Cassady, W. E., & Wagner, R. P. (1971) *J. Cell Biol. 49*, 1120–1121.
Chiu, N., Chiu, A., & Suyama, Y. (1975) *J. Mol. Biol. 99*, 37–50.
Dimitriadis, G. J., & Georgatsos, J. G. (1974) *FEBS Lett. 46*, 96–100.
Faye, G., Kujawa, C., & Fukuhara, H. (1974) *J. Mol. Biol. 88*, 185–203.
Fradin, A., Grühl, H., & Feldmann, H. (1975) *FEBS Lett. 50*, 185–189.
Grivell, L. A., & Metz, V. (1973) *Biochem. Biophys. Res. Commun. 55*, 125–131.
Heckman, J. E., Alzner-Dewerrd, B., & RajBhandary, U. L. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 717–721.
Küntzel, H., Ali, I., & Blossey, C. (1974) in *The Biogenesis of Mitochondria* (Kroon, A. M., & Saccone, C., Eds.) pp 71–78, Academic Press, London.
Macino, G., & Tzagoloff, A. (1979) *Mol. Gen. Genet. 169*, 183–188.
Martin, N. C., & Rabinowitz, M. (1978) *Biochemistry 17*, 1628–1634.
Martin, N. C., Rabinowitz, M., & Fukuhara, H. (1977) *Biochemistry 16*, 4672–4677.
Martin, R. P., Schneller, J. M., Stahl, A. J. C., & Dirheimer, G. (1976a) in *Genetics and Biogenesis of Chloroplasts and Mitochondria* (Bücher, T., et al., Eds.) pp 755–758, Elsevier, Amsterdam.
Martin, R. P., Schneller, J. M., Stahl, A. J. C., & Dirheimer, G. (1976b) *Biochem. Biophys. Res. Commun. 70*, 997–1002.
Martin, R. P., Schneller, J. M., Stahl, A. J. C., & Dirheimer, G. (1977) *Nucleic Acids Res. 4*, 3497–3510.
Martin, R. P., Sibler, A. P., Schneller, J. M., Keith, G., Stahl, A. J. C., & Dirheimer, G. (1978) *Nucleic Acids Res. 5*, 4579–4592.
Mitra, S. K. (1978) *FEBS Lett. 91*, 78–80.
Nirenberg, M., & Leder, P. (1964) *Science 145*, 1399–1407.
Oliver, S. G., McCready, S. J., Holm, C., Sutherland, P. A., McLaughlin, C. S., & Cox, B. S. (1977) *J. Bacteriol. 130*, 1303–1309.
Raba, M., Limberg, K., Burghagen, M., Gross, H. J., Katze, J. R., Simsek, M., Heckman, J. E., & RajBhandary, U. L. (1979) *Eur. J. Biochem.* (in press).
Schneller, J. M., Martin, R. P., Stahl, A. J. C., & Dirheimer, G. (1975a) *Biochem. Biophys. Res. Commun. 64*, 1046–1053.
Schneller, J. M., Accoceberry, B., & Stahl, A. J. C. (1975b) *FEBS Lett. 53*, 44–48.
Schneller, J. M., Stahl, A. J. C., & Fukuhara, H. (1975c) *Biochimie 57*, 1051–1057.
Schneller, J. M., Schneller, C., Martin, R. P., & Stahl, A. J. C. (1976) *Nucleic Acids Res. 3*, 1151–1165.
Smith, C. J., Teh, H. S., Ley, A., & D'Obrenan, P. (1973) *J. Biol. Chem. 248*, 4475–4485.
Suyama, Y., & Hamada, J. (1976) in *Genetics and Biogenesis of Chloroplasts and Mitochondria* (Bücher, T., et al., Eds.) pp 763–770, Elsevier, Amsterdam.
Swanson, R. F. (1971) *Nature (London) 231*, 31–34.
Waters, L. C., & Mullins, B. C. (1977) *Prog. Nucleic Acid Res. Mol. Biol. 20*, 131–160.
Weissenbach, J., Dirheimer, G., Falcoff, R., Sanceau, J., & Falcoff, E. (1977) *FEBS Lett. 82*, 71–76.
Wesolowski, M., & Fukuhara, H. (1979) *Mol. Gen. Genet. 170*, 261–275.

Amino Acid Sequence of the Ribosomal Protein L21 of *Escherichia coli*†

Ingeborg Heiland* and Brigitte Wittmann-Liebold

ABSTRACT: The primary structure of protein L21 from the 50S subunit of *Escherichia coli* ribosomes has been completely determined by sequencing the peptides obtained by digestion of L21 with trypsin before and after modification of the arginine residues with 1,2-cyclohexanedione, *Staphylococcus aureus* protease, thermolysin, and pepsin. Automated Edman degradation using a liquid-phase sequenator was carried out on the intact protein as well as on a fragment arising from cleavage with cyanogen bromide. Protein L21 consists of a single polypeptide chain of 103 amino acids of molecular weight 11 565. An estimation of the secondary structure of protein L21 and a comparison with other *E. coli* ribosomal protein sequences are presented.

Protein L21 is part of the large subunit of *Escherichia coli* ribosomes. The protein L21 does not bind directly to ribosomal RNA but is bound to the 23S RNA in the presence of L20 (Roth and Nierhaus, unpublished experiments). The protein is located in the neighborhood of protein L17 as established by cross-linking experiments (Kenny & Traut, 1979). It belongs to a group of six proteins present in a ribonucleoprotein core remaining after controlled trypsin and RNase digestion of the 50S subunit (Kühlbrandt & Garrett, 1978).

In reconstitution experiments it was found that the 50S particle without protein L21 has full activity in the poly(U)-dependent poly(Phe) synthesis (Spillman et al., 1977). It was also found that protein L21 is not essential for the peptidyltransferase activity (Schulze and Nierhaus, unpublished experiments).

Using a mutant with an altered protein L21 (Dabbs, 1978), it was recently possible to map the gene of this protein at 68 min of the *E. coli* chromosome (Kitakawa et al., 1979; Takata, 1978).

† From the Max-Planck-Institut für Molekulare Genetik, Abt. Wittmann, Berlin-Dahlem, West Germany. *Received June 22, 1979.*

0006-2960/79/0418-4605$01.00/0

Table I: Enzymatic and Chemical Cleavage of Protein L21[a]

expt	enzyme and source	amount of protein (mg)	amount of enzyme or reagent	buffer used[g]	time and temp (°C)
1	TPCK-trypsin[b]	3.0	60 μg	2.0 mL of buffer 1	4 h; 37
2	*Staphylococcus aureus V8* protease[c]	3.0	100 μg	2.0 mL of 0.05% ammonium acetate (pH 4.0)	48 h; 37
3	TPCK-trypsin after blocking of Arg residues with 1,2-cyclohexanedione	1.2	24 μg	1.5 mL of buffer 1	4 h; 37
4	thermolysin[d]	2.0	20 μg	1.5 mL of buffer 2	24 h; 50
5	pepsin[d]	1.2	24 μg	1.0 mL of 0.06 N HCl	4 h; 37
6	carboxypeptidase A[e]	0.1	20 units	170 μL of buffer 2	0–40 min; 37
7	carboxypeptidase B[e]	0.1	20 units	210 μL of buffer 2	0–90 min; 37
8	CNBr[b]	10.0	2 × 10 mg	1.5 mL of 70% formic acid	2 × 48 h; 25
9	BNPS-skatole[f]	0.15	0.6 mg	150 μL of 50% acetic acid	48 h; 25

[a] Cleavage with α-chymotrypsin (TLCK treated; Merck, Darmstadt, Germany) was performed in buffer 2 with an enzyme/substrate ratio of 1:50 for 1.5–4 h at 37 °C. [b] Merck, Darmstadt, Germany. [c] Miles, Slough, U.K. [d] Serva, Heidelberg, Germany. [e] Boehringer, Mannheim, Germany. [f] Pierce Chemical Co., Rockford, IL. [g] Buffer 1: H_2O adjusted to pH 7.8 with 0.05% NH_3. Buffer 2: 0.2 M morpholine formate, pH 8.0.

The present paper describes the determination of the complete primary structure of protein L21. Using the new sequencing method with a colored derivative of phenyl isothiocyanate (Chang et al., 1978), it was possible to scale down the amount of protein necessary for the amino acid sequence analysis. The increase of sensitivity over the conventional manual technique was 5–10-fold. As shown in Figure 3, all regions of the protein chain were sequenced at least twice.

Materials and Methods

If not stated otherwise, chemicals were of analytical grade purchased from Merck (Darmstadt, Germany). Solvents and reagents for the liquid-phase sequenator, for the dansyl–Edman degradation, and for the DABITC[1] method were purified as described previously (Wittmann-Liebold, 1973; Wittmann-Liebold et al., 1975; Chang et al., 1978).

Isolation of Protein L21. Protein L21 was isolated according to Hindennach et al. (1971) and kindly provided by Dr. H. G. Wittmann. Its identity and purity were checked by two-dimensional polyacrylamide gel electrophoresis (Kaltschmidt & Wittmann, 1970).

Amino Acid Composition. Samples of the protein were hydrolyzed in 5.7 N HCl–0.02% 2-mercaptoethanol in vacuo at 110 °C for 20, 48, and 72 h. Cysteine was determined as cysteic acid after oxidation of the protein with performic acid according to Hirs (1956). Tryptophan was determined after hydrolysis with 3 N mercaptoethanesulfonic acid (Pierce Chemical Co., Rockford, IL) at 110 °C for 20 h (Penke et al., 1974). The analyses were performed on a Durrum D-500 amino acid analyzer (Palo Alto, CA) at a sensitivity of 2.0 A. Peptides were hydrolyzed for 20 h and analyzed at a sensitivity of 0.5 or 0.2 A. In addition, tryptophan was determined from ultraviolet absorption spectra (Edelhoch, 1967) and from fluorescence spectra of the protein.

Tryptophan-containing peptides were located on peptide maps with the Ehrlich reagent (Smith, 1953) or after performic acid oxidation by their fluorescence under UV light at 366 nm.

Enzymatic and Chemical Cleavage. The cleavage conditions and sources of enzymes are summerized in Table I. Blocking of the arginyl residues with 1,2-cyclohexanedione prior to tryptic digestion was carried out according to Toi et al. (1967). Subsequent cleavage of peptides was performed under the same conditions as described above, and the details are given under Results.

Isolation of Peptides. The peptide mixtures derived from cleavages of protein L21 with the various enzymes were fractionated on Sephadex G-50 (superfine) columns as described under Results, except that the thermolysin peptides were separated on a Dowex 50 M71 microcolumn (Chen & Wittmann-Liebold, 1975). Peptides were purified from combined fractions by peptide mapping or by chromatography on cellulose thin-layer sheets (Macherey & Nagel, Düren, Germany) as described earlier (Heiland et al., 1976). They were stained by spraying either with ninhydrin (1.5 g of ninhydrin, 100 mL of acetic acid, and 900 mL of ethanol) or with 0.0005% fluorescamine (Pierce Chemical Co., Rockford, IL) in acetone after adjusting the pH of the plate by spraying with 5% pyridine in acetone (Vandekerckhove & Van Montagu, 1974). The peptides were eluted with 0.2 mL of 5.7 N HCl–0.02% 2-mercaptoethanol for amino acid analysis or with 50% acetic acid or 0.1 M NH_3 and 20% pyridine for sequencing.

Sequence Determination. Manual Methods. The tryptic peptides were sequenced by a micro-dansyl–Edman method described by Bruton & Hartley (1970). All other peptides were sequenced by a method using a colored derivative of phenyl isothiocyanate, 4-(*N*,*N*-dimethylamino)azobenzene 4′-isothiocyanate (DABITC) (Fluka, Ulm, Germany; recrystallized), described by Chang et al. (1978). Aspartic acid and asparagine as well as glutamic acid and glutamine were clearly identified by the DABITC method although isoleucine and leucine were not well resolved. They were determined by the dansyl–Edman degradation of the tryptic peptides and by amino acid analysis.

When approximately 1 nmol of peptide was sequenced by the DABITC method, smaller tubes (0.5 × 5 cm) sealed with parafilm and smaller amounts of reagents and solvents (25%) were used as described by Chang et al. (1978). The DABTH amino acids were dissolved in 2–3 μL of ethanol, and 50–100%

[1] Abbreviations and symbols used: TPCK, L-1-(tosylamido)-2-phenylethyl chloromethyl ketone; TLCK, N^{α}-*p*-tosyl-L-lysylchloromethane; BNPS-skatole, 2-(2-nitrophenylsulfenyl)-3-methyl-3-bromoindolenine; PTH, 3-phenyl-2-thiohydantoin; DABITC, 4-(*N*,*N*-dimethylamino)-azobenzene 4′-isothiocyanate; DABTH, 4-(*N*,*N*-dimethylamino)azobenzene 4′-thiohydantoin; DABTC, 4-(*N*,*N*-dimethylamino)-4′-(thiocarbamyl)-azobenzene; T, peptides resulting from tryptic digestion; CHD, peptides resulting from digestion with trypsin after blocking the arginine residues with 1,2-cyclohexanedione; SP, peptides resulting from digestion with *Staphylococcus aureus* protease; TH, peptides resulting from digestion with thermolysin; P, peptides resulting from digestion with pepsin; CH, peptides resulting from digestion with α-chymotrypsin; CNBr, peptides resulting from cleavage with cyanogen bromide; —, dansyl–Edman degradation; —, sequencing with the DABITC method; →, both manual methods; →, liquid-phase Edman degradation.

Table II: Amino Acid Composition of Protein L21[a]

	residues derived from sequence	residues derived from hydrolysis
Cys		[b]
Asx	4	4.04
Thr	5	4.75
Ser	3	3.11
Glx	15	15.09
Pro	1	1.18
Gly	12	12.19
Ala	7	7.35
Val	14	13.79[c]
Met	2	1.64
Ile	7	6.87[d]
Leu	3	3.36
Tyr	2	1.59
Phe	5	4.78
His	4	3.90
Lys	10	10.16
Arg	8	8.20
Trp	1	1[e]

[a] Amino acid composition as determined after hydrolysis at 110 °C for 20 h (averaged and uncorrected values from several runs). [b] Determined after oxidation as cysteic acid. [c] Value after hydrolysis for 72 h. [d] Value after hydrolysis for 48 h. [e] Determined by spectral methods and hydrolysis with mercaptoethanesulfonic acid.

of the solution was taken for their identification by thin-layer chromatography.

A second synthetic marker (DABTC–ethanolamine) was used, the position of which is below DABTC–diethylamine and just below DABTH–glutamic acid [for the identification on thin-layer sheets see Chang et al. (1978)]. DABITC amino acids were synthesized according to Chang et al. (1976).

Liquid-Phase Edman Degradation. Degradation of about 200 nmol of L21 and of 200 nmol of the larger fragment derived from cyanogen bromide cleavage was performed according to the method of Edman & Begg (1967) in an improved Beckman sequenator (Wittmann-Liebold, 1973) provided with an automatic conversion device (Wittmann-Liebold et al., 1976). A program with double coupling (in 0.25 M quadrol buffer, pH 9.0) and double cleavage per cycle was used. The cyanogen bromide fragment was dried in the cup and reacted with 200 μg of Braunitzer's reagent IV (Braunitzer et al., 1971). It was dissolved in 0.2 mL of water with the addition of 0.3 mL of *N*,*N*-dimethylamino-1-propyne for 40 min at 54 °C and degraded in 0.1 M quadrol buffer. In this case the lysine residues could not be determined because of their reaction with the Braunitzer's reagent. A total of 20–50% of the released PTH amino acids from the protein run and 30%–100% from the fragment run were identified by thin-layer chromatography using three successive solvent systems (Wittmann-Liebold et al., 1975). Aliquots of degradation cycles 7, 11, 13, and 14 of the fragment run were identified by high-pressure liquid chromatography (Strickland et al., 1978).

Results

Amino Acid Composition. The amino acid composition of L21 is presented in Table II. No cysteic acid was detected after oxidation. Tryptophan was determined on the amino acid analyzer after hydrolysis with mercaptoethanesulfonic acid as well as by spectral analysis of the intact protein as described under Materials and Methods.

Cleavage of the Protein. Isolation and Sequence of the Peptides. Peptides are numbered according to their order in the protein.

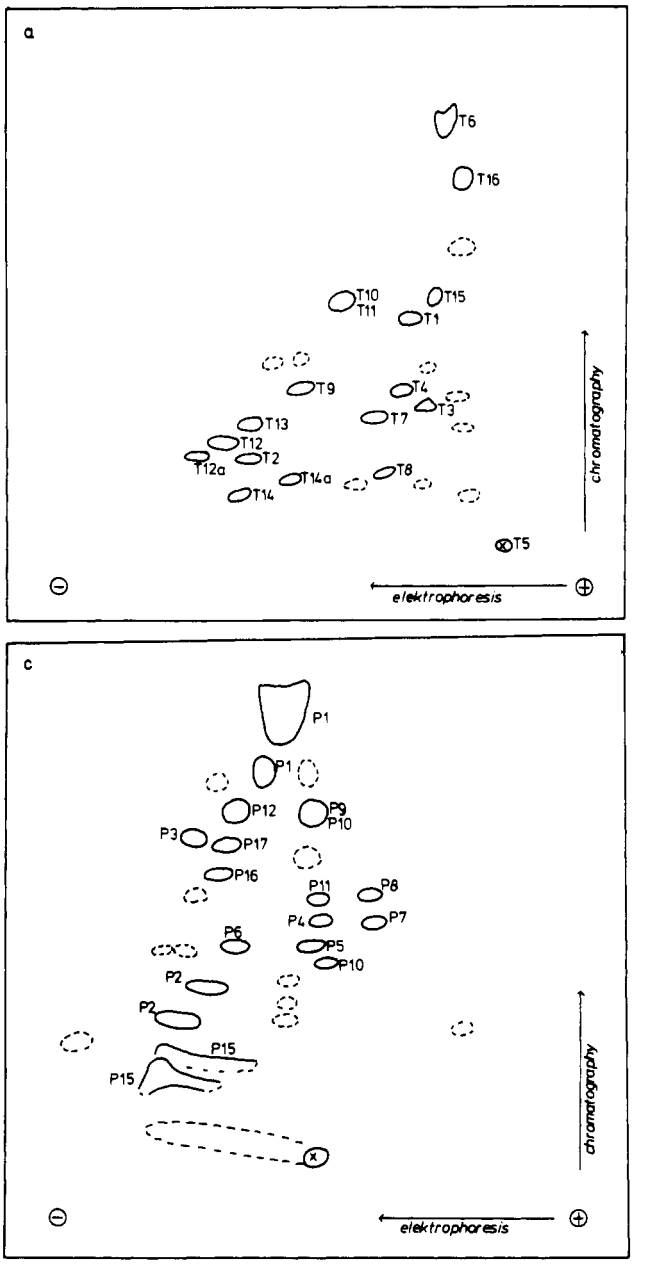


FIGURE 1: Thin-layer peptide maps from enzymatic digests of protein L21. × = origin; dotted lines indicate minor amounts of peptides generated by incomplete or unspecific digestion; (a) tryptic digest of the oxidized protein; (c) peptic digest. Figure 1b can be found in the supplementary material.

Tryptic Peptides. The tryptic thin-layer peptide map is shown in Figure 1a. In addition, the tryptic peptides were separated by gel filtration on Sephadex G-50 as shown in Figure 2a (see paragraph at end of paper regarding supplementary material) and purified by peptide mapping. T5 was insoluble in 10% acetic acid, and the pure peptide was isolated by centrifugation and washing before gel filtration on Sephadex G-50. The amino acid compositions of the tryptic peptides are presented in Table IIIa (see supplementary material). The tryptic peptides were sequenced by the micro-dansyl–Edman method. Since the identification of the PTH derivatives of the acidic amino acids and their amides on silica gel thin-layer sheets can give ambiguous results, the determination of Glu or Gln and Asp or Asn was performed by the DABITC method. The sequence methods applied are indicated by arrows.[1]

T1 (Residues 1–10).

Met-Tyr-Ala-Val-Phe-Gln-Ser-Gly-Gly-Lys

This peptide was completely sequenced by dansyl–Edman degradation.

T2 (Residues 11–13).

Gln-His-Arg

This peptide was not obtained in a pure form from peptide maps, and during the column run the N-terminal glutamine was converted into pyrrolidonecarboxylic acid. Thus, this peptide was not accessible to Edman degradation. The sequence was determined by degradation of the intact protein by a liquid-phase sequenator.

T3 (Residues 14–21).

Val-Ser-Glu-Gly-Gln-Thr-Val-Arg

The peptide was fully sequenced by both manual methods.

T4 (Residues 22–24).

Leu-Glu-Lys

The sequence was determined by dansyl–Edman degradation.

T5 (Residues 25–48).

Leu-Asp-Ile-Ala-Thr-Gly-Glu-Thr-Val-Glu-Phe-Ala-Glu-Val-Leu-
Met-Ile-Ala-Asn-Gly-Glu-Glu-Val-Lys

Cleavage of 8 nmol of T5 with α-chymotrypsin resulted in two peptides (positions 1–15 and 16–24 of peptide T5) which were separated by peptide mapping and identified by amino acid analysis (see Table IIIa, supplementary material). The sequence of peptide T5 was determined by dansyl–Edman degradation.

T6 (Residues 49–60).

Ile-Gly-Val-Pro-Phe-Val-Asp-Gly-Gly-Val-Ile-Lys

The peptide was obtained in a yield too low for further sequencing. The sequence was obtained from an identical CHD peptide (see below: CHD4).

T7 (Residues 61–68).

Ala-Glu-Val-Val-Ala-His-Gly-Arg

T8 (Residues 69–71).

Gly-Glu-Lys

T9 (Residues 72–73).

Val-Lys

T10 (Residues 74–76).

Ile-Val-Lys

T11 (Residues 77–78).

Phe-Arg

T12 (Residues 79–81).

Arg-Arg-Lys

T12a (Residues 81 and 85).

Lys

Lys was also determined by amino acid analysis without prior hydrolysis.

T13 (Residues 81–84).

Lys-His-Tyr-Arg

T13a (Residues 82–84).

His-Tyr-Arg

T14 (Residues 85–90).

Lys-Gln-Gln-Gly-His-Arg

T15 (Residues 91–97).

Gln-Trp-Phe-Thr-Asp-Val-Lys
T15-TH1

Peptide T15 was detected as a blue fluorescent spot under UV light at 366 nm and gave a pink color with Ehrlich's reagent. Subsequent cleavage of peptide T15 with thermolysin resulted in a peptide with a N-terminal Phe which was isolated by peptide mapping and sequenced by the DABITC method.

T16 (Residues 98–103).

Ile-Thr-Gly-Ile-Ser-Ala

T16 must be the C-terminal tryptic peptide of L21 since it lacks Arg or Lys. Furthermore, carboxypeptidase A, when incubated with the intact protein, released Ala, Ser, and Ile, whereas carboxypeptidase B released no amino acid.

Peptides Obtained from Tryptic Digestion after Blocking the Arg Residues with 1,2-Cyclohexanedione (CHD). The peptides were separated by gel filtration on Sephadex G-50 as shown in Figure 2b (see supplementary material) and purified by additional peptide mapping. Amino acid compositions of CHD peptides are presented in Table IIIb (see supplementary material). Besides the peptides corresponding to tryptic peptides (T1, T6, T9, T10, and T11), two overlapping peptides CHD2 and CHD5 could be isolated. Because of the irreversibility of the modification reaction, the blocked Arg residues are indicated as Arg* which elutes between His and Lys on the amino acid analyzer. The peptides were sequenced by the DABITC method, with the exception of CHD2 which was not accessible to Edman degradation because of the conversion of the N-terminal glutamine into pyrrolidonecarboxylic acid. This peptide was clearly identified by its amino acid composition. The peptide CHD4 corresponding to T6 was totally sequenced (see also Tryptic Peptides).

CHD2 (Overlapping T2–T3–T4).

Gln-His-Arg*-Val-Ser-Glu-Gly-Gln-Thr-Val-Arg*-Leu-Glu-Lys

CHD5 (Overlapping T7–T8).

Ala-Glu-Val-Val-Ala-His-Gly-Arg*-Gly-Glu-Lys

Peptides Obtained by Staphylococcus aureus Protease Digestion. The peptides were separated by gel filtration on Sephadex G-50 as shown in Figure 2c (supplementary material). The first peak contained the pure peptide SP9 whereas all other peptides were purified by additional chromatography or peptide mapping. The amino acid compositions of the peptides are presented in Table IIIc (see supplementary material). The peptides gave the necessary information for further alignment of tryptic peptides in protein L21 (Figure 3).

SP1 (Residues 1–16).

Met-Tyr-Ala-Val-Phe-Gln-Ser-Gly-Gly-Lys-Gln-His-Arg-Val-Ser-Glu

SP2 (Residues 17–23).

Gly-Gln-Thr-Val-Arg-Leu-Glu

SP3 (Residues 24–31).

Lys-Leu-Asp-Ile-Ala-Thr-Gly-Glu

SP4 (Residues 32–34).

Thr-Val-Glu

SP5 (Residues 35–37).

Phe-Ala-Glu

SP6 (Residues 38–46).

Val-Leu-Met-Ile-Ala-Asn-Gly-Glu-Glu

SP7 (Residues 47–62).

Val-Lys-Ile-Gly-Val-Pro-Phe-Val-Asp-Gly-Gly-Val-Ile-Lys-Ala-Glu

SP8 (Residues 63–70).

Val-Val-Ala-His-Gly-Arg-Gly-Glu

SP9 (Residues 71–103).

Lys-Val-Lys-Ile-Val-Lys-Phe-Arg-Arg-Arg-Lys-His-Tyr-Arg-Lys-Gln-
Gln-Gly-His-Arg-Gln-Trp-Phe-Thr-Asp-Val-Lys-Ile-Thr-Gly-Ile-Ser-
Ala

The first 25 residues of peptide SP9 were sequenced by the DABITC method. In addition, aliquots (7 nmol) of peptide SP9 were cleaved with trypsin, thermolysin, and pepsin. The peptides were isolated on peptide maps and identified by their amino acid composition (see Table IIId, supplementary material) as shown in Figure 3.

Thermolysin Peptides. A peptide map of a thermolysin digest of protein L21 is shown in Figure 1b (see supplementary material). The amino acid composition of the peptides (Table IIIe, supplementary material) and the sequences determined are given in the supplementary material. The peptides were separated on a Dowex 50 M71 column (Figure 2d, supplementary material). All peptides had to be purified by additional thin-layer chromatography or peptide mapping.

Peptic Peptides. The peptide map of the peptic peptides is shown in Figure 1c, and for the separation of the peptides on Sephadex G-50 see Figure 2e, supplementary material. The peptides provided additional information for overlaps and confirmed the previously sequenced regions. Most of the peptic peptides were sequenced completely or at least partially. Their amino acid compositions are given in Table IIIf and can, together with the sequence results, be found in the supplementary material.

Cyanogen Bromide Fragments. The cleavage of protein L21 with cyanogen bromide (see Table I) resulted in the two fragments which were separated on Sephadex G-50 (see Figure 2f in the supplementary material). The amino acid composition of both fragments was in good agreement with the amino acid composition of the whole protein and with the established sequences (Table IIIg, supplementary material). Since the N-terminal sequence of the protein was determined by liquid-phase degradation and by sequence analyses of peptides from several digestions, only the C-terminal fragment was sequenced by means of a liquid-phase sequenator up to position 39 of that fragment (see Figure 3). In addition, this fragment was digested with trypsin, α-chymotrypsin, and pepsin. The tryptic peptides were separated by peptide mapping and clearly identified by their amino acid composition. The peptic peptides were similarly separated and were partially sequenced to establish their identities. The chymotryptic peptides which were not well separated on peptide maps were purified on Sephadex G-50 and additional thin-layer chromatography. They were identified by their amino acid compositions and by sequencing them with the DABITC method. The results are given in Table IIIg (supplementary material).

Cleavage with BNPS-skatole. The peptides obtained after cleavage with BNPS-skatole (see Table I) were separated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis. Besides the uncleaved protein, only one fragment, slightly smaller than the intact protein (see Figure 2g, supplementary material), was stained with Coomassie blue, indicating that the tryptophan residue must be near one end of the protein. This is in agreement with the established sequence.

Liquid-Phase Edman Degradation. The N-terminal region of protein L21 was established by liquid-phase Edman degradation on the intact protein, giving clear results up to position 49 (see Figure 3). The results of the liquid-phase sequenator runs on CNBr2 are shown in Figure 3.

Complete Sequence of Protein L21. The complete primary structure of protein L21, deduced from the results presented here, is shown in Figure 3.

All tryptic peptides were isolated, and all but T2 (residues 11–13) and T15 (residues 91–97) were sequenced completely. The sequence of T2 was established unambiguously from the degradation of the intact protein by a liquid-phase sequenator and from several other overlapping peptides. T15 contained tryptophan, as shown by the Ehrlich reagent and fluorescence under UV light. Its sequence could be completed by subsequent cleavage with thermolysin, resulting in the C-terminal peptide, lacking tryptophan, which was sequenced by the DABITC method.

The alignment of all tryptic peptides can be summarized as follows. The total sequence was identified in peptides from the staphylococcal protease digest. SP1 gives the order of T1, T2, and T3; SP2 gives the tryptic peptides T3 and T4. SP3 links T4 and T5. SP7 gives the order of T5, T6, and T7; SP8 overlaps T7 and T8. Finally, SP9 gives the order of T8–T16. The overlaps for T4–T5 by SP3 and T8–T9 by SP9 include only one residue, but other peptides determine the order of these tryptic peptides (peptide P4, which overlaps T4 and T5, and P15, which gives the order of T7–T15).

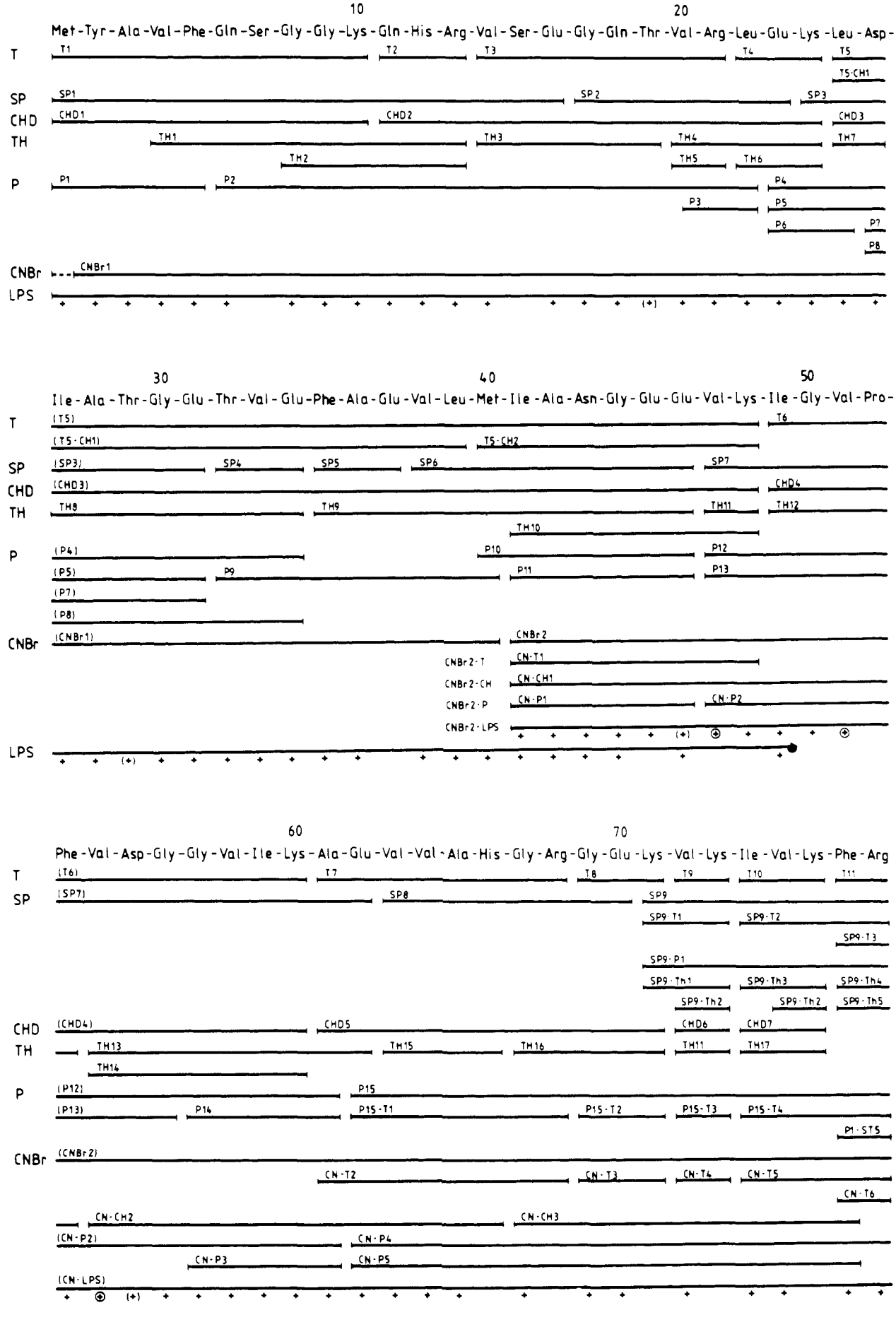
All other peptides confirmed the sequence shown in Figure 3. No peptides were found which were inconsistent with the presented sequence. The electrophoretic mobilities of all peptides on thin-layer cellulose were consistent with the assigned amide and acid residues.

Every position in protein L21 was sequenced several times. The total amount of protein used for sequence determination was 25 mg of the lyophilized material.

Discussion

Protein L21 consists of a single peptide chain of 103 amino acids (Figure 3) with the following amino acid composition: Asp_3, Asn_1, Thr_5, Ser_3, Glu_9, Gln_6, Pro_1, Gly_{12}, Ala_7, Val_{14}, Met_2, Ile_7, Leu_3, Tyr_2, Phe_5, His_4, Lys_{10}, Arg_8, and Trp_1. Its molecular weight is 11 565, which is in agreement with the mobility of protein L21 on sodium dodecyl sulfate gel electrophoresis. The amino acid sequence is in excellent agreement with the measured amino acid composition (Table II).

No peptide mixture derived from enzymatic digestion of the whole protein could be completely separated on peptide maps so that all of the peptides with the exception of the thermolytic peptides were isolated by gel filtration and additional peptide mapping. With the DABITC method, it was possible to


10
20
Met-Tyr-Ala-Val-Phe-Gln-Ser-Gly-Gly-Lys-Gln-His-Arg-Val-Ser-Glu-Gly-Gln-Thr-Val-Arg-Leu-Glu-Lys-Leu-Asp-
T
SP
CHD
TH
P
CNBr
LPS
30
40
50
Ile-Ala-Thr-Gly-Glu-Thr-Val-Glu-Phe-Ala-Glu-Val-Leu-Met-Ile-Ala-Asn-Gly-Glu-Glu-Val-Lys-Ile-Gly-Val-Pro-
CNBr2
CNBr2-T
CNBr2-CH
CNBr2-P
CNBr2-LPS
60
70
Phe-Val-Asp-Gly-Gly-Val-Ile-Lys-Ala-Glu-Val-Val-Ala-His-Gly-Arg-Gly-Glu-Lys-Val-Lys-Ile-Val-Lys-Phe-Arg

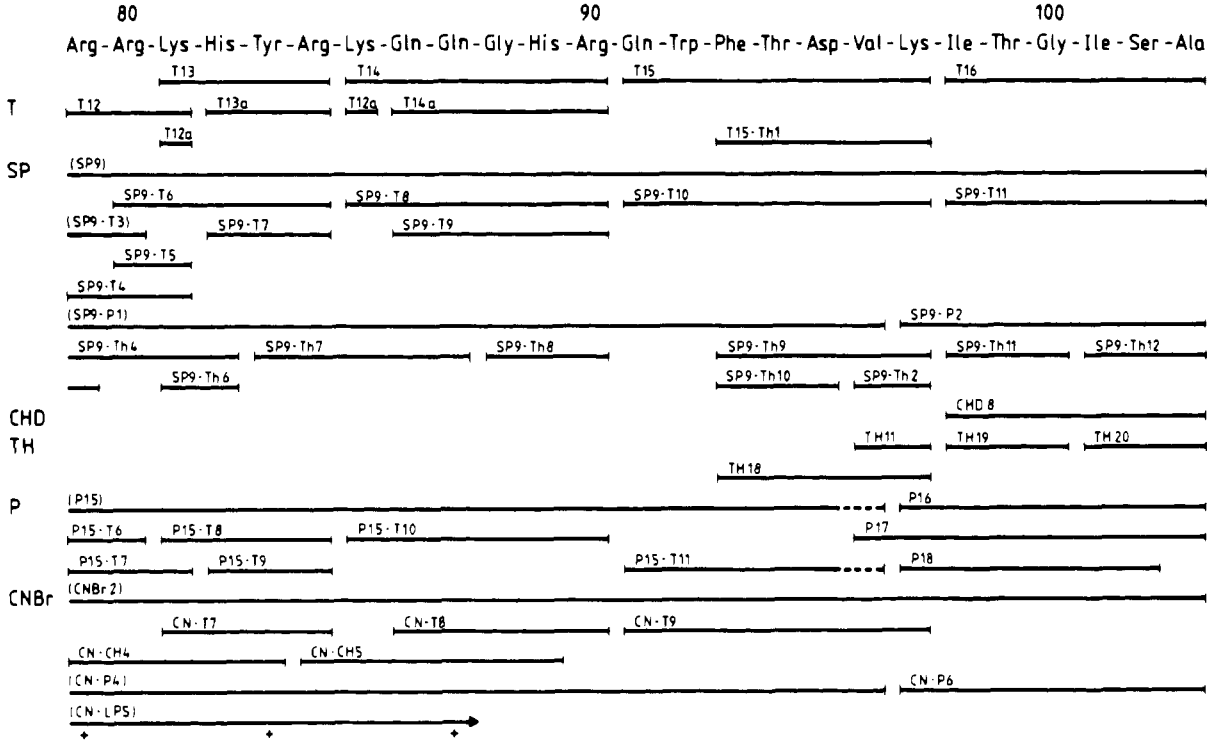


FIGURE 3: Primary structure of the ribosomal protein L21 of *E. coli*. →, liquid-phase sequenator; +, unambiguous identification; ⊕, identified by high-pressure liquid chromatography; (+), weaker evidence than in the case of +.

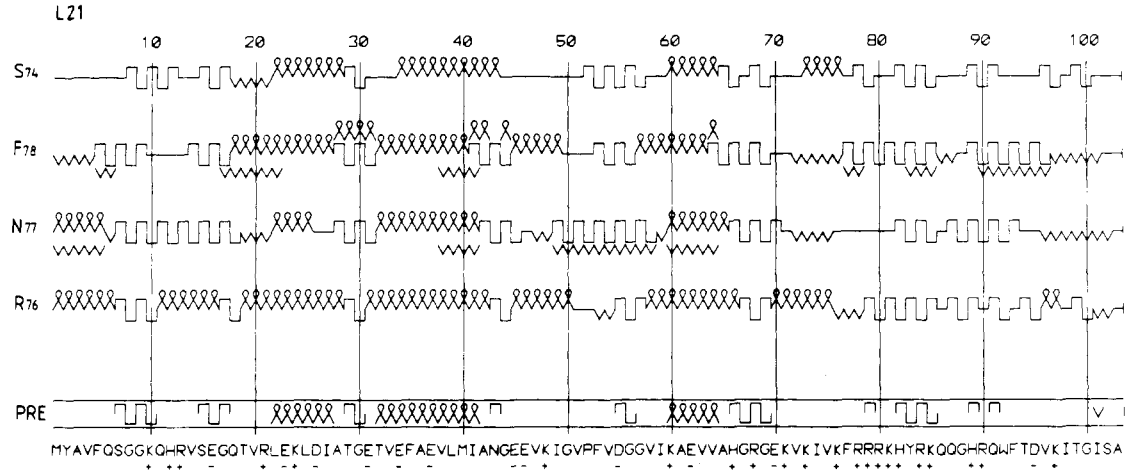


FIGURE 4: Predicted secondary structure of protein L21 according to four methods. S_{74}: Burgess (1974). F_{78}: Chou & Fasman (1977, 1978). N_{77}: Nagano (1977). R_{76}: Robson & Suzuki (1976). The symbols define residues in helix (α), extended (~), β-turn or loop (η), and random-coil (—) conformations.

sequence 5 nmol of peptide up to position 25 (see SP9), whereas in case of the micro-dansyl–Edman method at least 30 nmol of this peptide would have been necessary. Thus, it was possible to start with only 100 nmol of protein for a digestion and to isolate all peptides in sufficient amounts in a very pure state for amino acid analysis and sequencing. Basic amino acids are clustered in positions 71–90 whereas the acidic amino acids are distributed throughout the protein chain. Hydrophobic amino acids occur only within short regions (positions 51–54 and 63–65). Protein L21 shares only one identical pentapeptide and six tetrapeptides with other ribosomal proteins whose primary structures have been determined (see Table IV, supplementary material). This degree of homology is probably not statistically significant. Based on the amino acid sequence of L21, secondary structure calculations were performed as described recently (Dzionara et al., 1977). Protein L21 is predicted to contain 20% helix, 23% turns, and no β structure (see Figure 4).

Acknowledgments

We thank Dr. H. G. Wittmann for the sample of protein L21, A. Geissler for liquid-phase sequencing, Dr. P. Wooley for the spectral determination of tryptophan in the protein, and Dr. G. Allen for critically reading the manuscript.

Supplementary Material Available

The sequence of the thermolysin and peptic peptides, Table IIIa–g containing the amino acid composition of peptides, Table IV presenting homologies to other ribosomal proteins, Figure 1b showing the thermolysin peptide map, Figure 2a–f illustrating the separation of the different digests on columns, and Figure 2g showing the separation of the BNPS-skatole

fragments by gel electrophoresis (21 pages). Ordering information is given on any current masthead page.

References

Braunitzer, G., Schrank, B., Ruhfus, A., Petersen, S., & Petersen, U. (1971) *Hoppe-Seyler's Z. Physiol. Chem. 352*, 1730–1732.

Bruton, C. J., & Hartley, B. S. (1970) *J. Mol. Biol. 52*, 165–178.

Burgess, A. W., Ponnuswamy, P. K., & Scheraga, H. A. (1974) *Isr. J. Chem. 12*, 239–286.

Chang, J. Y., Creaser, E. H., & Bentley, K. W. (1976) *Biochem. J. 153*, 607–611.

Chang, J. Y., Brauer, D., & Wittmann-Liebold, B. (1978) *FEBS Lett. 93*, 205–214.

Chen, R., & Wittmann-Liebold, B. (1975) *FEBS Lett. 52*, 139–140.

Chou, P. Y., & Fasman, G. D. (1977) *J. Mol. Biol. 115*, 135–175.

Chou, P. Y., & Fasman, G. D. (1978) *Annu. Rev. Biochem. 47*, 251–276.

Dabbs, E. R. (1978) *Mol. Gen. Genet. 165*, 73–78.

Dzionara, M., Robinson, S. M. L., & Wittmann-Liebold, B. (1977) *Hoppe-Seyler's Z. Physiol. Chem. 358*, 1003–1019.

Edelhoch, H. (1967) *Biochemistry 6*, 1948–1954.

Edman, P., & Begg, G. (1967) *Eur. J. Biochem. 1*, 80–91.

Heiland, I., Brauer, D., & Wittmann-Liebold, B. (1976) *Hoppe-Seyler's Z. Physiol. Chem. 357*, 1751–1770.

Hindennach, I., Stöffler, G., & Wittmann, H. G. (1971) *Eur. J. Biochem. 23*, 7–11.

Hirs, C. H. W. (1956) *J. Biol. Chem. 219*, 611–621.

Kaltschmidt, E., & Wittmann, H. G. (1970) *Proc. Natl. Acad. Sci. U.S.A. 67*, 1276–1282.

Kenny, J. W., & Traut, R. R. (1979) *J. Mol. Biol. 127*, 243–263.

Kitakawa, M., Dabbs, E. R., & Isono, K. (1979) *J. Bacteriol.* (in press).

Kühlbrandt, W., & Garrett, R. A. (1978) *FEBS Lett. 94*, 207–212.

Nagano, K. (1977) *J. Mol. Biol. 109*, 251–274.

Penke, B., Ferenczi, R., & Kovács, K. (1974) *Anal. Biochem. 60*, 45–50.

Robson, B., & Suzuki, E. (1976) *J. Mol. Biol. 107*, 327–356.

Smith, I. (1953) *Nature (London) 171*, 43.

Spillmann, S., Dohme, F., & Nierhaus, K. H. (1977) *J. Mol. Biol. 115*, 513–523.

Strickland, M., Strickland, W. N., Brandt, W. F., Von Holt, C., Wittmann-Liebold, B., & Lehmann, A. (1978) *Eur. J. Biochem. 89*, 443–452.

Takata, R. (1978) *Mol. Gen. Genet. 160*, 151–155.

Toi, K., Brynum, E., Norris, E., & Hano, H. (1967) *J. Biol. Chem. 242*, 1036–1043.

Vanderkerckhove, J., & Van Montagu, M. (1974) *Eur. J. Biochem. 44*, 279–288.

Wittmann-Liebold, B. (1973) *Hoppe-Seyler's Z. Physiol. Chem. 354*, 1415–1431.

Wittmann-Liebold, B., Geissler, A. W., & Marzinzig, E. (1975) *J. Supramol. Struct. 3*, 426–447.

Wittmann-Liebold, B., Graffunder, H., & Kohls, H. (1976) *Anal. Biochem. 75*, 621–633.

Electrostatic Stabilization in Myoglobin. pH Dependence of Summed Electrostatic Contributions†

Stephen H. Friend and Frank R. N. Gurd*

ABSTRACT: The modified Tanford–Kirkwood theory of Shire et al. [Shire, S. J., Hanania, G. I. H., & Gurd, F. R. N. (1974) *Biochemistry 13*, 2967] for electrostatic interactions was applied to compute the free energy contributions from individual pairs of charge loci in sperm whale ferrimyoglobin. Such interaction energies depend not only on the fractional occupancy of each charge site but also on their static solvent accessibilities, in addition to the geometrical and other factors inherent to the treatment. The pH-dependent unfolding of the myoglobin in acid solution is strongly influenced by ionic strength in such a way that the native form is relatively destabilized by increased ionic strength. Since the titration behavior of the denatured form is less sensitive to ionic strength variation than that of the native form, it follows that the native form experiences net stabilization from intramolecular electrostatic interactions. The unfolded forms are likewise stabilized by ionic equilibria as a result of protonation of histidine residues that are masked in the native state but exposed in the denatured state. The summed electrostatic free energy of the native structure shows a broad maximum at about pH 6.5, in keeping with the observed thermal stability maximum [Acampora, G., & Hermans, J., Jr. (1967) *J. Am. Chem. Soc. 89*, 1543], with a net maximum stabilization of approximately 10.6 kcal/mol at 0.00 M, 10.0 kcal/mol at 0.01 M, and 8.6 kcal/mol at 0.10 M ionic strength. By difference it may be estimated that 5–6 kcal/mol of stabilization of the native protein structure can be ascribed to nonionic factors. The charge sites play a subtle dual role in both stabilizing and enhancing the water solubility of the protein.

Shire et al. (1974a,b, 1975) introduced into the Tanford–Kirkwood model of intramolecular electrostatic interactions (Tanford & Kirkwood, 1957; Tanford & Roxby, 1972) a term to take into account the degree of exposure to the solvent of each charged site. The treatment incorporates the static solvent accessibility values for each charged site derived from the crystallographic structural data (Lee & Richards, 1971) and employs a consistent set of intrinsic pK values. The first applications dealt with the overall titration curves of various ferrimyoglobins (Shire et al., 1974a,b, 1975), since then

† From the Department of Chemistry and the Medical Sciences Program, Indiana University, Bloomington, Indiana 47405. *Received April 13, 1979.* This work was supported by U.S. Public Health Service Research Grant HL-05556. This is the 109th paper in a series dealing with coordination complexes and catalytic properties of proteins and related substances. For the preceding paper see Gurd & Rothgeb (1979).

0006-2960/79/0418-4612$01.00/0