Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 1291–1294, February 1982
Medical Sciences

Primary structure analysis of the major internal protein p24 of human type C T-cell leukemia virus

(human RNA tumor virus/bovine leukemia virus/retrovirus proteins/amino acid sequence homology/T-cell growth factor)

STEPHEN OROSZLAN*, M. G. SARNGADHARAN†, TERRY D. COPELAND*, V. S. KALYANARAMAN‡, RAYMOND V. GILDEN*, AND ROBERT C. GALLO†

*Biological Carcinogenesis Program, Frederick Cancer Research Center, Frederick, Maryland 21701; †Laboratory of Tumor Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205; and ‡Department of Cell Biology, Litton Bionetics, Inc., Kensington, Maryland 20795

Communicated by Lloyd J. Old, November 16, 1981

ABSTRACT **A human type C retrovirus [human T-cell leukemia (lymphoma) virus; HTLV], recently isolated from young adult patients with cutaneous T-cell lymphoma or leukemia, was not detectably related to other known animal retroviruses in molecular hybridization studies, by comparison of reverse transcriptase and the major core protein p24. The p24 core protein was purified to homogeneity. The amino acid composition, the COOH-terminal amino acid, and the NH_2-terminal amino acid sequence of the first 25 residues of this major internal structural protein were determined. These results were then compared to the known structure of the internal core protein of other retroviruses. The compositional data reveal that HTLV p24 is chemically distinct from p30–p24 of other animal retroviruses, in agreement with the earlier immunological analyses. However, HTLV p24 shares the common NH_2-terminal proline and COOH-terminal leucine of all mammalian type C viral p30s. In addition, like bovine leukemia virus (BLV), HTLV lacks the common prolylleucylarginine tripeptide and the larger conserved region found near the NH_2 terminus of the other mammalian type C viral p30s. Alignment of the amino acid sequence of HTLV p24 with previously determined sequences of other retrovirus proteins, including BLV p24, reveals statistically significant sequence homology only to BLV. The results reported here demonstrate that HTLV p24 is related to but chemically distinct from the major core protein of other retroviruses. Similarly, previous results showed that there was no immunological crossreactivity of the p24 protein and reverse transcriptase of HTLV with other retroviruses, including BLV, and no nucleic acid sequence homology. However, the present results, combined with the common size of the p24 and reverse transcriptase, suggest that HTLV may be closer to BLV than any other known retrovirus.**

A type C retrovirus, designated human T-cell leukemia (lymphoma) virus (HTLV), was isolated from the cultured T cells of a patient (CR) with cutaneous T-cell lymphoma (1). Nucleic acid hybridization analysis showed that HTLV was an acquired virus (2), rather than an endogenous virus transmitted through the germ line (unpublished data). More recently, a second identical or closely related HTLV isolate was obtained from the T cells established in culture from a patient (MB) with cutaneous T-cell leukemia (3). Growth of these cells and isolation of HTLV were possible only by using a specific T-cell lymphokine, designated T-cell growth factor (4–6). Partially purified lectin-free T-cell growth factor stimulates the selective growth of antigen-stimulated normal T cells. On the other hand, long-term growth of several lymphoblastoid T-cell lines from human lymphocytic neoplasias have been achieved by using partially purified T-cell growth factor in the absence of any antigen or lectin stimulation (7). In retrospect, this unique capability to grow T-cell populations from specific classes of T-cell malignancies seems to have been critical for the isolation of the virus, as virus expression was detected only in cultured T cells and not in cultured B cells from the same patient (CR) (unpublished data).

HTLV is a unique retrovirus in man with no significant crossreactivity with other known mammalian retroviruses. This is indicated by results of nucleic acid hybridization studies (2), by biochemical and immunological comparisons of the reverse transcriptase (8), and by immunological analysis of the major structural protein p24 (9). High titer antibodies against HTLV have been detected in the sera of some patients with T-cell malignancies, whereas such activities have been negative in random normal subjects (10, 11). These findings suggest a potential relevance of this or related viruses in the etiology of certain human T-cell leukemias and lymphomas.

In our previous studies, primary structure analyses of the structural proteins of retroviruses provided evidence for their definitive genetic relatedness, which could not be detected by hybridization and immunological methods (12). Specifically, BLV p24 and FeLV p27, although unrelated by the above criteria, were found to have statistically significant sequence homology, allowing the suggestion that these contemporary viral proteins may have originated from a common progenitor. Here we report on the primary structure of core protein p24 of HTLV and its comparison to protein homologs of other mammalian retroviruses.

MATERIALS AND METHODS

Purification of HTLV p24. Density-banded HTLV was lysed with 0.5% Triton X-100 and 0.8 M NaCl in 20 mM Tris·HCl, pH 7.5/1 mM dithiothreitol/20% (vol/vol) glycerol/0.5 mM phenylmethylsulfonyl fluoride. Nucleic acids were removed from the clarified lysate by batch adsorption on DEAE-cellulose at 0.3 M NaCl. The unadsorbed proteins were dialyzed against 10 mM *N,N*-bishydroxyethylaminoethanesulfonic acid, pH 6.5/1 mM EDTA/0.1 mM phenylmethylsulfonyl fluoride and chromatographed on a column of phosphocellulose equilibrated with the same buffer eluting with a gradient between 0 and 0.6 M NaCl. Fractions containing the peak of p24, as judged by $NaDodSO_4$/polyacrylamide gel electrophoresis, were pooled and further purified by gel filtration on a Bio-Gel P-60 column. The nearly homogeneous protein thus obtained was subjected to reverse-phase high-performance liquid chromatography (HPLC) on a μBondapak C_{18} column (Waters Associates). (See the legend to Fig. 1 for details.)

Amino Acid Analysis. Protein samples were hydrolyzed *in vacuo* at 110°C for 24 hr with 6 M HCl/0.1% phenol. Amino

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

Abbreviations: HTLV, human T-cell leukemia (lymphoma) virus; BLV, bovine leukemia virus; FeLV, feline leukemia virus; HPLC, high-performance liquid chromatography.

acid analyses were performed on a Durrum 500 analyzer. A computer program (13) was used to analyze the compositional data as described (14).

Microsequence Analysis. Semiautomated microsequence analysis utilizing stepwise Edman degradation (15) was performed with a Beckman sequencer model 890C equipped with a cold trap accessory as described (16). Phenylthiohydantoin derivatives of amino acids were identified and quantitated by HPLC (17).

Carboxypeptidase Digestion. Digestion with carboxypeptidase A (Pierce) was performed as described (18).

RESULTS

Purity of HTLV p24 and Amino Acid Composition. The purity of HTLV p24 as obtained by ion-exchange chromatography and gel filtration was confirmed by $NaDodSO_4$/polyacrylamide slab gel electrophoresis and by HPLC (Fig. 1). For all the chemical analyses performed, HPLC-purified (final step) p24 was used. The computer-assisted amino acid compositional data of HTLV p24 is given in Table 1. The total number of amino acids (minus cysteine and tryptophan) is 236. The molecular weight calculated from the composition is 25,933. A comparison of these data with those of other retrovirus p24–p30 core proteins indicates an overall similarity, but a closer scrutiny reveals that the composition of HTLV p24 is closer to that of BLV p24 than to that of FeLV p27. When the HTLV p24 composition is compared with that of BLV p24, the major differences appear in the numbers for histidine, methionine, and leucine (which are greater) and for alanine and isoleucine (which are lesser) in HTLV p24 than in BLV p24. We wish to note that the number

Table 1. Amino acid composition of the major internal proteins of HTLV, BLV, and FeLV

Amino acid	Residues per protein		
	HTLV p24	BLV p24*	FeLV p27*
Lys	12	11	13
His	9	3	4
Arg	11	11	22
Asp	21	26	23
Thr	10	13	14
Ser	14	17	14
Glu	36	41	50
Pro	22	22	23
Gly	15	13	11
Ala	24	33	20
Val	7	9	10
Met	4	1	2
Ile	8	13	7
Leu	32	23	37
Tyr	6	5	6
Phe	5	3	7
Total†	236	244	263

* Taken from Oroszlan *et al.* (12).
† Minus Cys and Trp.

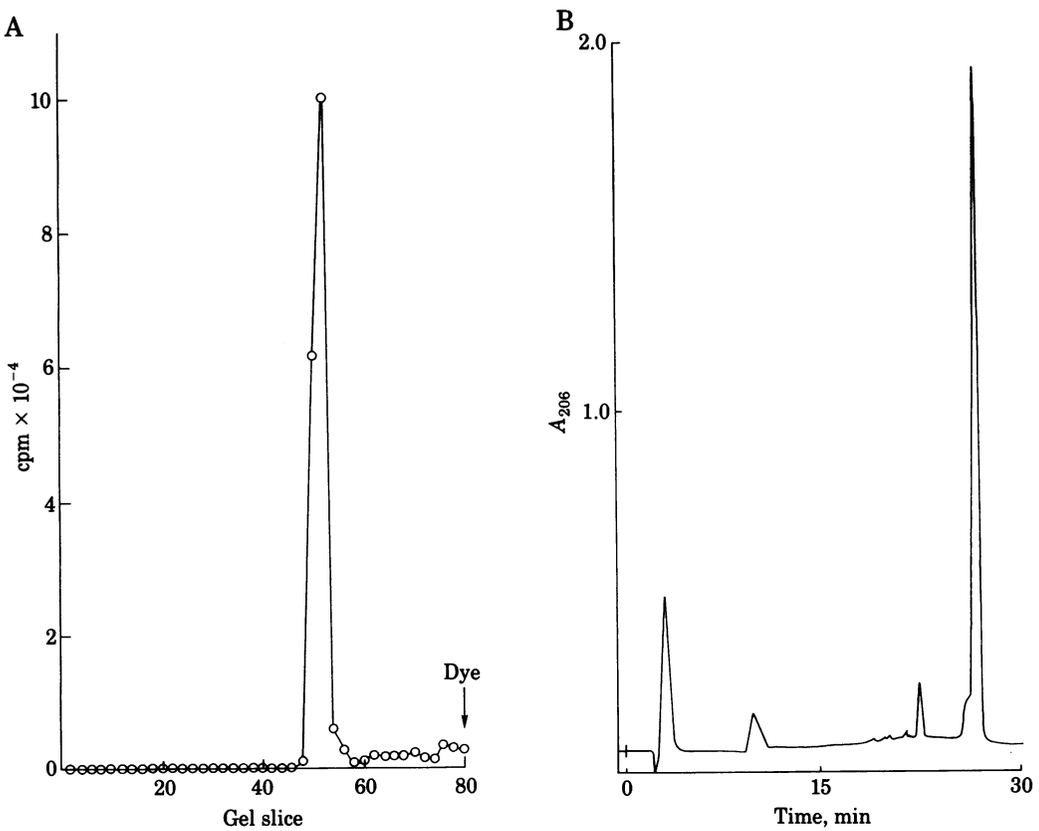

FIG. 1. $NaDodSO_4$/polyacrylamide gel electrophoresis (*A*) and HPLC (*B*) profiles of HTLV p24. (*A*) An aliquot of the Bio-Gel P-60 column fraction of p24 was labeled with ^{125}I (19) and analyzed by electrophoresis in the presence of $NaDodSO_4$ on a 12% polyacrylamide cylindrical gel as described by Laemmli (20). The gel was sliced into 1-mm slices, and the radioactivity was determined. A molecular weight of 24,000 was determined by comparison with marker proteins run on parallel gels. (*B*) Purified p24 dissolved in 0.01 M phosphate buffer (pH 7.5) was acidified with trifluoroacetic acid to pH 2 and applied to a μBondapak C_{18} column. The column was developed with a linear acetonitrile (0–60%) gradient consisting of 0.05% trifluoroacetic acid in water/0.05% trifluoroacetic acid in acetonitrile. The protein recovered under the major peak was used for the chemical analysis.

Table 2. Polarities of HTLV p24, BLV p24, FeLV p27*

Protein	% of total amino acids†		
	Polar	Intermediate	Nonpolar
HTLV p24	34.0	23.0	43.0
BLV p24	36.5	20.9	42.6
FeLV p27	41.1	18.6	40.3

* Calculated from the data of Table 1 as described by Capaldi and Vanderkoi (22).
† Minus Cys and Trp.

of residues of histidine is substantially greater in HTLV p24 than in the major internal proteins of any other retrovirus analyzed thus far (21). The polarities of HTLV p24, BLV p24, and FeLV p27 calculated from the respective amino acid composition are given in Table 2. The comparison shows that whereas HTLV p24 and BLV p24 are similar, both are less polar than FeLV p27. It is also clear, however, that by this criterion, HTLV p24 resembles retrovirus major core proteins more closely than the other lower molecular weight *gag* gene products (21). This provides additional support for considering HTLV p24 as the structural homolog of mammalian retrovirus p30s.

NH_2-Terminal Amino Acid Sequence of HTLV p24. To determine the NH_2-terminal amino acid sequence of HTLV p24, 2 nmol of protein was degraded in a single microsequence analysis. The yields of the various amino acids at each sequence cycle are shown in Fig. 2. Unambiguous identification of phenylthiohydantoin derivatives of amino acids was possible up to residue 25, except at positions 10 and 17, where no amino acid could be assigned. The NH_2-terminal amino acid sequence of HTLV p24 is given in Fig. 3. A comparison with our library of retroviral protein sequences revealed that HTLV p24, while sharing the common proline NH_2 terminus of all retrovirus major core proteins, has a unique NH_2-terminal amino acid sequence distinct from any other retrovirus gene product for which analytical data are available.

COOH-Terminal Amino Acids. Digestion of HTLV p24 with carboxypeptidase A for 20 min yielded 0.8 nmol of leucine per

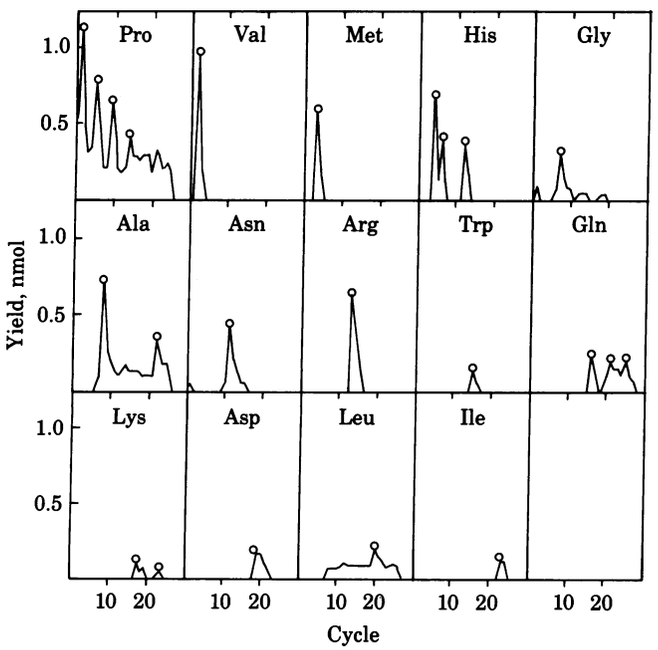

FIG. 2. Yields of amino acids obtained by degrading 2 nmol of HTLV p24. The data plotted were from analysis of phenylthiohydantoin amino acids by HPLC. Yields are normalized to 100% of the sample.

```
1          5          10         15
Pro-Val-Met-His-Pro-His-Gly-Ala-Pro-X-Asn-His-Arg-Pro-Trp-

      20         25
Gln-X-Lys-Asp-Leu-Gln-Ala-Ile-(Lys)-Gln-
```

FIG. 3. NH_2-terminal amino acid sequence of HTLV p24. X, unidentified.

nmol of protein. This finding shows that HTLV p24 and BLV p24 (12) have an identical COOH terminus—namely, leucine, which is shared with all other mammalian retrovirus p30s (21).

NH_2-Terminal Sequence Homology of HTLV p24 with BLV p24. To determine whether significant sequence homology exists between HTLV p24 and the major core proteins of the known retroviruses, we carried out a statistical analysis. The computer program ALIGN of Dayhoff (23) was used to compare sequences. This program is based on the mutation data matrix for scoring amino acid comparisons. The alignment of the NH_2-terminal sequence of HTLV p24 with sequences of BLV p24 and FeLV p27 is shown in Fig. 4. It can been seen that HTLV p24 and BLV p24 sequences can be aligned, resulting in 9 common amino acid residues out of 24 possible comparisons. To obtain such an alignment, the introduction of only a single break was necessary in the BLV p24 sequence. An alignment score (number of standard deviations of real score above random score) of 6.5 was obtained for the two p24s. This indicates that the probability is only 0.4×10^{-10} that the similarity between these sequences is the result of chance alone. This highly significant value indicates evolutionary relatedness between HTLV p24 and BLV p24. In contrast, there are only three identities obtained between HTLV p24 and FeLV p27, and this low alignment score ($\approx$2.8) is not considered to be statistically significant. For the BLV p24/FeLV p27 pair, however, our previous analysis (12) gave a statistically significant (>3) alignment score of 3.8. A comparison of the HTLV p24 sequence with the other known retrovirus p30s, or in fact with any other *gag* gene product, indicates that HTLV p24 is more closely related to BLV p24 than to any other retrovirus protein sequence.

DISCUSSION

HTLV is a unique human retrovirus isolated from the T-cell cultures of some patients with leukemias or lymphomas involving relatively mature T cells. It has been identified as an exogenously acquired virus with no immunological crossreactivity (9) or nucleic acid sequence homology (2) with known animal retroviruses. An initial limited epidemiological survey has detected significant antibody titers to HTLV antigens in the sera of some patients with T-cell malignancies and that of one close healthy relative of one of the virus- and seropositive patients (10, 11). This stimulated study of possible distant relationships

```
          1    5    10   15   20   25
HTLV      P V M H P H G A P X N H R P W Q X K D L Q A I K Q-

BLV       P I I * S E G N R N R H R A W A L R E L Q D I K K-

FeLV      P L * * R E G P N N R P Q Y W P F S A * S D L Y N-
```

FIG. 4. Alignment of the NH_2-terminal sequence of HTLV p24 with those of BLV p24 and FeLV p27.

that might exist between HTLV and other animal retroviruses. Although morphologic characteristics and mode of reproduction indicate HTLV to be similar to type C retroviruses, many biochemical and immunologic properties show a clear distinction for HTLV from other animal type B, type C, and type D retroviruses. Chemical analysis and direct sequence determination of viral genes or gene products provide the basis for quantitative estimates of relatedness and allow comparison at a level less stringent and, therefore, more sensitive than that afforded by immunologic means.

The study reported here concerns only the major core protein p24 of HTLV. The results of chemical analysis of HTLV p24 showed overall similarities with other retroviral p30s but, at the same time, showed some distinct differences. For instance, the total number of lysine and arginine residues is much lower than in other mammalian type C viral p30s, and the number of histidine and methionine residues is much greater (21). The amino acid composition of HTLV p24, and particularly the distribution between polar and nonpolar groups, is more similar to BLV p24 than to any other mammalian retroviral p30 (Table 2; ref. 21). However, there is more similarity to p30s than to any other retroviral *gag* proteins.

HTLV p24 shares the common NH_2-terminal proline and COOH-terminal leucine of all mammalian type C viral p30s. This may indicate conservation of progenitor cleavage sites on the precursor polyprotein (18). However, like BLV p24, HTLV p24 lacks the NH_2-terminal tripeptide Pro-Leu-Arg and the highly conserved region near the NH_2-terminus of the mammalian type C retroviral p30s. A computer-assisted alignment of the first 25 amino acid residues of HTLV p24 with those of BLV p24 resulted in 9 common amino acid residues out of a possible 24 comparisons, yielding an alignment score of 6.5. Statistically, this is a highly significant value that emphasizes the evolutionary relationship between HTLV and BLV. The extrapolation from structural homology to biologic function is noncontroversial at the protein level; however, whether this indicates similar biologic significance of the viruses (oncogenic potential) remains to be established by the ongoing epidemiological studies. The current data do serve to further justify detailed studies of HTLV.

The degree of homology detected between these two viral proteins is not sufficient to cause immunological crossreactivity, thus confirming our earlier observations (9). However, this does not rule out the existence of homologous regions further in the molecule of sufficient length to qualify as potential antigenic determinants. Such regions may not be antigenic in the native molecule because of specific steric characteristics and, therefore, may not contribute to crossreactivity in conventional test systems. Synthetic peptides representing such homologous sequences could be used to generate "interspecies" immune reagents applicable in screening biological material for related viral expression. Further, knowledge of the amino acid sequence of this major HTLV protein provides an opportunity to synthesize corresponding oligodeoxyribonucleotide sequences useful in cloning the provirus of HTLV—a task that has been hampered to date by the poor yield of the virus.

Note Added in Proof. Since the submission of this paper, we raised antiserum to purified BLV p12 (the basic nucleic acid binding protein). This serum reacted not only with BLV p12 but also with its HTLV homolog, indicating antigenic relatedness between the bovine and human viruses.

The authors wish to acknowledge the excellent technical assistance provided by M. Jarvis, Y. Kim, and G. Smythers.

1. Poiesz, B. J., Ruscetti, F. W., Gazdar, A. F., Bunn, P. A., Minna, J. D. & Gallo, R. C. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 7415–7419.
2. Reitz, M. S., Poiesz, B. J., Ruscetti, F. W. & Gallo, R. C. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 1887–1891.
3. Poiesz, B. J., Ruscetti, F. W., Reitz, M. S., Kalyanaraman, V. S. & Gallo, R. C. (1981) *Nature (London)* **294,** 268–271.
4. Morgan, D. A., Ruscetti, F. W. & Gallo, R. C. (1976) *Science* **193,** 1007–1008.
5. Mier, J. W. & Gallo, R. C. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 6134–6138.
6. Gootenberg, J. E., Ruscetti, F. W., Mier, J. W., Gazdar, A. F. & Gallo, R. C. (1981) *J. Exp. Med.* **154,** 1403–1418.
7. Poiesz, B. J., Ruscetti, F. W., Mier, J. W., Woods, A. M. & Gallo, R. C. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 6815–6819.
8. Rho, H. M., Poiesz, B. J., Ruscetti, F. W. & Gallo, R. C. (1981) *Virology* **112,** 355–360.
9. Kalyanaraman, V. S., Sarngadharan, M. G., Poiesz, B. J., Ruscetti, F. W. & Gallo, R. C. (1981) *J. Virol.* **38,** 906–915.
10. Kalyanaraman, V. S., Sarngadharan, M. G., Bunn, P. A., Minna, J. D. & Gallo, R. C. (1981) *Nature (London)* **294,** 271–273.
11. Posner, L. E., Robert-Guroff, M., Kalyanaraman, V. S., Poiesz, B. J., Ruscetti, F. W., Bunn, P. A., Minna, J. D. & Gallo, R. C. (1981) *J. Exp. Med.* **154,** 333–346.
12. Oroszlan, S., Copeland, T. D., Henderson, L. E., Stephenson, J. R. & Gilden, R. V. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 2996–3000.
13. Boyer, S. H., Noyes, A. N., Boyer, M. L. & Marr, K. (1973) *J. Biol. Chem.* **248,** 992–1003.
14. Henderson, L. E., Copeland, T. D., Smythers, G. W., Marquardt, H. & Oroszlan, S. (1978) *Virology* **85,** 319–322.
15. Edman, P. & Begg, G. (1967) *Eur. J. Biochem.* **1,** 80–91.
16. Copeland, T. D., Grandgenett, D. P. & Oroszlan, S. (1980) *J. Virol.* **36,** 115–119.
17. Henderson, L. E., Copeland, T. D. & Oroszlan, S. (1980) *Anal. Biochem.* **102,** 1–7.
18. Oroszlan, S., Henderson, L. E., Stephenson, J. R., Copeland, T. D., Long, C. W., Ihle, J. N. & Gilden, R. V. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 1404–1408.
19. Greenwood, F. C., Hunter, W. M. & Glover, J. S. (1963) *Biochem. J.* **89,** 114–123.
20. Laemmli, U. K. (1970) *Nature (London)* **227,** 680–686.
21. Oroszlan, S. & Gilden, R. V. (1980) in *Molecular Biology of RNA Tumor Viruses,* ed. Stephenson, J. R. (Academic, New York), pp. 299–344.
22. Capaldi, R. A. & Vanderkoi, G. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 930–932.
23. Dayhoff, M. O. (1976) *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, DC), Vol. 5, Suppl. 2, pp. 1–8.