Exhibit 1

THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 261, No. 7, Issue of March 5, pp. 3116–3121, 1986
Printed in U.S.A.

Structural Characterization of Human Erythropoietin*

(Received for publication, August 26, 1985)

Por-Hsiung Lai, Richard Everett, Fung-Fang Wang‡, Tsutomu Arakawa, and Eugene Goldwasser‡

From Amgen, Thousand Oaks, California 91320 and the ‡Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois 60637

Erythropoietin is the primary regulator of red blood cell formation in mammals. Because of its extreme scarcity, very little information is available regarding structural features of this important glycoprotein. We report here the primary structure of human urinary erythropoietin, determined by protein sequencing. In addition, the sites of glycosylation, assignment of disulfide bonds, and the circular dichroism of the hormone analyzed for secondary structure in comparison with the prediction from the sequence are presented.

Although the existence of a humoral factor regulating red blood cell formation was postulated as early as 1906 (1), very little information about the structure of erythropoietin (EPO[1]) has been published because of its very limited availability. The hormone derived from plasma of anemic sheep was purified in 1971 (2), but too little was obtained to learn more than its apparent molecular weight (3), amino acid composition, and the fact that it was glycosylated (4). Human EPO purified from the urine of patients with aplastic anemia (5) consists of two distinguishable forms with differing content of carbohydrate; one termed α containing 31% and one termed β with 24% carbohydrate. All of the carbohydrate appears to be *N*-linked (6). The apparent M_r was estimated to be 34,000 and some information about its domain structure has been published (7).

MATERIALS AND METHODS AND RESULTS[2]

DISCUSSION

Primary Structure—The complete amino acid sequence for the human EPO protein (565 μg of EPO used) is shown in Fig. 1. The sequenced region of the intact protein and the various fragments used to establish the order of sequences obtained by cleavage of the protein are also indicated. As shown in Fig. 1, 77 residues of sequence information could be obtained with only 100 μg of protein by using the technique of *in situ* CNBr cleavage. This sequencing technique should have application in the structural analysis of proteins which are only available in minute quantities. Proteins containing more than 1 methionine residue can also be analyzed by this technique if primary amine-specific reagents such as fluorescamine or *o*-phthalaldehyde are used to strategically block unwanted peptides in the CNBr peptide mixture at a point where a proline residue is at the exposed NH_2 terminus (8).

The amino acid sequence of human EPO, shown in Fig. 1, contains 166 residues and has a calculated M_r = 18,398 for the protein moiety. It contains three more basic amino acids than acidic ones. Charged residues constitute 27% of the total and are irregularly distributed, except that no charged residues occur in region 77–88 and both the NH_2- and C-terminal ends are relatively highly charged. It is interesting to note that although glycine and proline residues which are known to be strong breakers of α-helix and β-sheet structures, are randomly distributed through most of the molecule, no such residues occur in regions 4–27 and 130–150. A high degree of α-helix structures may be possible in these regions.

In a previous report Yanagawa *et al.* (9) published the sequence of the 30 NH_2-terminal residues of human urinary EPO purified by an immunoaffinity method. This sequence differs from the one in this paper at residues 5 (Leu instead of Lys) and 14 (Arg instead of Ile); residues 3 and 24 were not specified; we find Pro at position 3, and assigned Asn for position 24. In our studies, this region of the EPO molecule has been sequenced four times using intact protein as well as tryptic and *Staphylococcus aureus* V8 protease peptides. In addition, our data for these positions are confirmed by the DNA sequence of the human gene (10, 11).

Glycosylation Sites—The sequence we report has three possible glycosylation sites at positions 24, 38, and 83, according to the presence of Asn-X-Ser/Thr (12). The assignment of Asn at these positions was also supported by the evidence that amino acid composition analysis of peptide T30 and T38 indicated the presence of glucosamine the *N*-acetylated species of which is the linking sugar on the asparagine residue (data not shown). Sequence analysis of peptides T28 and 2S63 indicated a serine at position 120 and no identifiable PTH for position 126. However, amino acid composition analysis revealed the presence of 2 serine residues in this fragment. Analysis of the DNA sequence indicated that a serine is present at position 126 (10, 11). One possible explanation for these results is that position 126 is a glycosylated serine. In fact, our preliminary results indicated that galactosamine whose precursor, *N*-acetylgalactosamine, is the linking sugar at hydroxy amino acids was detected in the composition analysis of peptides T28 and 2S63 (data not shown).

* This work was supported in part by Grant HL 21676 from the National Heart, Lung and Blood Institute, National Institutes of Health. 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.

[1] The abbreviations used are: EPO, erythropoietin; HPLC, high performance liquid chromatography; TFA, trifluoroacetic acid; PTH, phenylthiohydantoins; PTC, phenylthiocarbamyl; TPCK, L-1-tosylamide-2-phenylethyl chloromethyl ketone.

[2] Portions of this paper (including "Materials and Methods," "Results," Figs. 2 and 3, Tables 1 and 2, Footnote 3, and additional references) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20814. Request Document No. 85M-2899, cite the authors, and include a check or money order for $6.40 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

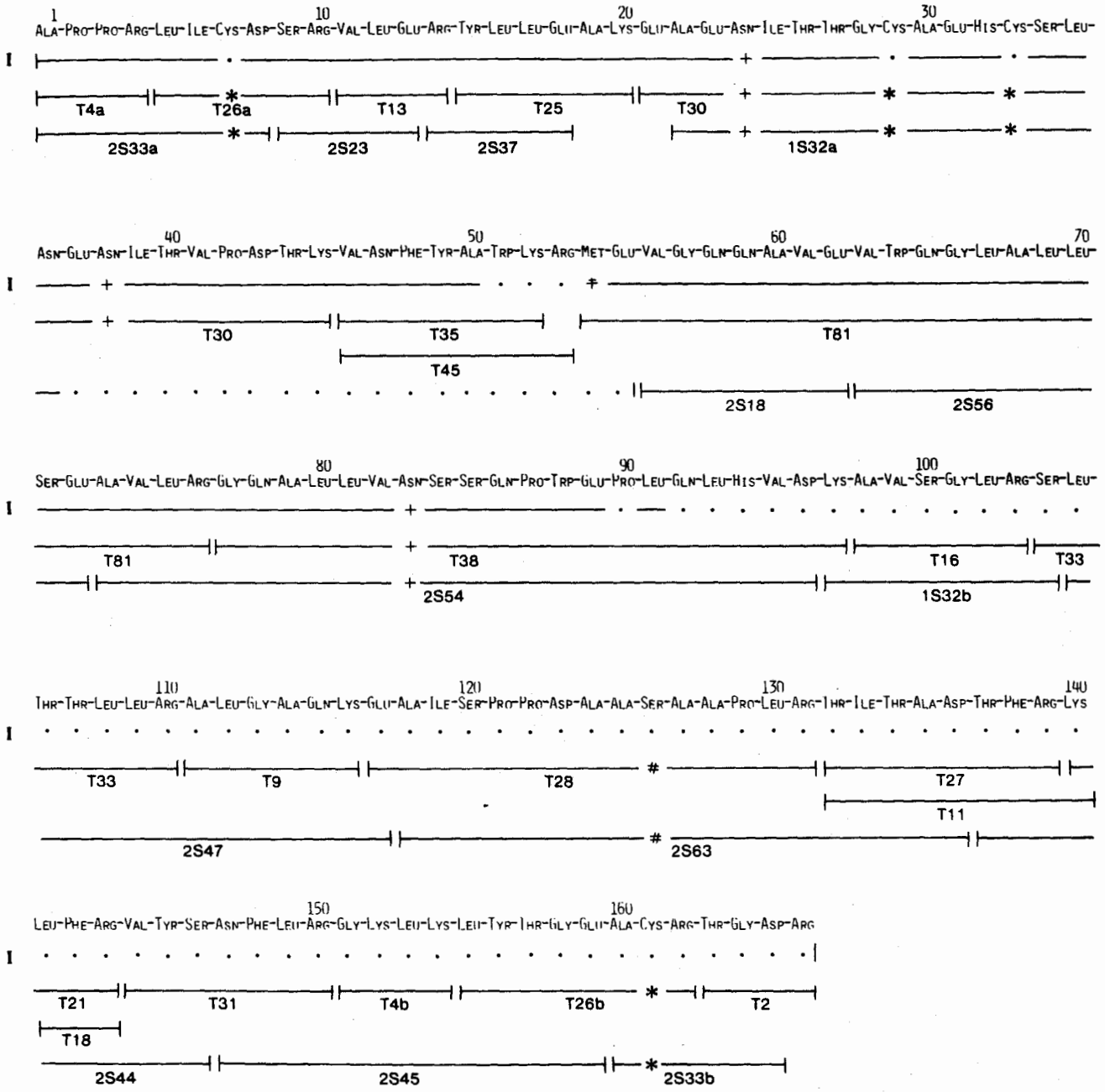

FIG. 1. **Amino acid sequence analysis of human erythropoietin.** Sequences analyzed with intact urinary EPO and peptides derived from urinary protein are indicated by *solid lines* under the residues comprising the protein or the peptide. The *letter code* indicates source of peptides and the cleavage method used to produce the peptides: *I*, intact protein; *T*, trypsin; *S*, *S. aureus* V8 protease; *1S*, digest C; *2S*, digest D. Results for sequence analysis with intact protein represents data obtained from two separate determinations. The number of the peptide identifies it in the respective HPLC chromatogram. The *solid line* indicates the results of automated Edman degradation. *Dots on the line* indicate respective residues which are not identified by automated Edman degradation. The letter *a* or *b* behind a peptide number indicates the pair of peptides which were co-isolated from peptide maps. * indicates identification of cysteine residues: Cys 7 and Cys 161 form one cystine and Cys 29 and Cys 33 form another cystine. The + indicates glycosylated asparagine. The # indicates a tentatively assigned glycosylated serine. The ‡ indicates identification of the methionine residue on the basis of cleavage chemistry and sequence analysis of T81.

Determination of Disulfide Bonds—As shown in Fig. 1, EPO contains 4 cysteine residues. Although during sequence analysis no PTH-cysteine could be detected for these 4 residues, we have assigned cysteine residue for positions 7, 29, 33, and 161 based on the following observations.

1. Two pairs of peptides, T26a/T26b and 2S33a/2S33b, copurified with equal recovery from high performance liquid chromatography of digests A and D, respectively. 2. PTH-cystine (eluted between PTH-threonine and PTH-glutamine under the described analytical conditions) could be detected at the seventh step of Edman degradation of both peptide pairs T26a/T26b and 2S33a/2S33b (Table 2), when degradation products were promptly analyzed. 3. PTH-cystine was detected at the thirteenth step of Edman degradation of

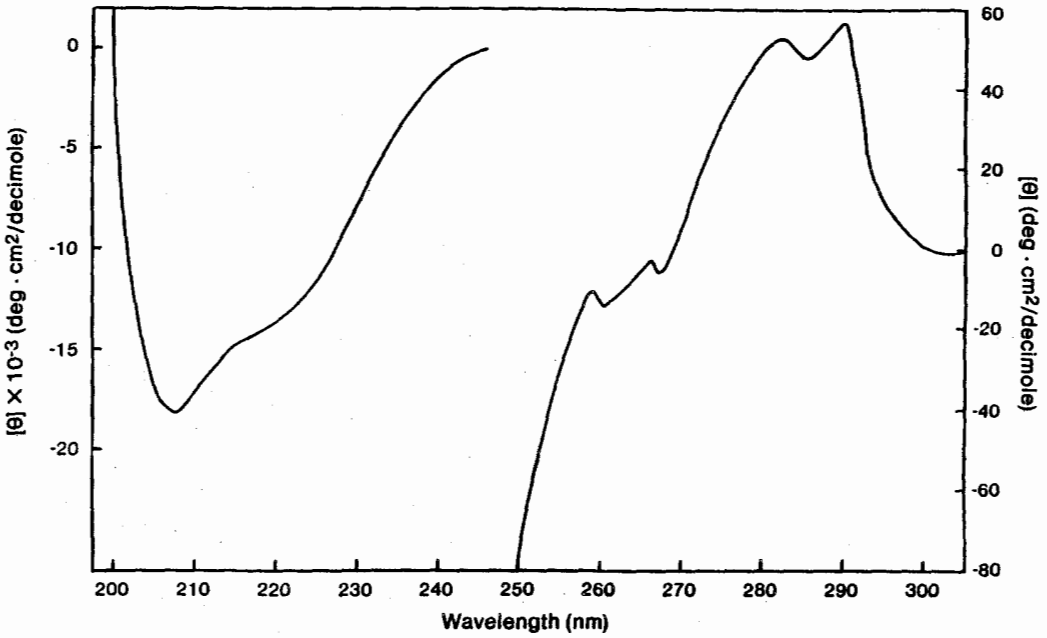

FIG. 4. The near- and far-UV spectra of human erythropoietin.

TABLE III
Analysis of secondary structure of human EPO

Method	Helix	Antiparallel and parallel β-sheet	Turns and others
	%	%	%
Chou and Fasman (14)	36	28	36
Garnier *et al.* (15)	42	21	37
CD (this work)	50	0	50

peptide T30 (Table 2) and at the twelfth step for peptide 1S32a (not shown in Table 2).

Based on these same observations, we also concluded that human EPO contains two disulfide bonds, one formed between Cys 7 and Cys 161, the other between Cys 29 and Cys 33. This conclusion also supports the previous report (6) that no free thiol is present in the EPO molecule. It is interesting to note that the second disulfide bond is sandwiched between two nearby glycosylation sites, i.e. Asn 24 and Asn 38.

Secondary Structure—The near- and far-UV CD spectra are shown in Fig. 4. The far-UV spectrum shows a minimum at 207.5 nm and a shoulder around 218 nm. The secondary structure of the protein was examined according to the method of Greenfield and Fasman (13). The α-helix content was calculated to be about 50% from the observed mean residue ellipticity at 208 nm. It seems that the remaining structure is mainly random and no obvious β-sheet structure could be observed.

Analysis of the sequence by a computer program based on the method of Chou and Fasman (14) suggests an α-helix content of about 36% and a β-sheet content of about 28%. Similar analysis by the method of Garnier *et al.* (15) predicts the α-helix content to be 42% with a β-sheet content of about 21%. Analysis of secondary structure of human EPO by prediction and CD measurements is summarized in Table III. The agreement with respect to α-helix is satisfactory but we do not yet know exactly whether there is any significant β-structure. However, it is noteworthy that the absence of obvious β-sheet structure may be expected from the distribution of proline, aspartic acid, and glutamic acid residues in the EPO molecule as shown in Fig. 1. These residues are highly unfavorable for β-sheet structure (16).

As reported in this study, EPO contains two disulfide bonds; however, the CD analysis of EPO showed no apparent CD signals between 300 and 350 nm where disulfide CD usually can be observed as a broad band (17). It may be possible that the EPO disulfide bonds have unfavorable configurations and give no CD extrema in this wavelength range or there is microenvironmental perturbation caused by interfering groups such as carbohydrate moieties.

The near-UV CD spectrum in Fig. 4 shows two strong positive bands at 282 and 290 nm and two weak negative bands at 260.5 and 267 nm. The observed positive CD bands can be assigned to the ^1L_b transition of tryptophan (18). The negative CD bands may be assigned to the transition of phenylalanine. Because of overlapping with the strong tryptophan transitions, the tyrosine CD bands, the maximum of which is usually located between 275 and 282 nm, are not apparent. These CD results clearly indicate that the protein has a distinct tertiary structure, providing asymmetric environments for the aromatic residues (17).

In an effort to understand possible structural relations between EPO and other known protein and nucleic acid sequences, we used a computer homology search which covers the Genbank and Dayhoff data bases. This analysis revealed no easily discernible homology with any proteins. Comparison with the recently published (19) sequence of another hemopoietic growth stimulator, colony-stimulating factor, also shows no significant homology.

Acknowledgments—We acknowledge Nowell Stebbing and Dan Vapnek for critical review of the manuscript and Joan Bennett for preparation of the manuscript.

REFERENCES

1. Carnot, P., and Deflandre (1906) *C. R. Acad. Sci.* **1434,** 384–387
2. Goldwasser, E., and Kung, C. K.-H. (1971) *Proc. Natl. Acad. Sci. U. S. A.* **68,** 697–698

3. Goldwasser, E., and Kung, C. K.-H. (1972) *J. Biol. Chem.* **247,** 5159–5160
4. Goldwasser, E. (1981) in *Control of Cellular Division and Development* (Cunningham, D., Goldwasser, E., Watson, J., and Fox, F., eds) pp. 487–494, A. R. Liss, New York
5. Miyake, T., Kung, C. K.-H., and Goldwasser, E. (1977) *J. Biol. Chem.* **252,** 5558–5564
6. Dordal, M. S., Wang, F. F., and Goldwasser, E. (1985) *Endocrinology* **116,** 2293–2299
7. Wang, F. F., Kung, C. K.-H., and Goldwasser, E. (1985) *Endocrinology* **116,** 2286–2292
8. Brauer, A. W., Oman, C. L., and Margolies, M. N. (1984) *Anal. Biochem.* **137,** 134–142
9. Yanagawa, S., Hirade, K., Ohnota, H., Sasaki, R., Chiba, H., Ueda, M., and Goto, M. (1984) *J. Biol. Chem.* **259,** 2707–2710
10. Jacobs, K., Shoemaker, C., Rudersdorf, R., Neill, S. D., Kaufman, R. J., Mufson, A., Seehra, J., Jones, S. S., Hewick, R., Fritsch, E. F., Kawakita, M., Shimizu, T., and Miyake, T. (1985) *Nature* **313,** 806–810
11. Lin, F. K., Suggs, S., Lin, C.-H., Browne, J., Smalling, R., Egrie, J., Chen, K., Fox, M., Stabinsky, Z., Badrawi, S., Lai, P.-H., and Goldwasser, E. (1985) *Proc. Natl. Acad. Sci. U. S. A.,* **82,** 7580–7584
12. Neuberger, A., Gottschalk, A., Marshall, R. D., and Spiro, R. G. (1972) in *The Glycoproteins: Their Composition, Structure and Function* (Gottschalk, A., ed) pp. 450–490, Elsevier/North Holland, Amsterdam
13. Greenfield, N., and Fasman, G. D. (1969) *Biochemistry* **8,** 4108–4116
14. Chou, P. Y., and Fasman, G. D. (1977) *J. Mol. Biol.* **115,** 135–175
15. Garnier, J., Osguthorpe, D. J., and Robson, B. (1978) *J. Mol. Biol.* **120,** 97–120
16. Chou, P. Y., and Fasman, G. D. (1978) *Annu. Rev. Biochem.* **47,** 251–276
17. Timasheff, S. N. (1970) *Enzymes* **2,** 371–443
18. Strickland, E. H. (1974) *CRC Crit. Rev. Biochem.* **2,** 113–175
19. Wong, G. G., Witek, J. S., Temple, P. A., Wilkens, K. M., Leary, A. C., Luxenberg, D. P., Jones, S. S., Brown, E. L., Kay, R. M., Orr, E. C., Shoemaker, C., Golde, D. W., Kaufman, R. J., Hewick, R. M., Wang, E. A., and Clark, S. C. (1985) *Science* **228,** 810–815

Additional references are found on p. 3120.

Supplemental Material to
Structural Characterization of Human Erythropoietin
by
Por-Hsiung Lai, Richard Everett, Fung-Fang Wang, Tsutomu Arakawa and Eugene Goldwasser

MATERIALS AND METHODS

Human urinary erythropoietin was purified as previously described (1,2).

Protease digestion and peptide separation. α-EPO, 120 μg, was lyophilized in a reacti-vial (Pierce) and dissolved in 25 μl of 10 mM calcium chloride, 0.1 M Tris-Cl, pH 8.0. TPCK-treated trypsin (Worthington) (2.4 μg) was added and the digestion was carried out at 37°C for 25 minutes. The reaction was stopped by adding phenylmethylsulfonylfluoride (Sigma) to a final concentration of 0.2 mM. This digest is designated digest A. Another tryptic digest was prepared using 80 μg of α-EPO (digest B). This digestion was performed at 37°C for 6 hours in 100 μl of 0.1 M ammonium bicarbonate, pH 8.0, using 2 μg of TPCK-treated trypsin (Worthington). Digestion of 185 μg of α-EPO (digest C) with 6.5 μg of S. aureus V8 protease (Miles) was performed in 0.1 M ammonium bicarbonate, pH 8.0 at 37°C for 42 hr. A second batch of S. aureus V8 protease digest using 50 μg of α-EPO (digest D) was prepared similarly.

All protease digests were separated by reverse-phase HPLC immediately after digestion. Peptides were eluted by a gradient formed by an aqueous mobile phase and an organic mobile phase. The aqueous mobile phase was either 0.06% TFA in water (solvent A) or 0.1% TFA in water (solvent B). The organic mobile phase was one of the following solvents: 0.05% TFA in 70% acetonitrile (solvent C), 0.1% TFA in 90% acetonitrile (solvent D), or 0.1% TFA in 80% acetonitrile (solvent E). Fractions were manually collected, dried, and kept at -20°C.

Digest A was separated on a Varian 5000 liquid chromatograph system equipped with a Synchropak RP-P (0.41 x 5 cm) column (Syn Chrom). The peptides were eluted with a linear gradient of 100% solvent A: 0% solvent C to 40% solvent A: 60% solvent C over a period of 60 minutes. The flow rate was 0.5 ml/min. The column was monitored at 220 nm by a Jasco Unidec-100-III UV Spectrophotometer.

Digests B and D were separated on a Vydac 5 μm C4 HPLC column (0.46 x 25 cm) using a Waters gradient HPLC system. The peptides were eluted with a linear gradient of 97% solvent B: 3% solvent D to 45% solvent B: 55% solvent D over a period of 95 minutes. The flow rate was 0.8 ml/min.

Digest C was separated on a Rainin Microsorb Short-Ones 3μm, C18 HPLC column (0.46 x 10 cm) using a Waters gradient HPLC system. The peptides were eluted with a linear gradient of 100% solvent B to 35% solvent B: 65% solvent E over 60 min then to 100% solvent E over 20 min. The flow rate was 1 ml/min.

Amino Acid Sequence Determinations

Automated sequence analyses (3, 4) of intact protein and peptide fragments isolated by HPLC were performed with a gas-phase sequenator using either a standard protein program or a new program designated MHNVAC supplied by M. Hunkapiller of Applied Biosystems. All peptide solutions were made in 50% formic acid (Fluka) in water before being applied to the sequenator. The amount of peptide samples used for sequence analysis varied from 30- to 80% of the materials recovered in the peptide fractions obtained from HPLC of protein digests. In the later phase of this study, the polybrene treated glass-fiber disc described previously (3) was replaced with a TFA-activated glass-fiber disc for the analysis of intact protein. The procedure[3] for the activation of the glass-fiber disc is as follows: the glass fiber disc is immersed in TFA in a covered glass container and kept for one hour at 22°±2°C. This TFA is then decanted and the activated disc is first air dried and further dried under vacuum over KOH. The protein sample previously reduced with 2-mercaptoethanol at 35°C for 30 min is applied in 50% formic acid directly onto the activated filter without polybrene. The disc loaded with sample is further dried under argon before sequencing.

The PTH-amino acid obtained from each sequenator cycle was identified by reverse-phase HPLC (5).

In situ CNBr (Eastman Kodak) cleavage of the remainder of the protein molecule after initial extended sequence analysis was performed as follows: after extended sequence analysis, the sequenator was paused at the end of a cycle leaving the PITC coupled protein uncleaved and the delivery tubing was disconnected. The cartridge reaction cell with the sample disc was removed from the reaction chamber. 30 μl of 10% (w/v) CNBr solution in 70% formic acid was quickly applied to the sample disc which was kept in place in the top piece of the cartridge. After loading CNBr solution, the cartridge reaction cell containing the disc was sealed using Teflon tape, wrapped with aluminum foil and placed in the reaction chamber of sequenator for one hour at 44°C.

[3] S. Kent, unpublished procedure.

Cartridge was reassembled at the end of CNBr cleavage. Before resuming sequence analysis using the same program, the sample disc was dried with argon for 10 min, washed with S2 (ethylacetate) for 2 min and then dried with argon for another 10 min.

Determination of Protein Disulfide Structure

Assignments of protein disulfide bonds were based on results of sequence analysis of peptide fragments and native and reduced intact protein, and HPLC mapping of tryptic and S. aureus V8 protease digests.

Prediction of Secondary Structure from Sequence

The amino acid sequence for human EPO was taken from this study. The prediction methods used are those of Chou and Fasman (6) and Garnier et al. (7).

Peptide Compositional Analysis by PTC-Amino Acids Methods

Compositional analysis of peptide hydrolysates derived from T28 and 2S63 were performed according to the improved method (8) of a published procedure (9). In this method, PTC-galactosamine is eluted between PTC-serine and PTC-glycine.

CD Measurements and Analysis

Circular dichroic spectra were determined at room temperature on a Jasco J-500C spectropolarimeter. Spectral band width was set at 1 nm. Cuvettes used were 0.1 and 1 cm in light path length for 190 to 260 nm and, 240 to 340 nm, respectively. The solvent spectrum was manually subtracted from the protein spectrum. CD measurements were made with the purified EPO in 2 mM K_2HPO_4-KH_2PO_4 (pH 7.0) at a protein concentration of 0.3 mg/ml. The results were expressed as mean residue ellipticity, [θ], calculated from the mean residue weight of 111. This value was obtained as the molecular weight of polypeptide/number of amino acid residues.

RESULTS

Primary Structure

Two separate runs of automated sequence analysis of the intact protein were performed. In one run, 30 μg of native α-EPO was carried out through 50 cycles, and 42 residues were positively assigned. Since a previous study (10) indicated that α-EPO contains only two methionine residues, it was possible to obtain more sequence information by performing in situ CNBr cleavage on the remaining part of the EPO molecule after initial extended N-terminal sequencing. Thus, 100 μg of 2-mercaptoethanol reduced α-EPO were sequenced on a TFA activated filter through 47 cycles. Forty-five residues were positively assigned after prompt HPLC analyses of products derived from Edman degradation cycles. After CNBr cleavage of the residual material, automated sequence analysis was continued through another 44 cycles, and degradation products were again promptly analyzed by HPLC. Only one major PTH-amino acid residue could be detected in each cycle analyzed except one which was later assigned to a glycosylation site (Asn 83). A total of 35 out of 44 cycles were positively identified such that a single sequence could be assigned. This sequence represents the region covering residues 55–89. The results of sequencing intact protein and sequence analysis after in situ CNBr cleavage are shown in Table 1. Although amino acid composition analysis (10) indicated the presence of two methionine residues, the in situ CNBr cleavage of the protein lacking the N-terminal 47 residues did not yield three fragments. Only one fragment, residues 55-166 was sequenced as judged from the sequencing results (Table 1). Fragment Phe_{48}-Tyr-Ala-Trp-Lys-Arg-Met_{54} was not sequenced possibly due to loss of the peptide during washing or due to destruction, during CNBr cleavage, of the PTC moiety which was left on the residual fragment when the sequenator was paused for CNBr treatment. In the latter case, this peptide would not have a free α-amino group for coupling reaction.

Most of the tryptic peptides analyzed were obtained from digest A (Figure 2). From this digest, 17 tryptic fractions were isolated and identified. Two fractions, i.e., fractions 4 and 26, are mixtures of two peptides of equal recovery. Fraction 4 consists of two tetrapeptides, i.e., T4a and T4b; the sequence of the latter peptide was determined by subtracting the known residues from the previously sequenced N-terminal region. The sequences of the two peptides collected in fraction 26, i.e., T26a and T26b were similarly determined.

The primary structure of EPO was established by aligning sequences of tryptic peptides with those obtained from the intact protein and the S. aureus V8 protease peptides as shown in Figure 1. The peptide contains residues 54-76 (T-81), was not isolated by HPLC of digest A. Since it consists of many hydrophobic residues, it was probably lost on the reverse phase column. A separation procedure that involves weaker hydrophobic interaction between peptide and column matrix was designed for the specific isolation of this peptide. Digest B was chromatographed on a C_4 HPLC column and peptide T81 was found in the last fraction eluted from the C_4 column (chromatogram not shown). Another peptide shown in Figure 1 which was also isolated from digest B is T45. T45 which had an intact Lys-Arg bond at its carboxy terminal was not found in digest A. It is interesting to note that 142 residues out of 166 residues (86%) of the whole molecule were entirely sequenced with the peptides

isolated from digest A. The average repetitive yield of sequencing runs was about 94%. All small peptides could be completely sequenced if the glass fiber disc contained polybrene. Results of sequence analysis of all peptide fragments used in establishing the primary structure of the protein are summarized in Table 2.

Sequence data obtained from peptides isolated from HPLC of digest D (Figure 3) provided further necessary information for reconstruction of the complete sequence by overlapping the tryptic fragments. From digest D, 10 S. aureus V8 protease peptide fractions were isolated and identified. Fraction 33 in Figure 3 contains two peptides of equal recovery, i.e., 2S33a and 2S33b as shown in Figure 1. Peptide 2S47 overlaps tryptic peptides T33, T9, and T28. Peptide 2S63 established the overlap between fragments T28 and T27. Fragments T27, T21 and T31 are overlapped by peptide 2S45. The rest of the C-terminal tryptic peptides including T31, T4b, T26b and T2 were overlapped by two peptides, i.e., 2S45 and 2S33b obtained from digest D. Fragments T38, T16, and T33 were overlapped by the S. aureus V8 protease peptide, 1S32b which together with peptide 1S32a were isolated from digest C (chromatogram not shown).

All of the residues were assigned positions by sequencing and positive identification except the asparagines at positions 24, 38, and 83 and one serine at position 126 which was identified and assigned by composition analysis of peptide T28 (data not shown). Ninety-nine percent of the residues have been assigned after more than one determination. The only two residues which were assigned based on single determination are Arg63 and Arg166. Determination of the C-terminal residue was based on sequence analysis and alignment of peptide T2 and confirmed by DNA sequencing. We did not detect any peptide whose sequence is not shown in Figure 1. Attempts to confirm the C-terminal residue by carboxypeptidase digestion failed possibly due to a sterically hindered C-terminal end.

Two peptide bonds which do not involve any aspartic acid or glutamic acid residues but involve serine residues were unexpectedly hydrolyzed by S. aureus V8 protease as evidenced by isolation and identification of peptides 1S32b and 2S47 and 2S44 and 2S45 which are linked by Ser104 - Leu105 bond and Ser146 - Asn147 bond, respectively.

REFERENCES

1. Miyake, T., Kung, C.K.-H. and Goldwasser, E. (1977) J. Biol. Chem. 252, 5558-5564.
2. Dordal, M.S., Wang, F.F. and Goldwasser, E. (1985) Endocrinol., 116, 2293-2299.
3. Hewick, R.M., Hunkapiller, M.W., Hood, L.E., and Dreyer, W.J. (1981) J. Biol. Chem. 256, 7990-7997.
4. Lai, P.-H. (1984) Anal. Chim. Acta 163, 243-248.
5. Hunkapiller, M.W. and Hood, L.E. (1983) Science 219, 650-659.
6. Chou, P.Y. and Fasman, E.D. (1977) J. Mol. Biol. 115, 135-175.
7. Garnier, J., Osguthorpe, D.J., and Robson, B. (1978) J. Mol. Biol. 120, 97-120.
8. Lu, H.S., Klein, M.L., Everett, R.R. and Lai, P.-H. in Modern Methods in Protein Chemistry (L'Italien, J., ed., Plenum, New York) in press.
9. Heinrikson, R.L. and Meredith, S.C. (1984) Anal. Biochem. 136, 65-74.
10. Goldwasser, E. (1981) in Control of Cellular Division and Development (Cunningham, D., Goldwasser, E., Watson, J., and Fox, F., eds.) pp. 487-494. A.R. Liss, New York.

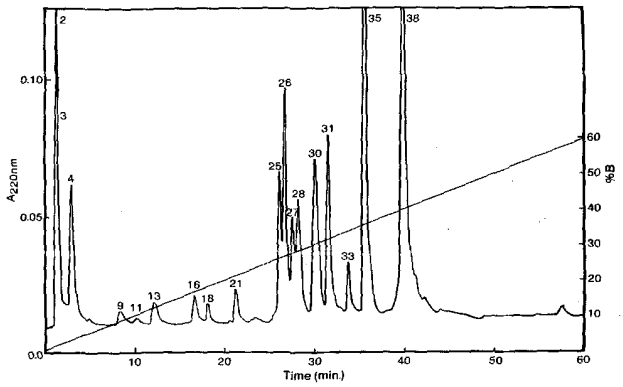

Figure 2. Tryptic map of Digest A.

Dash line indicates the gradient of solvent B. Number indicates source of the respective T-peptide.

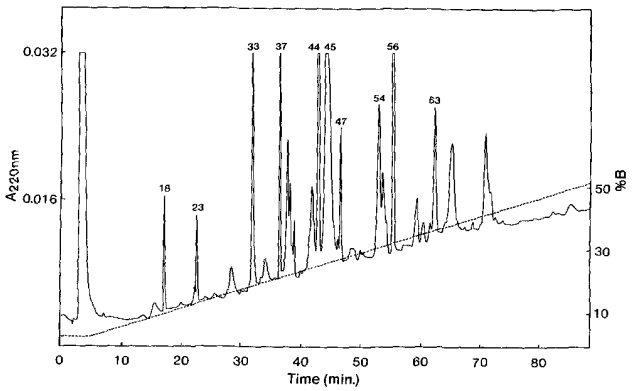

Figure 3. S. aureus V8 protease map of Digest D.

Dash line indicates the gradient of solvent B. Number indicates source of the respective S-peptide.

Table 1. Gas-phase sequence analysis of 100 µg of intact human EPO. No., residue number shown in Figure 1; I.D., PTH-amino acid identification by the one-letter code; yields are in pmols; NC, not calculated, identified qualitatively; (C), tentative assignments for cysteine based on absence of assignable PTH-amino acid; (N), tentative assignments for glycosylated Asn according to the rules of Asn-X-Ser/Thr;

(1) Sequenator was paused for in situ CNBr cleavage at methionines after sequencing through this residue.

(2) Known from cleavage chemistry and from T8l.

(3) Automated sequence analysis resumed after CNBr cleavage.

Table 1

No.	I.D.	Yield	No.	I.D.	Yield	No.	I.D.	Yield
1	A	520	31	E	93	61	V	259
2	P	378	32	H	NC	62	E	148
3	P	297	33	(C)	-	63	V	NC
4	R	NC	34	S	NC	64	W	NC
5	L	333	35	L	66	65	Q	144
6	I	297	36	N	35	66	G	131
7	(C)	-	37	E	81	67	L	239
8	D	96	38	(N)	-	68	A	249
9	S	24	39	I	26	69	L	242
10	R	NC	40	T	16	70	L	249
11	V	228	41	V	60	71	S	32
12	L	222	42	P	124	72	E	104
13	E	130	43	D	36	73	A	179
14	R	NC	44	T	NC	74	V	96
15	Y	120	45	K	52	75	L	202
16	L	230	46	V	81	76	R	NC
17	L	219	47	N(1)	23	77	G	78
18	E	125	48			78	Q	80
19	A	136	49			79	A	183
20	K	132	50			80	L	160
21	E	125	51			81	L	158
22	A	177	52			82	V	83
23	E	122	53			83	(N)	-
24	(N)	-	54	M(2)		84	S	24
25	I	101	55	E(3)	237	85	S	23
26	T	24	56	V	287	86	Q	48
27	T	58	57	G	217	87	P	NC
28	G	96	58	Q	238	88	W	NC
29	(C)	-	59	Q	220	89	E	44
30	A	102	60	A	284	90	X	-
						91	L	93

Table 2. Automated Edman degradation of peptides derived from tryptic and S. aureus V8 protease digests: yield of PTH-amino acids in pmols. No., residue number shown in Figure 1; I.D., PTH-amino acid identification by the one-letter code; NC, not calculated, identified qualitatively; -, not identified as PTH-amino acid.

(1) Identified as PTH-cystine, yield not calculated.

(2) Known as N-glycosylation site as described in text.

(3) Tentatively assigned as O-glycosylation site as described in text.

Table 2

No.	I.D.	2S33a	T26a	T13	T25	T30
1	A	173				
2	P	162				
3	P	165				
4	R	78				
5	L	150	220			
6	I	142	210			
7	C	NC(1)	-			
8	D	41	52			
9	S		49			
10	R		21			
11	V			450		
12	L			433		
13	E			156		
14	R			26		
15	Y				155	
16	L				149	
17	L				152	
18	E				96	
19	A				98	
20	K				45	
21	E					78
22	A					75
23	E					42
24	N					-(2)
25	I					36
26	T					32
27	T					39
28	G					46
29	C					-
30	A					49
31	E					22
32	H					NC
33	C					NC(1)
34	S					6
35	L					15
36	N					8
37	E					16
38	N					-(2)
39	I					12
40	T					7

No.	I.D.	T30	T35	T45	T81	2S56	T38
41	V	13					
42	P	8					
43	D	8					
44	T	4					
45	K	6					
46	V		210	58			
47	N		195	42			
48	F		139	45			
49	Y		105	22			
50	A		155	30			
51	W		58	NC			
52	K		54	12			
53	R			8			
54	M				55		
55	E				46		
56	V				43		
57	G				38		
58	Q				22		
59	Q				25		
60	A				19		
61	V				14		
62	E				8		
63	V				12	320	
64	W				NC	185	
65	Q				11	267	
66	G				20	251	
67	L				24	279	
68	A				21	243	
69	L				19	254	
70	L				21	260	
71	S				9	121	
72	E				5	44	
73	A				11		
74	V				9		
75	L				6		
76	R				NC		
77	G						235
78	Q						262
79	A						278
80	L						155

No.	I.D.	T38	1S32b	T16	T33	T9	2S47
81	L	181					
82	V	172					
83	N	-(2)					
84	S	60					
85	S	62					
86	Q	78					
87	P	35					
88	W	48					
89	E	38					
90	P	22					
91	L	39					
92	Q	27					
93	L	31					
94	H	9					
95	V	29					
96	D	12					
97	K	10	90				
98	A		121	210			
99	V		115	202			
100	S		20	35			
101	G		66	75			
102	L		75	178			
103	R		31	12			
104	S		12		40		
105	L				155		225
106	T				45		78
107	T				46		75
108	L				98		94
109	L				112		98
110	R				17		22

No.	I.D.	T9	2S47	T28	2S63	T27	2S44	T21
111	A	317	42					
112	L	305	28					
113	G	292	33					
114	A	279	39					
115	Q	267	35					
116	K	75	38					
117	E		10	92				
118	A			105	42			
119	I			101	50			
120	S			55	24			
121	P			88	48			
122	P			91	38			
123	D			49	35			
124	A			69	38			
125	A			77	35			
126	S			-(3)	-(3)			
127	A			77	33			
128	A			81	29			
129	P			NC	21			
130	L			15	18			
131	R			NC	15			
132	T				6	75		
133	I				18	155		
134	T				6	72		
135	A				15	136		
136	D				9	130		
137	T					48	45	
138	F					75	75	
139	R					22	40	
140	K						98	160

No.	I.D.	2S44	T21	T31	2S45	T46	T26b	2S33b	T2
141	L	86	155						
142	F	70	152						
143	R	32	22						
144	V	72		155					
145	Y	58		78					
146	S	12		45					
147	N			135	28				
148	F			128	45				
149	L			130	48				
150	R			42	14				
151	G				33	45			
152	K				42	48			
153	L				39	50			
154	K				38	12			
155	L				29		220		
156	Y				16		207		
157	T				5		45		
158	G				18		172		
159	E				5		68		
160	A						80	173	
161	C						NC(1)	-	
162	R						25	70	
163	T							55	48
164	G							122	65
165	D							33	44
166	R								21