressed proteins were purified with a nickel chelating column as previously described (*6*, *13*).

Chymotryptic Digest and Purification of HRP II and HRP II(C274S) Peptide Fragments. Purified HRP II(C274S) at 1 mg/mL was incubated with 1 μg/mL (0.1% w/w) sequencing grade chymotrypsin for time periods ranging from 30 min to 5 h at 25 °C in 100 mM HEPES pH 7.0. Two methods were used to purify the digested fragments. All purified fragments were loaded onto a Sep-Pak C-18 cartridge, eluted with 70% acetonitrile, and concentrated with a SpeedVac concentrator (Savant) prior to MS analysis.

Purification 1: Digested samples were loaded onto a 2 mL Bio-Rad Bio-scale S2 cation-exchange column with 100 mM sodium acetate, pH 4.8, running buffer. The digested fragments were eluted from the column with a gradient of 0−1 M $CaCl_2$ in the running buffer over 40 mL at 2 mL/min using a Bio-Rad Biologic HR Workstation chromatography system. Eluted fragments were detected by absorbance at 280 nm and collected in 1 mL fractions.

Purification 2: Digested samples were loaded onto a Pharmacia Biotech HiLoad 16/60 Superdex 75 prep grade prepacked gel filtration column with a 100 mM HEPES, 200 mM NaCl, pH 7.0, running buffer. The digested fragments were eluted from the column with an isocratic flow of running buffer over 140 mL at 1 mL/min using a Bio-Rad Biologic HR Workstation chromatography system. Eluted fragments were detected by absorbance at 280 nm and collected in 1 mL fractions. Each fraction was run on a 10−20% SDS−PAGE gel to determine the purity of the fragments. The most pure fractions containing each fragment were pooled together.

Mass Spectral Analysis of HRP II, HRP II(C274S), and Chymotrypsin Digest Fragments. Mass spectral analyses of HRP II, HRP II(C274S), and the chymotrypsin digest fragments were performed on a Bruker-Agilent Esquire quadrupole ion trap mass spectrometer with electrospray ionization. Analysis was performed off-line by flow injection analysis at 1 μL/min. An ultrafast microprotein analyzer HPLC (Michrom BioResources, Inc.) was used to elute a 20 μg sample of the full-length proteins from a Michrom Reliasil C-18 column (300 Å pore, 1 mm × 15 cm) using a 10−50% acetonitrile gradient in water with 0.1% TFA over 7 min at 0.75 mL/min. MS data were acquired on the eluent.

FT-ICR analysis was performed off-line on a 4.9 T ESI Bruker Apex III instrument with electrospray ionization. Data were acquired (broad-band spectra, 256K data points, external calibration) with Bruker Daltonics XMASS software.

Quantitative Amino Acid Analysis of the HRP II Chymotrypsin Digest Fragment. HRP II expressed in BL21(DE3) cells with 1 mM IPTG induction and HRP II expressed in BL21(DE3) pLysS cells with 5 μM IPTG induction were digested for 5 h with chymotrypsin, as described above. The small fragments from each digestion were separated from the rest of the digest products using size exclusion chromatography (purification 2). The concentration of the purified fragments was estimated by absorbance of the single tyrosine at 274 nm. Quantitative amino acid analysis was carried out at the Protein Chemistry Laboratory at Texas A&M University using a Hewlett-Packard AminoQuant II system. All analyses were performed in triplicate.

MS/MS Analysis of Chymotrypsin-Digested HRP II. HRP II at 1 mg/mL was digested with chymotrypsin for 5 h following the above protocol. Peptides from the digest solution were adsorbed onto a C-18 ZipTip (Millipore) and eluted with 4 μL of 50% acetonitrile/50% 0.1% aqueous TFA. A 0.7 μL aliquot of this solution was mixed with an equal volume of matrix (a 10 g/L solution of α-cyano-4-hydroxycinnamic acid in 50% acetonitrile/50% 0.1% aqueous TFA) and allowed to dry on the MALDI target.

Using an Applied Biosystem 4700 MALDI tandem TOF mass spectrometer a full scan mass spectrum of the digest was recorded over the range m/z 800−3500, then the protonated peptide ion at m/z 1566.9 was isolated, and a collision- induced dissociation (CID) spectrum was acquired. Air was used as the collision gas, and the collision energy was 1.0 keV (laboratory frame).

Synthesis of Methylated N-Terminal Peptides. The synthetic peptides α-Me-AKNAKGLNLNKRLL and A-(ϵ-Me-K)NAKGLNLNKRLL were prepared using standard FMOC solid-phase synthesis on an automated synthesizer (ABI 431). The mass and purity of the peptides were assessed by FTICR-

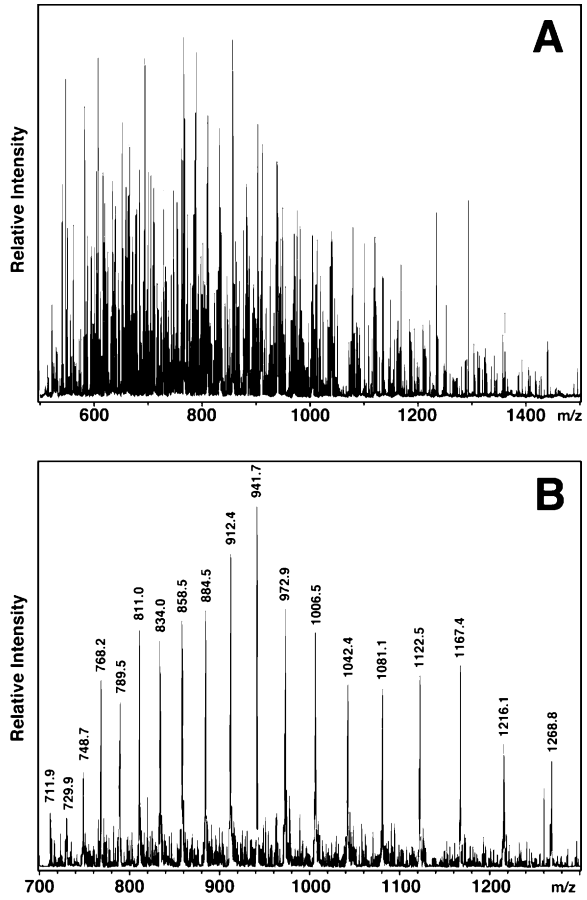

FIGURE 1: ESI mass spectrometry results for HRP II(C274S). Mass spectra are shown for the full-length recombinant HRP II(C274S) expressed in BL21(DE3) with 1 M IPTG induction (A) and the full-length HRP II(C274S) expressed in BL21(DE3) pLysS with 5 μM IPTG induction (B). Panel A shows a heterogeneous SerHRP II sample with estimated average molecular masses ranging from approximately 29033 to 29162 Da. Panel B shows a more homogeneous SerHRP II sample with an average molecular mass of 29162 Da.

MS. MS/MS analysis of these peptides was performed similarly to the chymotrypsin digest fragments, using approximately 0.5 pmol of the synthetic peptide.

RESULTS

Expression of HRP II and HRP II(C274S). Although the SDS−PAGE gel bands were compact and the reversed-phase HPLC peaks sharp, mass spectra of intact HRP II(C274S) (using pET21d-SerHRP II) expressed in BL21(DE3) competent cells with 1 mM IPTG induction revealed a highly heterogeneous mixture of proteins as can be seen in Figure 1A. The mass spectral results for purified recombinant HRP II revealed a similar heterogeneity (data not shown).

Digestion of heterogeneous HRP II(C274S) with chymotrypsin produced three proteolytic fragments as seen by SDS−PAGE, matching the predicted chymotrypsin digestion fragments shown in Table 1. Purification of the chymotrypic digest by cation-exchange chromatography separated the smallest fragment (a combination of the identical 79−108 and 109−138 fragments; see Table 1) from the other two expected fragments of larger mass. Mass spectral analysis of the smallest fragment showed the presence of two peptide species with average masses of 3174.6 and 3165.6 Da as shown in Figure 2, compared with an expected mass of 3183.3 Da. These lower masses are 9 and 18 Da less than the expected average mass.

Analysis of the same small fragment from HRP II by FTICR-MS supported the ESI-MS results shown above where the conclusion was that the sample was heterogeneous. The monoisotopic masses found for this fragment were 3181.4113, 3172.4048, 3163.4048, 3154.4067, and 3145.4064 Da (Figure 3). The species with the highest mass matched correctly with the expected monoisotopic mass at 3181.3767 Da. The remaining species are all spaced by an average value of 9.0012 ± 0.0036 Da, giving species ranging and differing by −9.0065 to −36.0050 Da compared to the highest mass component.

Altered Expression Conditions for HRP II(C274S) and HRP II. In an effort to decrease the heterogeneity of HRP II and HRP II(C274S), we attempted to slow the expression of HRP II(C274S) in *E. coli* BL21(DE3) cells by incrementally decreasing the IPTG concentration. This resulted in very little change to the overall heterogeneity of the purified protein, as assessed by mass spectrometry. Consistently, the extent of protein expression in these cells remained very high, even in the absence of IPTG.

Expression of HRP II(C274S) in BL21(DE3) pLysS competent cells with 5 μM IPTG successfully lowered expression levels, resulting in a more homogeneous protein as shown in the mass spectrum in Figure 1B. HRP II(C274S) expressed in this way had an average mass of 29160.2 Da, compared with an expected average mass of 29277.3 Da. The mass difference is consistent with cleavage of the N-terminal Met, leaving a +14.1 Da difference between the expected average mass minus Met (29146.1 Da) and the average mass determined by MS.

Mass spectral analysis of the smallest chymotryptic fragment after purification by cation-exchange chromatography also revealed a more homogeneous HRP II(C274S) (Figure 4). Approximately 15% of the population, calculated by peak area, was found to have an average mass of 3174.2 with the remaining population having an average mass of 3183.2 Da, which matches the expected average mass for that fragment. There was no evidence for a species at 3165.6 Da, which was found previously in the more heterogeneous samples.

To increase the homogeneity of HRP II similarly to that achieved for HRP II(C274S) above, the HRP II gene was inserted into a pET21d vector and expressed under the same conditions as HRP II(C274S). Mass spectral analysis of the purified HRP II revealed a more homogeneous protein, similar to the results for HRP II(C274S). The average mass determined for HRP II was 29180.0 Da, compared to an expected average mass of 29162.1 Da without the N-terminal Met residue (data not shown). Similar to HRP II(C274S), the measured mass was higher than the expected average mass, in this case by 18 Da.

Mass Spectral Analysis of Digested HRP II(C274S) and HRP II Fragments. Using an S75 16/60 gel filtration column, the chymotryptic digest fragments of HRP II and HRP II(C274S) expressed in BL21(DE3) pLysS were well resolved, as shown in Figure 5. Mass spectral analysis of fragment 139−277 revealed that the four C-terminal residues

Table 1: Expected Chymotrypsin Digestion Fragments of HRP II and HRP II(C274S)[a]

residues	sequence	average mass (Da)	monoisotopic mass (Da)
1-78	MAKNAKGLNLNKRLLHETQAHVDDAHHAHHVADAHHAHHAADAHHAHHAADAHHA–HHAADAHHAHHAADAHHAHHAAY	8481.05	8475.96
2-78	AKNAKGLNLNKRLLHETQAHVDDAHHAHHVADAHHAHHAADAHHAHHAADAHHAH–HAADAHHAHHAADAHHAHHAAY	8349.86	8344.92
79-108	AHHAHHAADAHHAHHASDAHHAADAHHAAY	3183.25	3181.38
109-138	AHHAHHAADAHHAHHASDAHHAADAHHAAY	3183.25	3181.38
139-277	AHHAHHAADAHHAADAHHATDAHHAHHAADAHHATDAHHAADAHHATDAHHAADA–HHAADAHHATDAHHAADAHHATDAHHAADAHHAADAHRATDAHHAHHAADAHHAA–AHHATDAHHATDAHHAAAHHEAATH(**C/S**)LRH	14499.83 (C) 14483.77 (S)	14491.83 (C) 14475.32 (S)
1-277	Full Sequence	29293.34 (C) 29277.28 (S)	29275.98 (C) 29260.00 (S)
2-277	Full Sequence minus Met	29162.14 (C) 29146.08 (S)	29144.94 (C) 29128.96 (S)

[a] The expected digestion fragments for HRP II and HRP II(C274S) are shown along with the expected mass for each fragment based on the DNA sequence. Sequences for fragments 79−108 and 109−138 are identical, resulting in only three different fragments from a chymotrypsin digest. Masses for fragment 139−277 are shown for both the wild type and C274S mutant, as indicated by the residue in parentheses.

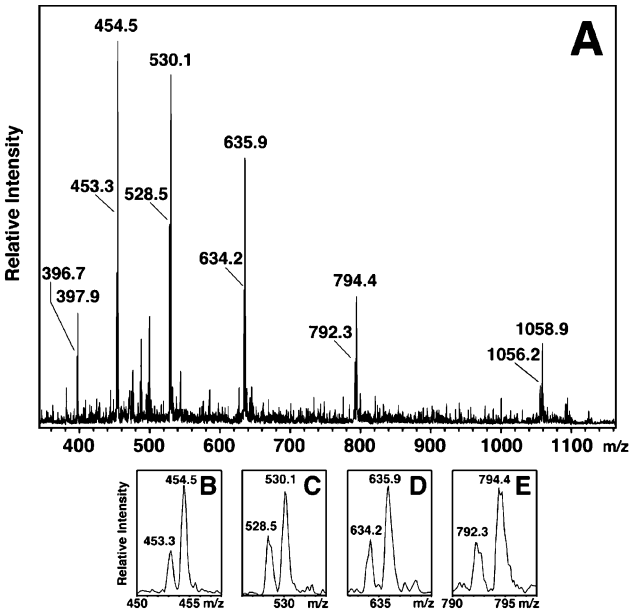

FIGURE 2: ESI mass spectrometry results for a chymotrypsin digest fragment of HRP II(C274S). Panel A shows the full spectrum for the small digest fragment. Two species are readily apparent, corresponding to average masses of 3174.6 and 3165.6 Da. Panels B−E show specific *m*/*z* peaks. The relative peak size for the *m*/*z* values corresponding to the two major species can be readily compared.

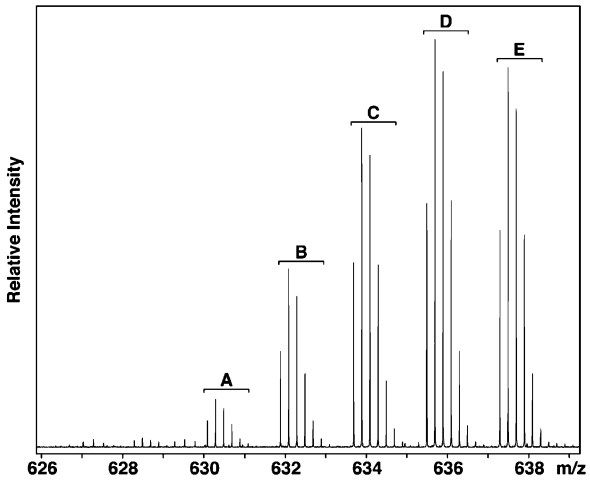

FIGURE 3: FTICR-MS results for a chymotrypsin digest fragment of HRP II. The results reveal the presence of five species, sequentially differing from each other by an average of 9.0012 ± 0.0036 Da. The masses determined were (A) 3145.4064, (B) 3154.4067, (C) 3163.4048, (D) 3172.4048, and (E) 3181.4113. These species indicate up to five His to Gln substitutions. A small population centered around *m*/*z* 628.5 may represent a species with six His to Gln substitutions.

were cleaved, resulting in the average masses of 13991.6 and 13990.9 Da for HRP II(C274S) and HRP II, respectively. These results showed no difference from the expected masses.

Mass spectral analysis results for the purified 2−78 fragment were dependent upon the digestion time. Digestion times greater than 1 h gave an average mass of 7667.0 Da, indicating cleavage between Leu 8 and Asn 9 (expected mass for residues 9−78 of 7667.7 Da). A shorter, half-hour digestion time resulted in the expected 2−78 fragment. ESI-MS results found the fragment to have an average mass of 8364.3 Da (Figure 6), compared to the expected average mass of 8349.9 Da. The difference shows an additional 14.4 Da present on the 2−78 fragment. In addition, the unexpected cleavage between residues 8 and 9 further localizes the additional mass to residues 2 through 8.

FTICR-MS analysis of a 1 h digest product containing the same 2−78 fragment gave a monoisotopic mass of 8359.0554 Da, which is 14.1386 Da larger than the expected monoisotopic mass of 8344.9168 (data not shown). This difference closely matches the expected monoisotopic mass change of 14.0151 Da for the substitution of hydrogen (−H) with a methyl (−CH_3) group.

Quantitative Amino Acid Analysis of Chymotryptic Digests. Quantitative amino acid analysis results for the 3182 Da fragments of HRP II expressed in BL21(DE3) cells with 1 mM IPTG (fragment A) and HRP II expressed in BL21(DE3) pLysS with 5 μM IPTG (fragment B) are shown in Figure 7. Both HRP II fragments A and B contained the expected number of Asx [3.08 ± 0.25 (fragment A) and 2.67 ± 0.02 (fragment B) versus 3 (expected)], Ala [13.65 ± 0.06 (A)

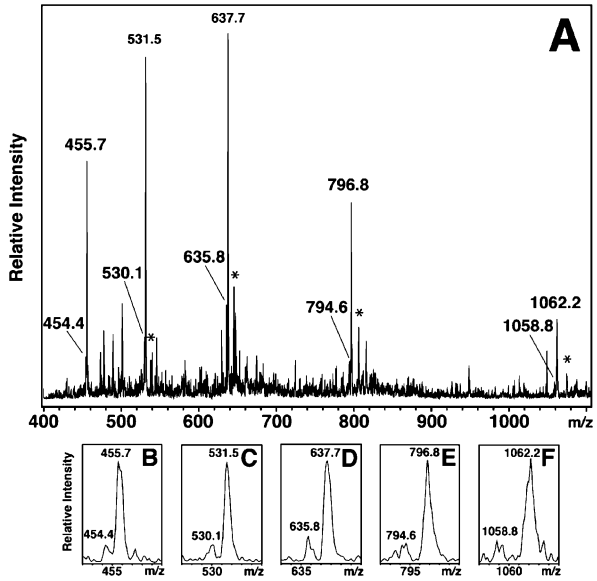

FIGURE 4: ESI mass spectrometry results for a more homogeneous small chymotrypsin digest fragment of HRP II(C274S). Panel A shows the full spectrum for the small digest fragment. Two species are readily apparent, corresponding to average masses of 3183.2 and 3174.2 Da. The peaks indicated by asterisks correspond to *m*/*z* values corresponding to peptide plus potassium (3222 Da). Panels B−F show specific *m*/*z* peaks for comparison of the relative peak sizes.

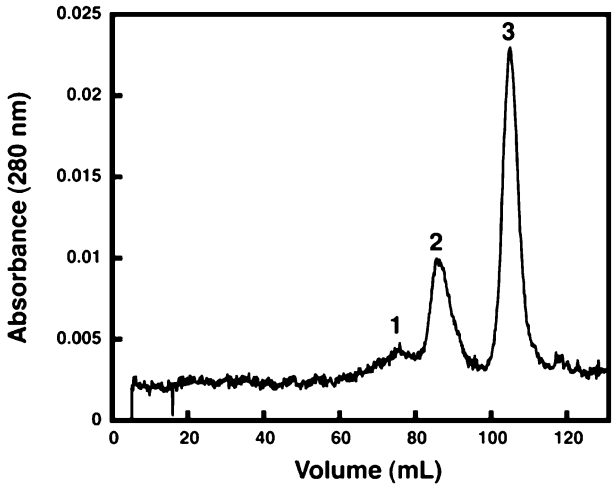

FIGURE 5: Separation of HRP II(C274S) chymotrypsin digest fragments by size exclusion chromatography. Peak 1 corresponds to the fragment containing residues 139−277, peak 2 corresponds to residues 2−78, and peak 3 corresponds to the identical fragments containing residues 79−108 and 109−138. Fragment 139−277 contains no chromophore (as shown in Table 1), resulting in the very small size of peak 1. The relative sizes of peak 2 and peak 3 relate to the fact that peak 3 contains two identical digest fragments.

and 13.46 ± 0.07 (B) versus 13], and Tyr [0.98 ± 0.07 (A) and 1.09 ± 0.04 (B) versus 1]. The number of His and Glx residues did not match the expected values. Fragment A contained 9.31 ± 0.64 His and 0.96 ± 0.07 Glx whereas fragment B contained 11.58 ± 0.30 His and 0.22 ± 0.01 Glx. The expected values were 12 His and 0 Glx (see sequence in Table 1). These results show a larger than predicted presence of Glx in fragment A than in fragment B and a lower than predicted presence of His.

MS/MS Analysis of the HRP II N-Terminus. HRP II that had been expressed more slowly to minimize heterogeneity was digested with chymotrypsin for 5 h in order to produce

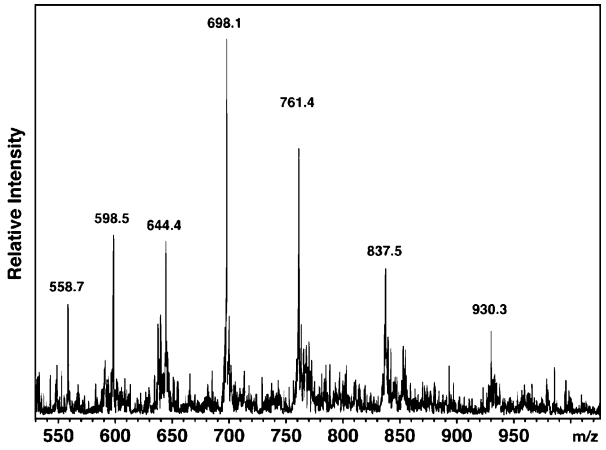

FIGURE 6: ESI-MS spectrum of the N-terminal digest fragment of HRP II(C274S). The digest was carried out for 0.5 h. Indicated *m*/*z* values correspond to an average mass of 8364.30 Da, 14.44 Da higher than the average mass predicted from the amino acid sequence.

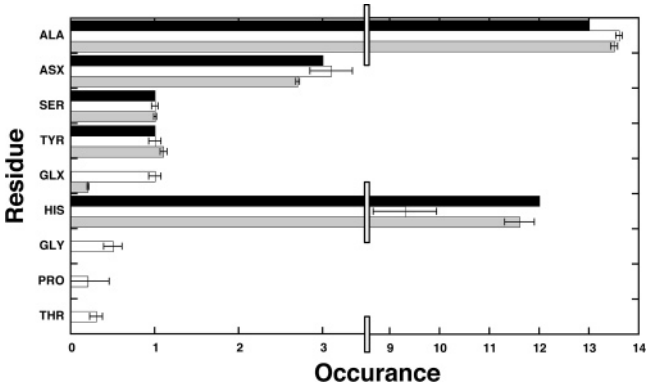

FIGURE 7: Comparison of quantitative amino acid analysis results for the small chymotrypsin digest fragments. QAA results for the small chymotrypsin digest fragment from heterogeneously expressed HRP II(C274S) (fragment A in text) (light gray) and the more homogeneously expressed HRP II(C274S) (fragment B in text) (dark gray) are shown with the expected composition (black).

small peptide fragments. One of these fragments was then analyzed by tandem mass spectrometry. A parent $[M + H]^+$ ion at *m*/*z* 1566.9 potentially corresponded to a methylated N-terminal sequence of HRP II (residues 2−15 with +14 Da). MS/MS of this ion generated the daughter ions shown in Figures 8 and 9B. The fragmentation pattern confirmed the sequence to be that of residues 2−15 of HRP II. A daughter ion at *m*/*z* 58 corresponded to a methylated Ala (a_1 fragment ion, Figures 8 and 9B).

MS/MS Analysis of the Synthetic Peptides. The identities of peptides N-Me-AKNAKGLNLNKRL (MeA-peptide) and A-ϵ-Me-KNAKGLNLNKRL (MeK-peptide) were confirmed by mass spectral analysis. The MS/MS of both of these peptides are shown in Figure 9A). MeK-peptide and MeA-peptide show similar fragmentation patterns as expected. However, only the MeA peptide produces a daughter ion at *m*/*z* 58.07, supporting the assignment of this *m*/*z* to a Me-Ala immonium ion.

DISCUSSION

Expression of HRP II and HRP II(C274S). As shown above, the mass spectral analysis of HRP II and HRP II(C274S) showed that the expressed protein was extremely

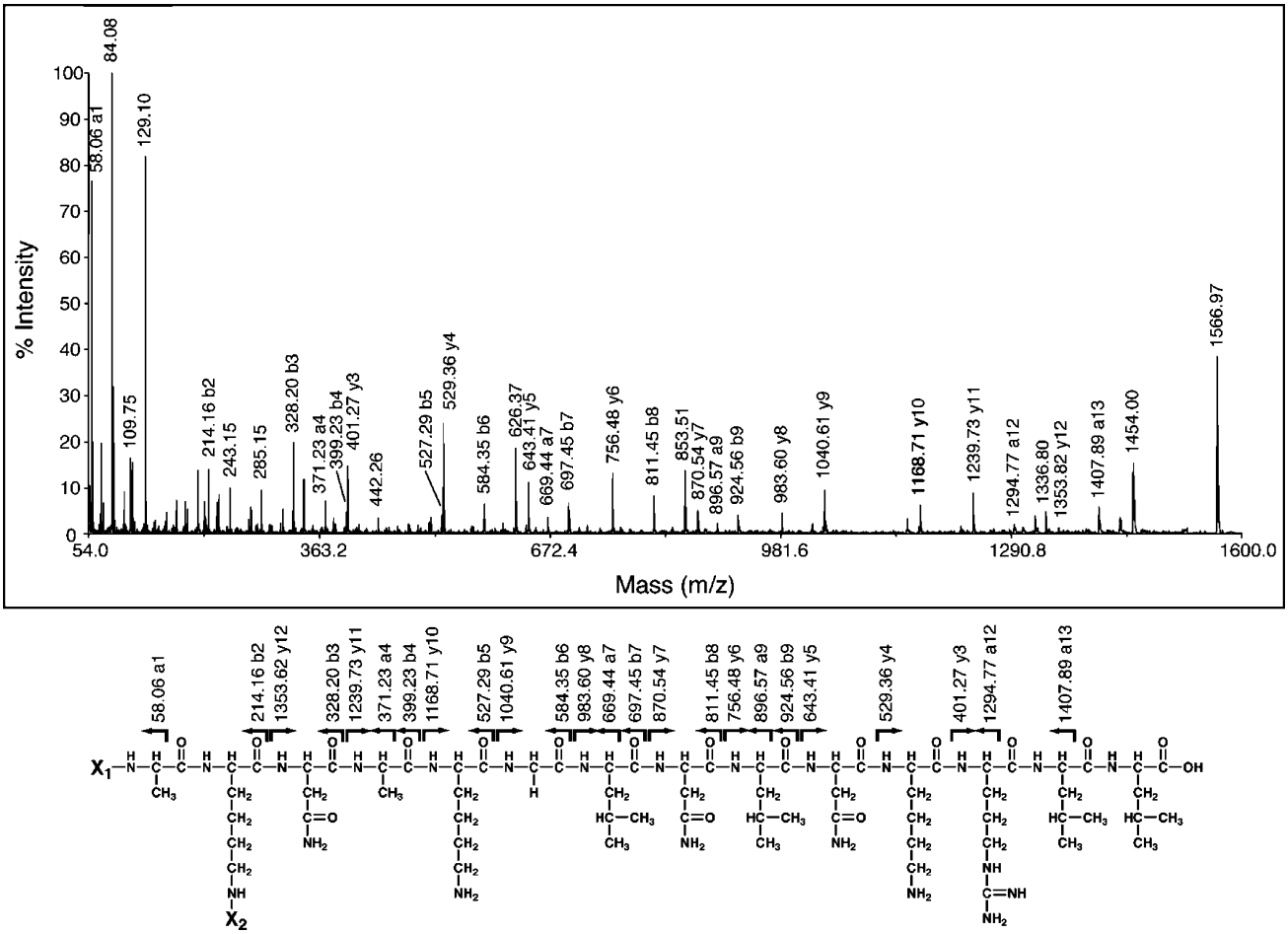

FIGURE 8: MS/MS results for the 1566.9 Da fragment from the chymotrypsin digest of HRP II(C274S). The fragmentation pattern for the 1566.9 Da fragment from HRP II(C274S) after a 5 h chymotrypsin digest is shown with many of the *m*/*z* values indicated above the corresponding peaks. Below the mass spectrum is the sequence corresponding to the 1566.9 Da peptide. A few of the fragments corresponding to *m*/*z* values found in the mass spectrum are indicated above the peptide with bent arrows starting at the fragmentation site and pointing to the correct terminus. The corresponding *m*/*z* value is indicated above the arrow. For the *N*-methylalanine peptide, X_1 is $-CH_3$ and X_2 is $-H$. For the methyllysine peptide, X_1 is $-H$ and X_2 is $-CH_3$. As noted in the text, only one fragment indicated (*m*/*z* 58.06) has a dependence on which X corresponds to the $-CH_3$.

heterogeneous (Figure 1). Initial studies to determine the cause of the heterogeneity were performed with HRP II(C274S) because previous studies indicated the potential for the formation of disulfide-linked dimers in wild-type HRP II, which could potentially complicate interpretation of the resulting data (*24*). To probe the nature of the heterogeneity, HRP II(C274S) was digested into four smaller fragments with chymotrypsin, two of which are exactly the same, resulting in three distinct peptides (Table 1). The resulting smaller peptides simplified the search for the cause of the heterogeneity. After purification by cation-exchange chromatography, MS analysis of the smallest chymotryptic fragment (a mixture of the identical peptides formed from residues 79−108 and 109−138; see Table 1) from HRP II(C274S) revealed heterogeneity similar to that found in the full-length protein. As seen in Figure 2, discrete populations corresponding to masses of 3174.6 (−8.7 Da from the expected mass) and 3165.6 Da (−17.7 Da) were found in the sample.

A change in mass of −9 Da could correspond to a substitution of glutamine or lysine for histidine in the sequence. Given the histidine-rich nature of HRP II, a random number of these substitutions for His throughout the protein would result in a heterogeneous mixture with masses varying by 9 Da, similar to that found for the small digest fragment. To differentiate between substitution by Gln or Lys, FT-ICR studies were performed on the small digest fragment (residues 79−108/109−138) (Figure 3). Five masses were found for the peptide, one matching the expected exact mass, with the other four sequentially differing by an average of 9.0012 ± 0.0036 Da. The mass change matches that expected for a His to Gln substitution. The expected difference in exact mass between His and Gln is 9.000 Da, whereas the expected difference between His and Lys is 8.964 Da, lower than what was determined for these peptides, indicating that the mass change is due to Gln, not Lys, substitution for His.

Quantitative amino acid analysis (QAA) was used to further confirm the substitution of glutamine for histidine. The small chymotryptic fragments (residues 79−108 and 109−138) were analyzed because of the high histidine content (40%, 12/30 residues) and lack of any glutamine or glutamate residues. The content of alanine, aspartate, serine, and tyrosine determined for the small peptide from heterogeneous HRP II was very close to the expected values, with some minor, unexpected amounts of proline, glycine, and threonine (0.5 or less per fragment) (Figure 7). The peptide was also found to contain 0.98 ± 0.04 Glx compared to an expected Glx content of zero. Correspondingly, the peptide contained 9.31 ± 0.64 His compared to 12 expected His

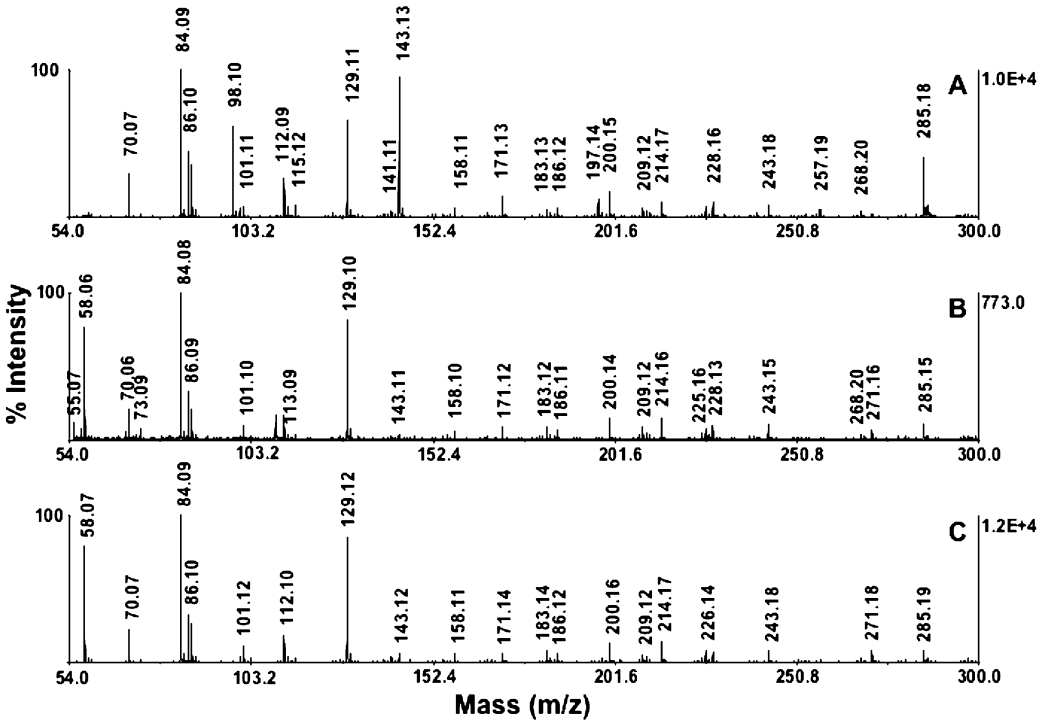

FIGURE 9: Comparison of the MS/MS between the synthetic methylated peptides and the HRP II(C274S) digest peptide. The fragmentation patterns for the synthetic Me-Lys peptide (A) and the synthetic Me-Ala peptide (C) are compared directly with the small HRP II(C274S) peptide resulting from a 5 h chymotrypsin digest. The *m*/*z* 58 corresponding to a fragment of *N*-methylalanine is present only in spectra B and C, whereas the *m*/*z* 98 corresponding to a fragment of methyllysine is only present in spectrum A.

residues. The finding of ~1 Glx per peptide by QAA compares favorably with the ESI-MS results suggesting an average of approximately 1.3 His to Gln substitutions per peptide. The substitution of Gln for His should have resulted in a Glx content value equal to the difference between the expected and determined values for His content. Although these values were not found to be equal (1 Glx versus 2.7 His), the detection of a decrease in His residues along with the subsequent appearance of Glx supports the hypothesis that the heterogeneity is a result of the substitution of some histidine residues with glutamine residues.

Because of the repetitiveness of the sequence, a substitution of Gln for His could arise through mutation to the plasmid. However, a mutation was not found after subsequent plasmid sequencing. Others have reported amino acid misincorporation during translation, including histidine to glutamine (*14*−*16*). These substitutions have been mainly attributed to translational errors stemming from the incorrect binding of charged tRNA (*17*). The substitution of glutamine for histidine has been hypothesized to occur because the codons differ only at the wobble position (CAC and CAU for His versus CAG and CAA for Gln), allowing charged $tRNA^{Gln}$ to bind at His codons under certain conditions. Studies showing an increase in the misincorporation of glutamine at histidine codons under $tRNA^{His}$ starvation conditions lend support to this hypothesis (*15*, *18*). The opposite, misincorporation of histidine at glutamine codons under $tRNA^{Gln}$ starvation conditions, has also been reported (*17*). Because the initial conditions used to express HRP II in *E. coli* produce large quantities of HRP II, it is likely that $tRNA^{His}$ starvation occurs during HRP II overexpression, creating a strain on the cellular $tRNA^{His}$ levels and allowing for the incorrect binding of $tRNA^{Gln}$ at His codons.

The extent of the amino acid substitution in recombinant HRP II expressed in *E. coli* appears to be very high. The small digest fragment (30 amino acids) described above contains 12 histidine residues (see Table 1), of which approximately 1.3 were found by ESI-MS to be substituted with glutamine. Overall, this equates to ~0.11 glutamine misincorporations per histidine residue in the population of this small fragment. If this frequency is taken as an indication of the average misincorporation frequency throughout the full-length HRP II, then approximately 10 histidine residues (of the 99 total) per HRP II would have been substituted with glutamine. These values are about half of the estimation made by Parker et al. for histidine to glutamine substitution under induced histidine starvation condition (0.23 per histidine residue) (*18*). However, our findings are much higher than the histidine to glutamine substitution frequencies reported by Lu et al. in *E. coli* expressed recombinant human granulocyte colony stimulating factor (rhG-CSF) under normal growth conditions, which they report to be in line with the translational errors of 2×10^{-3} to 2×10^{-4} found in normally growing *E. coli* (*19*). The failure of this rhG-CSF to overwhelm the translational machinery and increase the error frequency is not surprising, considering the more typical amino acid composition when compared to HRP II. However, the fact that the substitutions of glutamine for histidine are found under normal growth conditions indicates the possibility of $tRNA^{Gln}$ substituting for $tRNA^{His}$.

Altering Expression Conditions for HRP II. If the misincorporation of glutamine for histidine is the direct result of $tRNA^{His}$ starvation during HRP II overexpression, then slowing HRP II expression should prevent, or at least reduce, the occurrence of the error. Initial trials at slowing HRP II(C274S) expression were attempted with concentrations as

low as 5 μM IPTG, resulting in similarly heterogeneous samples. Complete exclusion of IPTG still resulted in high levels of HRP II expression with the same heterogeneity in the purified sample. Examination of the expression levels on a qualitative level indicated that the decrease, and even exclusion, of IPTG in the cell cultures appeared to have minor effects on the expression level of HRP II when expressed from the pET vector in BL21(DE3) cells.

An *E. coli* strain containing the pLysS plasmid allows for greater control of protein expression from the T7 promoter. Expression of HRP II(C274S) from the pET21d-HRP II(C274S) plasmid in these BL21(DE3)pLysS cells resulted in lowered protein expression levels. Using 5 μM IPTG at induction, the resulting HRP II(C274S) was much more homogeneous (Figure 1B). After accounting for cleavage of the N-terminal methionine, the mass spectral results indicated HRP II(C274S) was 14.1 Da higher than expected. After resequencing the plasmid DNA to ensure that the predicted amino acid sequence was correct, the additional mass was suggestive of a methylated amino acid.

Mass spectral analysis of the small chymotryptic fragment from this more homogeneous HRP II(C274S) was performed to gauge the extent to which glutamine misincorporation had been successfully inhibited. The results indicated only partial success, in that approximately 15% of the fragment population had one histidine to glutamine substitution with no indication of any with two substitutions (Figure 4). Further analysis of this fragment by QAA was also consistent with incomplete, but still significant, success in preventing glutamine substitutions (Figure 7). As with the heterogeneous HRP II peptide, the QAA results for this more homogeneous peptide showed alanine, aspartate, serine, and tyrosine content to be very close to the expected values. However, unlike the heterogeneous HRP II peptide the His and Glx content in this peptide was found to be much closer to the expected levels. Only 0.22 ± 0.01 Glx was found per peptide, just slightly higher than the expected 0 Glx. Similarly, the results for His content were very close to the expected results; 11.58 ± 0.30 His per peptide were found compared to the expected 12 His per peptide. As with the heterogeneous peptide, these values found by QAA compare very well with values estimated by mass spectrometry. The mass spectral results indicated close to 0.15 Glx per peptide compared to 0.22 Glx per peptide found by QAA, corresponding to substitution frequencies of 0.013 and 0.018 per histidine, respectively.

Although both the mass spectral and QAA results indicate that histidine to glutamine substitution has not been completely prevented, the results do show that the occurrence is significantly lower than in the original population, where nearly 100% of the small digest peptides had at least one glutamine substitution and approximately 30% had two. The mass spectral results of the homogeneous HRP II(C274S) indicate that one glutamine substitution in 15% of the population equates to a substitution frequency of 0.013 per histidine in each fragment. From this substitution frequency, the average full-length HRP II(C274S) can be expected to have approximately one histidine to glutamine substitution per protein under these new expression conditions. This frequency of substitution is still ~100 times greater than the frequency found for endogenous cellular proteins, but indicates a 10-fold decrease from our original findings. The decrease in the frequency of substitution does, however, lend support to the hypothesis that the high expression levels of HRP II coupled with the unusually high content of His residues led to amino acid substitutions.

To achieve a more homogeneous HRP II expression similar to that shown above for HRP II(C274S), the HRP II gene was transferred from the pET3d vector to a pET21d vector. Expression in BL21(DE3) pLysS cells under the same conditions as for HRP II(C274S) produced a more homogeneous HRP II sample found by ESI-MS to have a cleaved N-terminal Met with an additional 18 Da compared to the expected average mass, corroborating with the initial suggestion of a methylated residue on HRP II(C274S).

Identification of the Methylation Site. Purification of the chymotrypsin digest fragments of HRP II and HRP II(C274S) by size exclusion chromatography provided an avenue to determine the site of methylation in the protein. As described above, the examination of the smallest digestion fragments (79−108 and 109−138) showed no unexpected mass increase (Figure 4). Turning to the two remaining fragments, ESI-MS revealed that the C-terminal fragments (139−277) had average masses of 13991.6 and 13990.9 Da for HRP II(C274S) and HRP II, respectively, corresponding to the unexpected, although not entirely surprising (*20*), cleavage after His 273, removing the four residues from the C-terminus. These results indicated that the methylation site was not within residues 139−273 but did not rule out the chance that one of the four C-terminal residues (CLRH or SLRH, depending on the protein) contained the additional mass.

Initial mass spectral examination of the N-terminal fragment after a 5 h chymotrypsin digest revealed that it was unexpectedly cleaved after residue Leu 8. The remaining fragment, residues 9−78, showed no additional mass, suggesting that the potentially methylated amino acid was located either on residues 2−8 or residues 274−277. Shortening the digest time to 30 min allowed for the purification of the full N-terminal fragment (residues 2−78). ESI-MS results of this complete N-terminal peptide revealed an additional 14.4 Da (Figure 6), pinpointing the additional mass within residues 2−8. FTICR-MS of this fragment confirmed the presence of an additional 14 Da matching the expected mass change for the substitution of a hydrogen with a methyl group.

The addition of 14 Da is suggestive of a methylated amino acid. Potential sites within the identified sequence (AKNAKGL) include the amino group of the N-terminal alanine and the ϵ-amino groups of the two lysine residues. MALDI-MS/MS analysis of a 5 h digest sample of SerHRP II identified a fragment at 1566.9 Da that matched a methylated fragment composed of residues 2−15. Although the corresponding daughter ions confirmed the identification of this fragment (Figure 8), they were ambiguous as to the exact residue that was methylated. All daughter ions that included the N-terminal alanine were +14 Da higher than predicted. However, all of these fragments also included the Lys at position 3, precluding the assignment of the methyl group to the N-terminal alanine. Only the presence of an ion peak at m/z 58 indicated the presence of an *N*-methylalanine residue, corresponding to an *N*-methylalanine immonium ion. An ion corresponding to the loss of *N*-methylalanine from the rest of the peptide was not found,

leaving the possibility that the *m*/*z* 58 might have arisen from some other fragmentation of the parent ion. However, the data show that the methyl group is positioned on either Ala 2 or Lys 3, with the most evidence pointing to Ala 2.

To confirm the site of methylation, synthetic peptides corresponding to residues 2−8 were prepared with either an α-Me-Ala or ϵ-Me-Lys, and CID spectra were obtained (Figure 8). The results show the fragmentation pattern of the synthetic *N*-methylalanine peptide to be nearly identical with the digest fragment from HRP II (Figure 9). The *m*/*z* 58 tentatively assigned to an *N*-methylalanine immonium ion is present in the *N*-methylalanine peptide spectrum but is clearly absent from the methyllysine peptide spectrum. Furthermore, *m*/*z* 98 corresponding to a methyllysine immonium ion is present in the spectrum of the methyllysine peptide but not the spectra corresponding to the *N*-methylalanine peptide or the HRP II peptide. These synthetic peptides establish the methylation site as the amino group of the N-terminal alanine.

Only a few N-methylated proteins have been found in *E. coli*, and most of those are ribosomal proteins (*21*, *22*). Similar to HRP II, a number of these N-methylated proteins have a Lys as the third amino acid. Another group has also identified an N-methylated methionine on a recombinant version of a malarial antigen expressed in *E. coli* (*23*). In this case the methionine is also followed by a basic residue, arginine. It is likely that recombinant HRP II has the proper recognition site for the methyltransferase(s) that identifies these proteins.

Although the findings described here may not be indicative of in vivo translation and processing of HRP II by *P. falciparum*, it is clear that, although extremely powerful, expression of recombinant proteins in *E. coli* can have unexpected results. The overexpression of the very repetitive HRP II sequence resulted in a high level of translational errors. Along with these errors, the consistent methylation of the N-terminal alanine was also unpredicted.

ACKNOWLEDGMENT

We thank Dr. A. M. Falick for MALDI-MS/MS work and H. H. Chang for PCR work on HRP II(C274S).

SUPPORTING INFORMATION AVAILABLE

One table containing the *m*/*z* values for Figure 3 and FTICR-MS data for the 2−78 fragment along with a corresponding table of *m*/*z* values. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. WHO (2002) *World Health Organization Report on Infectious Diseases: Scaling up the Response to Infectious Diseases; A Way Out of Poverty*, World Health Organization, Geneva, Switzerland.
2. Ursos, L. M., and Roepe, P. D. (2002) Chloroquine resistance in the malarial parasite, *Plasmodium falciparum*, *Med. Res. Rev. 22*, 465−491.
3. Meshnick, S. R. (1996) Is haemozoin a target for antimalarial drugs?, *Ann. Trop. Med. Parasitol. 90*, 367−372.
4. Goldberg, D. E., Slater, A. F., Cerami, A., and Henderson, G. B. (1990) Hemoglobin degradation in the malaria parasite Plasmodium falciparum: an ordered process in a unique organelle, *Proc. Natl. Acad. Sci. U.S.A. 87*, 2931−2935.
5. Rosenthal, P. J., and Meshnick, S. R. (1996) Hemoglobin catabolism and iron utilization by malaria parasites, *Mol. Biochem. Parasitol. 83*, 131−139.
6. Sullivan, D. J., Jr., Gluzman, I. Y., and Goldberg, D. E. (1996) *Plasmodium* hemozoin formation mediated by histidine-rich proteins, *Science 271*, 219−222.
7. Pandey, A. V., Bisht, H., Babbarwal, V. K., Srivastava, J., Pandey, K. C., and Chauhan, V. S. (2001) Mechanism of malarial haem detoxification inhibition by chloroquine, *Biochem. J. 355*, 333−338.
8. Fitch, C. D., Chevli, R., Banyal, H. S., Phillips, G., Pfaller, M. A., and Krogstad, D. J. (1982) Lysis of *Plasmodium falciparum* by ferriprotoporphyrin IX and a chloroquine-ferriprotoporphyrin IX complex, *Antimicrob. Agents Chemother. 21*, 819−822.
9. Chou, A. C., and Fitch, C. D. (1981) Mechanism of hemolysis induced by ferriprotoporphyrin IX, *J. Clin. Invest. 68*, 672−677.
10. Wellems, T. E., and Howard, R. J. (1986) Homologous genes encode two distinct histidine-rich proteins in a cloned isolate of *Plasmodium falciparum*, *Proc. Natl. Acad. Sci. U.S.A. 83*, 6065−6069.
11. Panton, L. J., McPhie, P., Maloy, W. L., Wellems, T. E., Taylor, D. W., and Howard, R. J. (1989) Purification and partial characterization of an unusual protein of *Plasmodium falciparum*: histidine-rich protein II, *Mol. Biochem. Parasitol. 35*, 149−160.
12. Gornall, A. G., Bardawill, C. J., and David, M. M. (1949) Determination of serum proteins by means of the biuret reaction, *J. Biol. Chem. 177*, 751−766.
13. Choi, C. Y., Cerda, J. F., Chu, H. A., Babcock, G. T., and Marletta, M. A. (1999) Spectroscopic characterization of the heme-binding sites in *Plasmodium falciparum* histidine-rich protein 2, *Biochemistry 38*, 16916−16924.
14. Parker, J. (1989) Errors and alternatives in reading the universal genetic code, *Microbiol. Rev. 53*, 273−298.
15. O'Farrell, P. H. (1978) The suppression of defective translation by ppGpp and its role in the stringent response, *Cell 14*, 545−557.
16. Scorer, C. A., Carrier, M. J., and Rosenberger, R. F. (1991) Amino acid misincorporation during high-level expression of mouse epidermal growth factor in *Escherichia coli*, *Nucleic Acids Res. 19*, 3511−3516.
17. Ulrich, A. K., Li, L. Y., and Parker, J. (1991) Codon usage, transfer RNA availability and mistranslation in amino acid starved bacteria, *Biochim. Biophys. Acta 1089*, 362−366.
18. Parker, J., Pollard, J. W., Friesen, J. D., and Stanners, C. P. (1978) Stuttering: high-level mistranslation in animal and bacterial cells, *Proc. Natl. Acad. Sci. U.S.A. 75*, 1091−1095.
19. Lu, H. S., Fausset, P. R., Sotos, L. S., Clogston, C. L., Rohde, M. F., Stoney, K. S., and Herman, A. C. (1993) Isolation and characterization of three recombinant human granulocyte colony stimulating factor His→Gln isoforms produced in *Escherichia coli*, *Protein Expression Purif. 4*, 465−472.
20. Keil, B. (1992) *Specificity of Proteolysis*, Springer-Verlag, New York.
21. Apostol, I., Aitken, J., Levine, J., Lippincott, J., Davidson, J. S., and Abbott-Brown, D. (1995) Recombinant protein sequences can trigger methylation of N-terminal amino acids in *Escherichia coli*, *Protein Sci. 4*, 2616−2618.
22. Stock, A., Clarke, S., Clarke, C., and Stock, J. (1987) N-terminal methylation of proteins: structure, function and specificity, *FEBS Lett. 220*, 8−14.
23. Lahnstein, J., Dyer, S., Goss, N., Duncan, M., and Norton, R. (1993) in *Techniques in Protein Chemistry*, Academic Press, New York.
24. Schneider, E. L., and Marletta, M. A. (2005) Heme binding to the histidine-rich protein II from *Plasmodium falciparum*, *Biochemistry 44*, 987−995.

BI048571H