We report here a novel finding that norvaline can be incorporated in place of leucine in recombinant human hemoglobin expressed in Escherichia coli. The presence of the norvaline was confirmed by several analytical methods such as amino acid analysis, peptide mapping, electrospray mass spectrometry, and Edman protein sequencing. It appears that substitution is distributed across both the β- and di-α-globins in purified recombinant hemoglobin. The level of misincorporation correlated with the ratio of the free norvaline/leucine pool available in the cell culture. This suggests that the incorporation of norvaline for leucine occurs through misaminoacylation of tRNA\textsubscript{Leu}, similar to the misincorporation of norleucine found in many recombinant proteins expressed in E. coli.

The best known example of misincorporation of uncoded amino acids into recombinant proteins is substitution of norleucine for methionine. This phenomenon has been found in many recombinant proteins and is well documented in the literature (1–9). Norleucine can be substituted for methionine in proteins both at internal residues and at the initiating amino terminus. Bogosian (1) postulated that high level synthesis of leucine-rich proteins may promote the biosynthesis of norleucine in bacteria. Norleucine can then become randomly incorporated into cloned proteins at methionine residues. The incorporation of norleucine occurs through misaminoacylation (mischarging) during protein synthesis.

Norvaline can be intentionally incorporated into proteins either by semi-synthesis at designed positions (10–13) or by feeding microorganisms expressing heterologous proteins with norvaline (14). Also, it has been reported that norvaline is a natural component of an antifungal peptide produced by Bacillus subtilis (15). We report here that a low but readily detectable level of norvaline can be naturally misincorporated into recombinant hemoglobin in place of leucine. To determine the extent of Nva\textsuperscript{1} incorporation, and establish which of the 72 leucine residues were involved in the misincorporation, we undertook the task of detailed structural characterization of (putatively) norvaline-enriched samples of rHb1.1. We used a variety of analytical tools such as amino acid analysis, peptide mapping, Edman sequencing, and a several mass spectrometry techniques. Our results demonstrate that the misincorporation of norvaline is affected by fermentation conditions. The extent to which misincorporation occurs strongly correlates with the ratio of free norvaline to leucine found in the total pool of lysed bacterial cells in fermentation broth. In purified recombinant hemoglobin, the misincorporation appears to occur specifically at leucine positions and is distributed across both the β- and α-globins. We hypothesize that norvaline misincorporation may occur in other heterologous proteins expressed in Escherichia coli under conditions where norvaline can be produced and accumulated during the expression phase.

Recombinant hemoglobin (rHb1.1) expressed in E. coli is a human hemoglobin stabilized by the fusion of two α chains through a glycine linking the carboxyl terminus of the first α chain to the amino terminus of the second α chain (di-α-globin). At the amino termini of both the di-α- and β-globin chains of rHb1.1, the wild type valine residues have been replaced with methionine residues. In addition, rHb1.1 also contains the Presbyterian mutation (Asn\textsubscript{108} → Lys) in the β chain of the molecule which confers reduced oxygen affinity (17–19). Recombinant hemoglobin contains 72 leucine residues (12.5%) out of 575 total amino acids. In contrast, typical E. coli proteins contain on average 8.4% of leucine (16). That distinction classifies hemoglobin as a leucine-rich protein, which under certain conditions appears to trigger norleucine and norvaline production in E. coli in a manner similar to that postulated by Bogosian (1).

**EXPERIMENTAL PROCEDURES**

**Materials**—Amino acid standards for amino acid analysis were from Pierce. Reagents for OPA amino acid analysis were from Hewlett-Packard. AccQ-Tag chemistry package was purchased from Waters. Acetone, and 4-vinylpyridine were from Aldrich. Proteolytic enzymes were used as follows: V8 was from Pierce and trypsin was from Worthington, and Prozyme cartrige of immobilized trypsin was purchased from PerSeptive Biosystems. Restriction enzymes and DNA ligase were purchased from New England Biolabs. All sequencing reagents for the Porcon sequecer were from Beckman. Horse heart myoglobin, norvaline, and formic acid were purchased from Sigma. HPLC-grade acetonitrile and water were from Burdick and Jackson, and general chemicals and buffer components were obtained from Amresco.

**Expression and Purification**—The expression system and purification methods used to produce rHb1.1 have been described (17–20).

**Rare Leucine Codon Replacement**—The leucine codons at positions 110 and 141 in the β-globin were replaced by polymerase chain reaction amplification of the carboxyl-terminal region of β using two DNA primers in which the CUA codons were replaced by a sequence creating CUG codons. The primers were CBG111 (5'-GGTTAAGTACTGTTTGCGTT-3') and CBB112(5'-AAAACAGAGGTCATGTGGATTAGTCGACAGCGCGTTAGC-3'). The polymerase chain reaction product was digested with restriction enzymes Sau3A1 and HindIII and ligated into the identically digested β-globin chain in plasmid pSGB729, replacing the original coding sequence with the rare codons for leucine. These leucine codon mutations removed SpeI and Nhel restriction sites so clones were digested with these enzymes to eliminate any background carryover, and the digest was transformed into competent E. coli cells by the method of Hanahan (21). Correct codon replacement

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: Somatogen, Inc., 2545 Central Ave., Boulder, CO 80301.
clones were identified by resistance to SpeI and Nhel restriction digestion and confirmed by deoxy DNA sequencing using the Sequenase (U. S. Biochemical Corp.) kit and protocol. The strain resulting from the replacement of the two rare codons was SGE2770, containing plasmid pSGE761, identical to pSGE720 except for the A to G change in the two β- globin codons.

Protein Sequencing—The sequence of proteins and peptides was determined by automated Edman degradation chemistry using a PE Protein Sequence 180 analyzer. Peptide supports were used as recommended. Phenylthiohydantoin-derivatives were identified by reversed phase chromatography on a modified Hewlett-Packard 1090L HPLC using a Hewlett-Packard AminoQ column.

Amino Acid Analysis—Protein and peptides were subjected to gas phase hydrolysis at 165 °C for 1 h in the presence of HCl containing 1% phenol using a Savant AP100 AminoPrep Hydrolyzer. Amino acids were analyzed using pre-column derivatization with 6-aminoquinolyl-N-hydroxysuccinimidy carbamate (AQc). The standard AccQ-Tag method was used for derivatization of amino acids (22). AQc derivatives of amino acids were separated on a HP 1090IM using a NovaPac C18 (2 × 150 mm) column. Solvent A was diluted 1/100 into Eluant A from Waters (1000 ml of water, 85 ml of concentrated Eluant A). Solvent B was 60% acetonitrile in HPLC grade water. Separation was performed using the following gradient conditions: 0 min, 0% B; 0.5 min, 2% B; 13.0 min, 7.0% B; 17.0 min, 10.0% B; 23.0 min, 21.0% B; 27.0 min, 22.0% B; 32.0 min, 42.0% B; 35.0 min, 64.0% B; 46.0 min, 100% B. Column temperature was 35 °C and flow rate was 0.250 ml/min. The eluate was monitored at 254 nm and additionally by fluorescence detection with an excitation wavelength of 254 nm and an emission wavelength of 395 nm. A typical load was 2.5 μl of hydrolyzed protein in 2.5 μl.

A modified OPA pre-column derivatization was used for quantitation of norvaline (22–24). Hydrolyzed protein and peptides were derivatized in-line with OPA and injected on an AminoQ RP column at 35 °C. To improve resolution of Nva from surrounding amino acids, the elution gradient was modified as follows: 0 min, 5% B, 0.45 ml/min flow; 0 min, 5% B, 0.45 ml/min flow; 10 min, 15% B, 0.45 ml/min flow; 10 min, 30% B, 0.45 ml/min flow; 12 min, 35% B, 0.45 ml/min flow; 20.0 min, 36% B, 0.45 ml/min flow; 30.0 min, 70% B, 0.45 ml/min flow; 32.0 min, 100% B, 0.90 ml/min flow; and 35 min, 100% B, 0.60 ml/min flow.

Analysis of Free Amino Acids in Lysate—1-ml samples of fermentation broth were collected and stored at 80 °C until analyzed. Prior to analysis, samples were thawed and lysed by sonication for 1 min on ice. Lysate was filtered through 3K Centricon membranes and analyzed for content of free amino acids using the modified OPA method described above.

Pyruvate Assay—Fresh cell broth was spun down and the supernatant was immediately frozen at −20 °C. Thawed supernatant was analyzed for pyruvate by HPLC using an Aminex HPX-87H column (Bio-Rad) employing isocratic elution with 0.004 N H₂SO₄ as recommended by the manufacturer.

Analytical C3 HPLC Separation—Recombinant hemoglobin was diluted to 2 mg/ml in HPLC-grade water. Where specified, globin was precipitated with 20 volumes of cold 0.6% HCl in acetone and solubilized in 2% (v/v) trifluoroacetic acid to a final concentration of 2 mg/ml. 100 μg of protein was injected onto a HP 1090IM using a Zorbax C83 HPLC analytical column (0.46 × 25 cm) at 40 °C. Solvent A was 0.1% (v/v) trifluoroacetic acid in water, and solvent B was acetonitrile with 0.1% (v/v) trifluoroacetic acid. The flow rate was regulated at 1 ml/min. Gradient conditions were as follows: 35% solvent B for 5 min, then the percentage of B was increased to 49% over 45 min. Eluant was monitored at 215, 280, and 400 nm.

Electrospray Mass Spectrometry—Electrospray mass spectra were obtained using a Vectra electrospray source and a model 201 single quadrupole mass spectrometer with a Finnigan MAT electrospray ionization source. The quadrupole was calibrated on the day of analysis with an MRFA tetrapeptide/apomyoglobin. The quadrupole mass spectrometer with a 2000 atomic mass unit range. The peptides were eluted at a flow rate of 1 ml/min. Solvent A was 0.1% (v/v) trifluoroacetic acid in water, and solvent B was 76% (v/v) acetonitrile with 0.1% (v/v) trifluoroacetic acid. The peptides were eluted isocratically with 5% solvent B for 5 min, and then solvent B was increased to 70% over 65 min. The absorbance was monitored at 215 and 400 nm.

Tryptic mapping was performed using Porozyme trypsin cartridges (25–26). Separation was performed using a Zorbax C18 column. The peptides were eluted starting with 1% solvent B at 25 min, and then solvent B was increased to 50% over 65 min. Solvent A was 0.1% (v/v) trifluoroacetic acid in water and solvent B acetonitrile with 0.1% (v/v) trifluoroacetic acid. The absorbance was monitored at 215 and 400 nm.

Oxygen Affinity Measurements—Oxygen equilibrium was measured using a Hemox analyzer, and values of pO₂ and n₅₀ were calculated as described previously (17–19).

RESULTS

C3 RP HPLC Chromatography and Initial Characterization

Peptide Mapping—Staphylococcus aureus V8 protease was used to map recombinant hemoglobin. V8 mapping of the entire hemoglobin molecule was performed without separating the globin chains. The protein was precipitated with acetone containing 0.6% (v/v) HCl and reduced and alkylated with vinylpyridine in specified experiments. After alkylation, the chains were desalted using a PD-10 disposable column (Sephadex G-25M, 2% (v/v) formic acid mobile phase. Dried globins (alkylated or nonalkylated) were reconstituted in 100 mM ammonium acetate, pH 4, containing 2 mM urea. The conditions for this digest were 1:30 enzyme/substrate (w/w) with an overnight incubation at room temperature. The following day, additional V8 protease was added to bring the final enzyme/substrate ratio to 1:25 (w/w), and the incubation was continued for an additional 2 h at room temperature. The products of this digestion were separated on a 4.6 × 250 mm Vydac C4 HPLC column at a flow rate of 1 ml/min. Solvent A was 0.1% (v/v) trifluoroacetic acid in water, and solvent B was 76% (v/v) acetonitrile with 0.1% (v/v) trifluoroacetic acid. The peptides were eluted isocratically with 5% solvent B for 5 min, and then solvent B was increased to 70% over 65 min. The absorbance was monitored at 215 and 400 nm.

Tryptic mapping was performed using Porozyme trypsin cartridges (25–26). Separation was performed using a Zorbax C18 column. The peptides were eluted starting with 1% solvent B at 25 min, and then solvent B was increased to 50% over 65 min. Solvent A was 0.1% (v/v) trifluoroacetic acid in water and solvent B acetonitrile with 0.1% (v/v) trifluoroacetic acid. The absorbance was monitored at 215 and 400 nm.

Oxygen Affinity Measurements—Oxygen equilibrium was measured using a Hemox analyzer, and values of pO₂ and n₅₀ were calculated as described previously (17–19).

RESULTS

C3 RP HPLC Chromatography and Initial Characterization

Fig. 1 shows C3 chromatograms of two samples of rHb1.1. In the case of sample B, we observed distinct shouldering in front of the β- and di-α-globin peaks. These shoulders were not visible in the globin peaks of control sample A. AQc amino acid analysis of the isolated main β- and di-α-globins.
and isolated shouldering fractions (see Fig. 1 for fraction collection) were in general agreement with the expected composition (Table I). Amino acid analysis of the di-a-globin showed low recovery of the methionine (3.4 versus expected 5 residues). This is due to approximately 40% methylation of the amino-terminal methionine in the di-a-chain (27). In the case of both shouldering fractions, recovery of methionine was discernibly increased relative to the respective main fraction, and in both shouldering fractions the methionine peak was atypically broad, possibly suggesting the presence of an unusual or modified amino acid in these fractions eluting closely to the methionine peak in our amino acid analysis. Later it was found that the AQC derivative of norvaline elutes very close to methionine causing overestimation of methionine content (see Fig. 4).

LC-MS analysis of hemoglobin sample B showed a mass loss of 14.86 ± 0.6 Da for the shouldering fraction on the beta-globin versus the main fraction and a mass loss of 10.66 ± 0.6 Da for the shouldering fraction on the di-a-globin versus the main fraction (Table II). It should be noted that the observed mass of the di-a-globin chain (30,329.3 ± 3.5) was significantly higher than expected (30,329.3). We ascribe this discrepancy to methylation of 40% the di-a-globin amino-terminal methionine (27), which translates to an expected 5.6 Da mass gain over the predicted mass for the di-a-globin.

V8 Mapping of Globins

A control sample A of rHb1.1 and sample B containing a new modification were mapped with V8 proteinase (Fig. 2A). Prior to the digestion both samples were alkylated with 4-vinylpyridine. V8 does not cleave at all available Glu sites and therefore generates a rather simple and consistent peptide set for rHb1.1 (Fig. 2B). V8 mapping of sample B showed small differences when compared with control sample A (Fig. 2A). A small difference peak was seen in front of the G peptide (carboxyl-terminal peptide from the beta-globin), and another small difference peak is visible in front of the M peptide (internal peptide from the di-a-globin). Mass analysis (Table III) of these difference peptides indicated that they each exhibited a mass loss of approximately 14 Da when compared with their respective neighboring peptides G and M. This mass loss was strikingly similar to the mass difference found between the leading shoulder fractions of the beta- and di-a-globins and their respective main peaks (Table II and Fig. 1). These difference peptides

| Amino acids | Residues/molecule | Di-a | Expected | Main peak | Front shoulder | Beta | Expected | Main peak | Front shoulder |
|-------------|------------------|------|----------|-----------|----------------|------|----------|-----------|----------------|
| Asp         | 24               | 23.08| 23.74    | 12        | 11.89          | 12.14|
| Ser         | 22               | 22.47| 22.59    | 5         | 4.86           | 4.97 |
| Glu         | 10               | 10.25| 11.60    | 11        | 10.94          | 11.24|
| Gly         | 15               | 15.99| 16.69    | 13        | 13.59          | 13.77|
| His         | 20               | 20.13| 19.64    | 9         | 9.42           | 8.95 |
| Arg         | 6                | 6.69 | 6.86     | 3         | 3.27           | 3.17 |
| Thr         | 18               | 17.12| 16.91    | 7         | 7.14           | 6.60 |
| Ala         | 42               | 40.28| 40.45    | 15        | 14.77          | 14.92|
| Pro         | 14               | 14.46| 14.35    | 7         | 7.21           | 7.09 |
| Tyr         | 6                | 6.86 | 6.11     | 3         | 3.60           | 3.38 |
| Val         | 25               | 25.11| 24.74    | 17        | 16.50          | 16.75|
| Met         | 5                | 5.31 | 4.81     | 2         | 1.81           | 2.83 |
| Lys         | 22               | 21.86| 22.22    | 12        | 11.95          | 11.73|
| Ile         | 0                | 0    | 0.45     | 0         | 0              | 0    |
| Leu         | 36               | 35.87| 33.88    | 18        | 17.43          | 16.88|
| Phe         | 14               | 14.42| 13.97    | 8         | 8.18           | 7.99 |

**TABLE II**

Mass spectrometry data on infused isolated fractions of the di-a- and beta-globins from sample B

The same fractions were used in amino acid analysis.

| Isolated protein fraction | Mass [Da] |
|---------------------------|-----------|
| Found                      | Expected  |
| Main beta-globin peak      | 15,912.4 ± 0.6 | 15,913.4 |
| Front shoulder on beta-globin peak | 15,898.1 ± 0.6 |
| Main di-a-globin peak      | 30,329.3 ± 3.5 | 30,323.9 |
| Front shoulder on di-a-globin peak | 30,319.5 ± 3.1 |

**Fig. 2. A, V8 map of sample A (control rHb1.1, top panel) and sample B (modified rHb1.1, bottom panel).** In the map of sample B two small difference peaks can be seen marked pre-G and pre-M. B, diagram of V8 cleavage sites on the beta and di-a chain from A.
were isolated and further characterized. V8 peptide mapping of the nonalkylated globin showed improved separation of the M and N peptides and additionally revealed a pre-N peptide eluting at the back side of the M peptide which was not evident in the map of the alkylated globins (chromatogram not shown). Mass analysis is included in Table IIIB.

**Characterization of the β-Globin Difference Peptide Eluting before G Peptide**

**Electrospray Mass Spectrometry**—Electrospray mass spectra of the pre-G difference peptide (which showed an asymmetric elution profile in the peptide map suggestive of more than one component) showed an envelope indicative of a single component with mass loss of 14 Da when compared with the G peptide (Table III).

**Edman Protein Sequencing**—Edman protein sequencing confirmed that the pre-G peptide has the same amino-terminal sequence as the G peptide. However, during further sequencing we observed partial substitution of norvaline for leucine at positions 110, 114, and 141 (cycles 9, 14, and 40). Fig. 3 shows an example chromatogram detecting both Leu and Nva at position 110 in cycle 9. The level of norvaline substitution in the pre-G peptide varies at each position with about 25% at position 110, more than 50% at position 141, but significantly lower amounts of substitution at positions 114, and 106, and virtually none at the leucine in position 105.

**Amino Acid Analysis**—The pre-G difference peptide was analyzed for amino acid composition. Amino acid analysis clearly indicated a new unusual amino acid eluting between S-pyridylethylcysteine and methionine (Fig. 4). The G peptide does not contain any methionine (for which it could be mistaken), further suggesting the presence of a new modified or unusual amino acid. By comparison with the retention times of a standard amino acid mixture spiked with norvaline, it was estab-
lished that the new peak in amino acid analysis corresponds to norvaline (Fig. 4). Elution of the norvaline next to methionine is consistent with the amino acid analysis of shoulders from the β- and di-α-globins.

Characterization of the Di-α-globin Difference Peptide Eluting in Front of M Peptide

Electrospray Mass Spectrometry—Electrospray mass spectra of pre-M the difference peptide showed an envelope indicative of a single component with mass loss of about 14 Da when compared with the M peptide (Table III).

Trypsin Digest and Edman Sequencing—The nominal M peptide was too large to establish the source of the observed difference in mass by sequencing. Therefore, mapping with trypsin was used to identify smaller sub-peptide(s) containing the difference. Trypsin digest of the pre-M peak revealed one difference peptide when compared with a trypsin map of the original M peptide (Fig. 5). The difference peptide was isolated and analyzed by electrospray mass spectrometry and Edman sequencing. Again this difference peptide revealed a 13.7-Da mass loss when compared with the disappearing reference sub-peptide from trypsin digestion of the M peptide. This new difference peptide was sequenced and was found to be a 28-mer (Leu100-Lys127 in di-α notation) containing 7 leucines: \(\text{LLSHCLLVTLAAHLPAEFTPAVHASLDK}\).

We did not observe significant substitution at the first 2 leucines; however, at the subsequent 5 leucines, we found evidence of partial norvaline substitution for leucine. The level of substitution ranged from 15 to 25%.

Amino Acid Analysis—The isolated and re-purified trypsin difference sub-peptide from pre-M was analyzed for amino acid composition. Amino acid analysis was in good agreement with the expected composition of the 28-mer, with a loss of 0.9

lecine residues and an indication of a new amino acid nearly coeluting with methionine. By comparison with the retention times of standard amino acids, it was established that the new peak in the amino acid analysis elutes at the same position as norvaline and is poorly resolved from methionine (data not shown).

Based on these results we conclude that the pre-M difference peptide (Leu100-Lys127) was from the modified M peptide from the first half of the di-α-globin in which leucines were substituted with different levels of norvaline yielding on average one substitution of norvaline for leucine per peptide. Similarly, the existence of a pre-N peak during mapping of the nonalkylated globins suggested that the same phenomena occurred on the second half of the di-α-globin. This may indicate that the corresponding repeating peptide (Leu242-Lys269) in the second part of the di-α-globin also contained partial substitution of leucine by norvaline, although we did not determine this experimentally. The pre-M (Arg31-Glu172) and pre-N (Arg173-Arg283) peptides in the sample B of rHb1.1 each constitute roughly 5% of the total of the respective M and N peptides.

Determination of Total Norvaline for Leucine Substitution

Our standard AQC amino acid analysis system did not serve well for the determination of the extent of norvaline for leucine substitution because of its poor resolution between norvaline and methionine. Therefore, we adopted another amino acid analysis method using OPA precolumn derivatization. In that system, norvaline is essentially base line resolved from the nearest amino acid, valine. Because the concentration of leucine is several orders of magnitude higher than norvaline, we used the UV signal for the quantitation of leucine and the fluorescent signal for the quantitation of norvaline. Fig. 6 (bottom panel) shows a standard amino acid mixture spiked with norvaline monitored with a fluorescence detector. Fig. 6 (top and middle panel) shows the level of norvaline versus leucine.
from analysis of two hydrolyzed samples of rHb1.1 hemoglobin. The results are included in Table IV.

**Distribution of Norvaline Substitution Measured by LC-MS**

To determine the extent of Nva incorporation and to establish which leucine residues within the amino acid sequence of β- and di-α-globin undergo norvaline misincorporation for leucine, we applied an orthogonal technique. Samples A and B were mapped with trypsin and analyzed by LC-MS. It was relatively straightforward to query complex LC-MS data for molecular weight information on certain tryptic peptides and on certain known (or possible) modifications to tryptic peptides by selective ion monitoring. For example, the expected mass for a substituted peptide and dividing it into the total area of all ions, i.e. 984 and 970. Percent relative abundance can be calculated by taking the integrated ion area for a substituted peptide and dividing it into the total area of all known forms of that peptide. If one compares the ionization ability of two peptides that have very similar primary amino acid sequence, it is quite conceivable that the relation of the two selected ion chromatographic intensities can give a reasonable estimate of percent relative abundances of these peptides.

Using this approach, it was possible to probe a number of sites within the amino acid sequence of the molecule and to determine the relative abundance of norvaline for leucine misincorporation. The leucine-containing tryptic peptides were chosen by the criteria that one or two leucine residues were present in the peptide. Analysis of peptides containing more than two leucine residues would become intractably complex. There are eight tryptic peptides containing a single leucine residue which should give a representative view of the pattern of norvaline for leucine misincorporation. Table V shows the percent leucine to norvaline misincorporation. These results were in agreement with Edman sequencing of pre-G and pre-M difference peptides.

The data for sample B indicated that norvaline can misincorporate for leucine residues throughout di-α- and β-globins, and there was variation between sites in the level of misincorporation. The control sample A only showed detectable leucine to norvaline substitution at position 141 (0.2%) in the β chain, which was one of the most highly substituted sites in the sample B.

**Codon Effects**

The two positions in the β-globin from sample B with one of the highest levels of norvaline misincorporation were Leu110 and Leu141. This was coincident with the only occurrence of CUA, the rarest of the six leucine encoding codons in *E. coli* (28). This suggested that the level of norvaline misincorporation could be tRNA-specific, possibly more likely at rarely used codons. To test this hypothesis we constructed a new version of the β-globin expression gene in which both of these CUA codons were changed to CUG. This is the most commonly used leucine codon in *E. coli* and constitutes 86% of the Leu codons used in the rHb1.1. This change resulted in no significant changes in the amount of soluble rHb1.1 produced, and virtually identical levels of norvaline misincorporation were measured (0.20 ± 0.11 versus 0.14 ± 0.00%). These results indicated that any potential codon-directed misincorporation did not play a significant role in overall levels of norvaline misincorporation. Probably very low levels at the other 68 leucine sites cumulatively comprised the major portion of the overall misincorporation. It is possible that low levels of norvaline misincorporation found in certain positions of purified hemoglobin may reflect structural requirements for folding of subunits, insertion of heme, and/or assembly of a functional trimeric recombinant hemoglobin. Therefore globins containing significant levels of misincorporation at those sites could fail to accumulate as soluble rHb1.1.

**Accumulation of Free Amino Acids in Whole Fermentation Broth**

Fig. 7 shows a set of typical time courses of accumulation of selected free amino acids found in lysed whole broth from a fermentation of *E. coli* engineered for high level expression of rHb1.1. In six fermentations the profiles were very similar; however, the amplitude and phase varied from fermentation to fermentation, resulting in different levels of misincorporation. The level of leucine decreased in the first 5 h and then remained relatively constant through the fermentation. In addition, we monitored the level of pyruvate since it has been indicated by others (1, 5, 29) as the primary substrate in biosynthesis of norvaline and norleucine. As can be seen in Fig. 7, stabilization of leucine level coincided with a rapid increase in pyruvate. Also, the accumulation of pyruvate was paralleled closely by norvaline. It appeared that norvaline and norleucine accumulation followed a bi-phasic profile. In the first 5 h following induction, norvaline and norleucine accumulated at a relatively slow rate, but after that norvaline accumulated significantly faster. Norleucine followed a similar trend, but the increased accumulation of norleucine occurred 1 to 2 h later than seen for norvaline.

**Kinetics of Norvaline Production and Misincorporation to rHb**

Fig. 7 shows a time course of rHb1.1 production and norvaline misincorporation. Hemoglobin expression began almost immediately following induction. Misincorporation, however, could not be detected before 6–8 h after induction. From that point on, the level of norvaline substituted for leucine continued to increase until the end of fermentation.
Hemoglobin from several fermentations was isolated at the end of fermentation and analyzed for norvaline misincorporation. At the same time, homogenized aliquots of whole fermentation broth were tested for free norvaline, norleucine, and leucine concentrations. When the level of norvaline substitution was plotted against the ratio of free norvaline to free leucine available at the end of fermentation, a high correlation was established with $r^2 = 0.90$ (Fig. 8). This indicated that the ratio of norvaline to leucine available in the fermentor had a very significant impact on misincorporation. Significant but weaker correlations of norvaline misincorporation with norvaline concentration or the inverse of leucine concentration were observed. This supports the argument that detectable levels of norvaline misincorporation can result from a significant increase in free norvaline or a decrease of leucine concentration, either of which can force this ratio to increase.

**DISCUSSION**

The results presented here show that *E. coli* can biosynthesize norvaline and incorporate it into recombinant hemoglobin at positions normally occupied by leucine. It appeared that the misincorporation in the purified sample of rHb1.1 was not equally distributed across the $\beta$ and the $\alpha$ chains, with certain leucine positions appearing more highly misincorporated than others. Either these positions were favored for substitution or the procedure employed for purification of soluble rHb1.1 may work as a selection mechanism against isoforms containing substitutions at certain other positions. At certain positions in hemoglobin norvaline misincorporation may disrupt folding and proper assembly. This hemoglobin may fail to accumulate as a soluble protein and/or may be destabilized such that it can be denatured during the heat deactivation step at the time of purification (20). Therefore the sites where misincorporation was detected in the purified functional recombi-
nant hemoglobin may not necessarily reflect the actual, presumably random distribution of misincorporation occurring during translation of the globins. The observed level of norvaline misincorporation did not affect oxygen affinity and cooperativity.

Our results indicate that the ratio of norvaline to leucine available during expression is the most critical factor in determining norvaline misincorporation (Fig. 8). This suggested that misincorporation of norvaline was the result of the mis-aminacylation of leucine transfer RNA with norvaline (2). Misincorporation produced by mis-aminacylation of tRNA would be expected to result in random distribution at leucine positions. However, our results indicate that the misincorporation in purified protein sample was not equally distributed across the β and the di-α chains. Two sites in the β-globin, Leu110 and Leu141, were correlated with the only occurrences of the rarest Leu codon “CUA” in the rHb1.1 genes, suggesting that the corresponding tRNA may have been more sensitive to mis-aminacylation. However, our experiments did not confirm a relationship between the rare codon usage for Leu employed in our expression construct and the level of norvaline misincorporation.

We believe that norvaline misincorporation into recombinant hemoglobin may not necessarily be unique; norvaline misincorporation is likely to be found in other recombinant proteins produced in E. coli. Similar misincorporation of norleucine for methionine is known to vary widely among different recombinant proteins expressed in E. coli (5). Therefore, it is conceivable that other proteins could demonstrate similar variation in norvaline misincorporation.

The leucine biosynthetic pathway in E. coli has been studied by a number of researchers. These groups have proposed norvaline and norleucine as potential by-products in this pathway (1, 29–31). They proposed that imperfect substrate fidelity could result in norvaline and norleucine production, originating from pyruvate through keto acid elongation and transamination processes. Our experiments have shown a reproducible, sequential order of such events (Fig. 7): accumulation of expressed hemoglobin, initial decrease in leucine concentration to a steady level with a simultaneous increase in pyruvate concentration, followed by accumulation of norvaline. Norvaline accumulation results in an increase of Nva/Leu ratio which is followed by norleucine misincorporation into recombinant hemoglobin. Later still a similar accumulation of norleucine is observed. These findings are consistent with Bogosian’s hypothesis that high level expression of leucine-rich recombinant proteins can stimulate production of norvaline and norleucine and cause misincorporation of these uncoded amino acids into proteins (1). However, it should be noted that we did not find norleucine for methionine misincorporation in the sample exhibiting norvaline misincorporation. Norleucine misincorporation should be influenced by the ratio of norleucine to methionine. Observed norleucine accumulation without significant methionine depletion was insufficient to produce detectable norleucine misincorporation. The difference in affinities of amino acids competing for the same tRNA will influence the ratio necessary to trigger detectable misincorporation of uncoded amino acids.

Based on our findings, we propose a model for norvaline misincorporation. In this model endogenous norvaline synthesis in E. coli alone is generally insufficient to produce detectable misincorporation into proteins. A crucial component for detectable misincorporation is the equilibrium between the supply and demand for leucine. In “normal” cells, leucine is constantly biosynthetically produced at the same time, norvaline (and norleucine) is synthesized to some extent as a by-product of the pathway. As the relative concentration of leucine is higher than norvaline, the ratio of Nva/Leu is low. A low Nva/Leu ratio helps the leucine tRNAs to properly charge and preclude significant norvaline misincorporation into proteins. Overexpression of a leucine-rich protein such as recombinant hemoglobin (12.5% leucine) can trigger norleucine and norvaline production in E. coli as postulated by Bogosian. Under this highly overleucine demand, the leucine biosynthesis pathway may begin to lose stringency. In such a “stressed” system, normal feedback signals are absent, and production and accumulation of norvaline as a by-product of the leucine pathway may increase. Our data also suggest that high levels of pyruvate may contribute to the increased production of norvaline relative to leucine. Since there is no clear mechanism to utilize norvaline, it will accumulate and the ratio of free Nva/Leu will rise. This allows leucine-specific tRNAs to increasingly mischarge with norvaline, resulting in the detectable incorporation of norvaline into proteins. Because the preponderance of norvaline accumulation happens late in the hemoglobin expression period (several hours after induction) when most of the cell proteins are already synthesized, most of the misincorporation occurs in hemoglobin produced late in the expression period. This would explain why we have been unable to find evidence of norvaline incorporation into native E. coli proteins from fermentations with readily detectable levels of norvaline in the expressed recombinant hemoglobin. From this proposed model, it would appear possible to carefully design a strategy to prevent norvaline misincorporation by appropriately supplementing the fermentation of E. coli with leucine. This strategy would be similar to the prevention of norleucine for methionine misincorporation by feeding E. coli with methionine (1). It should be noted, however, that addition of leucine may significantly influence physiology of the cells through leucine-responsive regulatory proteins (32). Alternatively, fermentation control strategies that limit the production of pyruvate appear to limit norvaline production by minimizing the concentration of the necessary substrate (pyruvate) for its biosynthesis. Limited production of norvaline results in minimizing norvaline misincorporation.

The requirement of very high purity and homogeneity for recombinant proteins that are to be used as human therapeutics presents a scientific and analytical challenge to demonstrate the purity and integrity of final products. Several very sensitive analytical methods can be used to detect low levels of contamination with other unknown proteins and/or misincorporation of encoded amino acids. For several decades norleucine and norvaline were used as convenient internal standards in amino acid analysis. The discovery of norleucine misincorporation precluded its use as a standard in analysis of recombinant proteins. Our findings raise the question as to whether there are other cases in which norvaline should no longer be used as an internal standard in the analysis of heterologous proteins.

Acknowledgments—We express our special thanks to Dr. Raymond E. Kaiser for many insightful suggestions as well as providing LC-MS data. We also thank our colleagues Debbie Abbott-Brown, Dr. Spencer J. Leach, B. Rognlie, I. Apostol, and J. Neway, manuscript in preparation.

### Table VI

| ID   | \( p_{\text{oc}} \) | \( \eta_{\text{max}} \) |
|------|----------------------|--------------------------|
| Sample A | 29.6 | 2.3 |
| Sample B | 29.4 | 2.4 |

\( p_{\text{oc}} \) and cooperativity of hemoglobins (\( \eta_{\text{max}} \))
Incorporation of Norvaline at Leucine Positions

Anthony-Cahill, Dr. Jeff Etter, Dr. Bruce Kerwin, Dr. Antony Mathews, and Dr. Justin Neway for their very helpful discussions during the discovery process and later during manuscript preparation. This work would not be possible without support of large group of Somatogen technical staff in manufacturing, pilot operations, and assay services, especially Daryl Odgen, Daniel Rice, and Tim Fattor.

REFERENCES
1. Bogosian, G., Violand, B. N., Dorward-King, E. J., Workman, W. E., Jung, P. E., and Kane, J. F. (1989) J. Biol. Chem. 264, 531–539
2. Barker, D. G., and Bruton, C. J. (1979) J. Mol. Biol. 133, 217–231
3. Semmes, O. J., Riehm, J. P., and Fao, K. R. (1985) Peptides (Elmsford) 6, 491–494
4. Lu, H. S., Tsai, L. B., Kenney, W. C., and Lai, P.-H. (1988) Biochem. Biophys. Res. Commun. 156, 733–739
5. Fiorino, A., Frigo, G., and Cucchetti, E. (1989) J. Chromatogr. 476, 83–92
6. Fleming, J., Terrell, L., Taylor, T., Miller, C., and Woodward, C. (1992) Hewlett-Packard Application Note 228–212, Hewlett-Packard Co.
7. Apostol, I., Aitken, J., Levine, J., Lippincott, J., Davidson, J. S., and Abbott-Brown, D. (1995) Protein Sci. 4, 2616–2618
8. Sharp, P. M., Crowe, E., Higgins, D. G., Shields, D. C., Wolfe, K. H., and Wright, P. (1988) Nucleic Acids Res. 16, 8207–8211
9. Suga, M., Inui, M., Goto, M., Kurusu, Y., and Yamada, D. (1991) Appl. Microbiol. Biotechnol. 35, 348–351
10. Kishino, M., Sugiyama, M., Takagi, T., and Chibata, I. (1977) J. Antibiot. 30, 111–117
11. Calvo, J. M., and Matthews, R. G. (1994) Microbiol. Rev. 58, 466–490