Accelerated Publications

Complete Amino Acid Sequence of Streptokinase and Its Homology with Serine Proteases†

Kenneth W. Jackson and Jordan Tang*

ABSTRACT: The complete amino acid sequence of streptokinase has been determined by automated Edman degradation of its cyanogen bromide and proteolytic fragments. The protein consists of 415 amino acid residues. Sequence microheterogeneity was found at two positions. The NH_2-terminal 245 residues of streptokinase are homologous to the sequences of several serine proteases including bovine trypsin and *Streptomyces griseus* proteases A and B. The sequence alignment suggests that the active-site histidine-57 has changed to a glycine in streptokinase. The other active-site residues, aspartyl-102 and serine-195, are, however, present at the expected positions. Streptokinase also contains internal sequence homology between the NH_2-terminal 173 residues and a COOH-terminal 162-residue region between residues 254 and 415. Moderate homology in predicted secondary structures also exists between these two regions. Although streptokinase is not a protease, these observations suggest that it has evolved from a serine protease by gene duplication and fusion. A COOH-terminal region of about 80 residues is apparently deleted from the second half of the duplicated structures. These observations further suggest that the three-dimensional structure of streptokinase likely contains two independently folded domains, each homologous to serine proteases.

Streptokinase is an extracellular protein produced by various strains of streptococci. Its activity was first reported by Tillet & Garner (1933), who discovered that this protein caused the lysis of blood clots. It is now well established that the fibrinolytic activity of streptokinase originates in its ability to activate plasma plasminogen (Milstone, 1941; Castellino, 1979). The mechanism of plasminogen activation by streptokinase is of particular interest for two reasons. First, the increased therapeutic use of streptokinase provokes a better understanding of its mode of action. Second, unlike most of the plasminogen activators, streptokinase is not a protease. It activates plasminogen by the formation of a strong stoichiometric complex with plasminogen (Werkheiser & Markus, 1964; Buck et al., 1968; Schick & Castellino, 1973; Bajaj & Castellino, 1977). The streptokinase–plasminogen complex is an efficient proteolytic activator of plasminogen. The product of plasminogen activation, plasmin, ultimately produces the blood clot lysis. The binding activation of serine protease zymogens by other nonprotease proteins is known for plasminogen by staphylokinase (Lack & Glanville, 1970), for prothrombin by staphylocoagulase (Hamker et al., 1975), and possibly for other zymogens in the clotting (Jackson et al., 1971), fibrinolytic (Saito et al., 1974; Davie & Hanahan, 1977), and complement activation (Reid & Porter, 1981) pathways. The streptokinase–plasminogen interaction, therefore, is a model for the understanding of protein–zymogen binding activation. For these reasons, we have undertaken the determination of the complete amino acid sequence of streptokinase which is reported in this paper.

Materials and Methods

Streptokinase was obtained from A. B. Kabi (Stockholm, Sweden) in vials which also contained serum albumin added as a stabilizer. The streptokinase was separated from the albumin by ammonium sulfate precipitation, followed by dialysis and lyophilization. This starting material gave essentially a single band in $NaDodSO_4$[1]–polyacrylamide gel electrophoresis with a molecular weight near 50000.

Cyanogen bromide cleavages (Steers et al., 1965) were carried out in several batches of about 100 mg of streptokinase each. The fragments were separated on a 2.5 × 100 cm column of Sephadex G-75 superfine, eluted with 1 M acetic acid (Morgan & Henschen, 1969). After their purification, the five CNBr peptides were further fragmented by trypsin (Worthington), clostripain (Boehringer-Mannheim), endoproteinase Lys-C (Boehringer-Mannheim), *Staphylococcus aureus* protease (Miles), or chymotrypsin (Worthington). Some proteolytic digestions were performed on CNBr fragments that were either citraconylated (Atassi & Habeeb, 1972) or phthalylated (Pechère & Bertrand, 1977) in order to increase the solubility and to produce peptides with lysine overlaps.

The peptides were purified by using one or more of the following methods: (a) molecular sieve chromatography on Sephadex G-75, G-25, and G-15 columns eluted with 0.1 M NH_4HCO_3, pH 8; (b) ion-exchange chromatography on

† From the Laboratory of Protein Studies, Oklahoma Medical Research Foundation, and the Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73104. *Received November 1, 1982.* This work was supported by Research Grant HL-17812 from The National Institutes of Health.

[1] Abbreviations: $NaDodSO_4$, sodium dodecyl sulfate; PTH, phenylthiohydantoin; HPLC, high-performance liquid chromatography.

0006-2960/82/0421-6620$01.25/0

1 10 20
NH_2-Ile-Ala-Gly-Pro-Glu-Trp-Leu-Leu-Asp-Arg-Pro-Ser-Val-Asn-Asn-Ser-Gln-Leu-Val-Val-
21 30 40
Ser-Val-Ala-Gly-Thr-Val-Glu-Gly-Thr-Asn-Gln-Asp-Ile-Ser-Leu-Lys-Phe-Phe-Glu-Ile-
41 50 60
Asp-Leu-Thr-Ser-Arg-Pro-Ala-His-Gly-Gly-Lys-Thr-Glu-Gln-Gly-Leu-Ser-Pro-Lys-Ser-
61 70 80
Lys-Pro-Phe-Ala-Thr-Asp-Ser-Gly-Ala-Met-Ser-His-Lys-Leu-Glu-Lys-Ala-Asp-Leu-Leu-
81 90 100
Lys-Ala-Ile-Gln-Glu-Gln-Leu-Ile-Ala-Asn-Val-His-Ser-Asn-Asp-Asp-Tyr-Phe-Glu-Val-
101 110 120
Ile-Asp-Phe-Ala-Ser-Asp-Ala-Thr-Ile-Thr-Asp-Arg-Asn-Gly-Lys-Val-Tyr-Phe-Ala-Asp-
121 130 140
Lys-Asp-Gly-Ser-Val-Thr-Leu-Pro-Thr-Gln-Pro-Val-Gln-Glu-Phe-Leu-Leu-Ser-Gly-His-
141 150 160
Val-Arg-Val-Arg-Pro-Tyr-Lys-Glu-Lys-Pro-Ile-Gln-Asn-Gln-Ala-Lys-Ser-Val-Asp-Val-
161 170 180
Glu-Tyr-Thr-Val-Gln-Phe-Thr-Pro-Leu/Asp-Asn-Pro-Asp-Asp-Asp-Phe-Arg-Pro-Gly-Leu-Lys-
181 190 200
Leu/Asp-Thr-Lys-Leu-Leu-Lys-Thr-Leu-Ala-Ile-Gly-Asp-Thr-Ile-Thr-Ser-Gln-Glu-Leu-Leu-
201 210 220
Ala-Gln-Ala-Gln-Ser-Ile-Leu-Asn-Lys-Asn-His-Pro-Gly-Tyr-Thr-Ile-Tyr-Glu-Arg-Asp-
221 230 240
Ser-Ser-Ile-Val-Thr-His-Asp-Asn-Asp-Ile-Phe-Arg-Thr-Ile-Leu-Pro-Met-Asp-Gln-Glu-
241 250 260
Phe-Thr-Tyr-Arg-Val-Lys-Asn-Arg-Glu-Gln-Ala-Tyr-Arg-Ile-Asn-Lys-Lys-Ser-Gly-Leu-
261 270 280
Asn-Glu-Glu-Ile-Asn-Asn-Thr-Asp-Leu-Ile-Ser-Leu-Glu-Tyr-Lys-Tyr-Val-Leu-Lys-Lys-
281 290 300
Gly-Glu-Lys-Pro-Tyr-Asp-Pro-Phe-Asp-Arg-Ser-His-Leu-Lys-Leu-Phe-Thr-Ile-Lys-Tyr-
301 310 320
Val-Asp-Val-Asp-Thr-Asn-Glu-Leu-Leu-Lys-Ser-Glu-Gln-Leu-Leu-Thr-Ala-Ser-Glu-Arg-
321 330 340
Asn-Leu-Asp-Phe-Arg-Asp-Leu-Tyr-Asp-Pro-Arg-Asp-Lys-Ala-Lys-Leu-Leu-Tyr-Asn-Asn-
341 350 360
Leu-Asp-Ala-Phe-Gly-Ile-Met-Asp-Tyr-Thr-Leu-Thr-Gly-Lys-Val-Glu-Asp-Asn-His-Asp-
361 370 380
Asp-Thr-Asn-Arg-Ile-Ile-Thr-Val-Tyr-Met-Gly-Lys-Arg-Pro-Glu-Gly-Glu-Asn-Ala-Ser-
381 390 400
Tyr-His-Leu-Ala-Tyr-Asp-Lys-Asp-Arg-Tyr-Thr-Glu-Glu-Glu-Arg-Glu-Val-Tyr-Ser-Tyr-
401 410
Leu-Arg-Tyr-Thr-Gly-Thr-Pro-Ile-Pro-Asp-Asn-Pro-Asp-Asp-Lys-COOH

FIGURE 1: Amino acid sequence of streptokinase. Positions 169 and 181 are structurally heterogeneous.

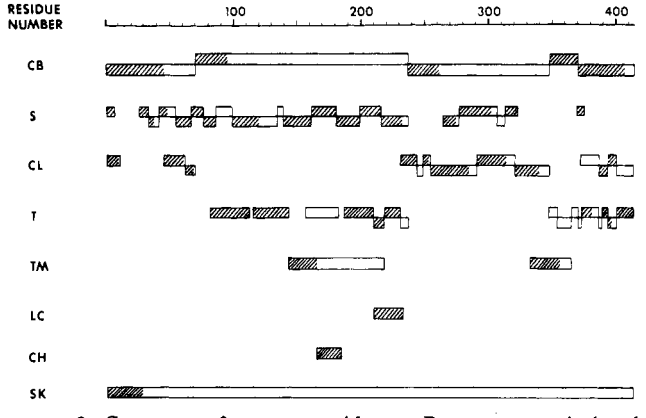

FIGURE 2: Summary of sequence evidence. Bars represent isolated fragments. The shaded areas represent sequenced regions. Peptides were generated by treatment with cyanogen bromide (CB), *S. aureus* protease (S), clostripain (CL), trypsin (T), trypsin digestion of lysine-modified streptokinase (TM), endoproteinase Lys-C (LC), and chymotrypsin (CH). Intact streptokinase is labeled SK.

DEAE-Sephadex, QAE-Sephadex, and SP-Sephadex; (c) HPLC on a reversed-phase Supelcosil LC-18 column (0.45 × 25 cm, from Supelco) by acetonitrile gradient elution from either 0.01 M potassium phosphate, pH 7.5, or 0.1% trifluoroacetic acid; (d) high-voltage paper electrophoresis as described by Chen et al. (1975). Amino acid analyses of the peptides were performed, after acid hydrolysis, on a Durrum D500 amino acid analyzer by using the method of Spackman et al. (1958).

The sequence determinations were carried out by automated Edman degradation in a Beckman sequencer, Model 890C. A dilute Quadrol program (Brauer et al., 1975) was used with 5 mg of polybrene as the carrier. The conversion of anilinothiazolinones to PTH-amino acids was accomplished with anhydrous methanolic HCl (Tarr, 1975) in a Sequemat P-6 autoconverter. The PTH-amino acids were identified by using a Waters Associates gradient HPLC, with a Waters Associates 5-μm C_{18} (8 × 100 mm) radial compression column and module. The solvent system and elution procedure were essentially that described by Henderson et al. (1980). The majority of PTH-amino acid samples were independently identified by thin-layer chromatography (Summers et al., 1973).

Results and Discussion

The complete amino acid sequence of streptokinase is given in Figure 1. The molecule contains 415 amino acid residues in a single polypeptide chain. This sequence was constructed from the peptide sequences from five CNBr fragments and proteolytic cleavages of intact streptokinase and the CNBr fragments, as summarized in Figure 2. The sequence of the NH_2-terminal CNBr fragment, CB1 (residues 1–70), was

Table I: Amino Acid Composition of Streptokinase

amino acid	no. of residues	amino acid	no. of residues
aspartic acid	41	methionine	4
asparagine	24	isoleucine	23
threonine	30	leucine	40
serine	25	tyrosine	22
glutamic acid	28	phenylalanine	15
glutamine	16	histidine	9
proline	21	lysine	32
glycine	20	arginine	20
alanine	21	tryptophan	1
valine	23	total residues	415

constructed from the Edman degradation of this fragment and from the sequence of CB1 peptides produced separately by clostripain and *Staphylococcus aureus* protease digestions. Fragment CB2 (residues 71–237) was the largest of five CNBr fragments. Most of the CB2 structure was obtained from the sequence of peptides produced from *S. aureus* protease and tryptic digests. Other information used in constructing the CB2 sequence was generated from Edman degradation of CB2, sequences of two peptides isolated from the digestion of CB2 separately with α-chymotrypsin and endoproteinase Lys-C, and the sequence of a tryptic peptide obtained from phthalylated streptokinase. The sequence of fragment CB3 (residues 238–347) was determined essentially from Edman degradation of the fragment and from the peptide sequence of its proteolytic digest by using *S. aureus* protease and clostripain. The structure of fragment CB4 (residues 348–370) was obtained by using only Edman degradation of the fragment. Fragment CB5 (residues 371–415) was nearly sequenced through by Edman degradation. A short COOH-terminal sequence was obtained from a tryptic peptide. The placement of CNBr fragments was achieved by the isolation and sequencing of methionine-containing peptides from streptokinase after separate digestions with clostripain, *S. aureus* protease, and

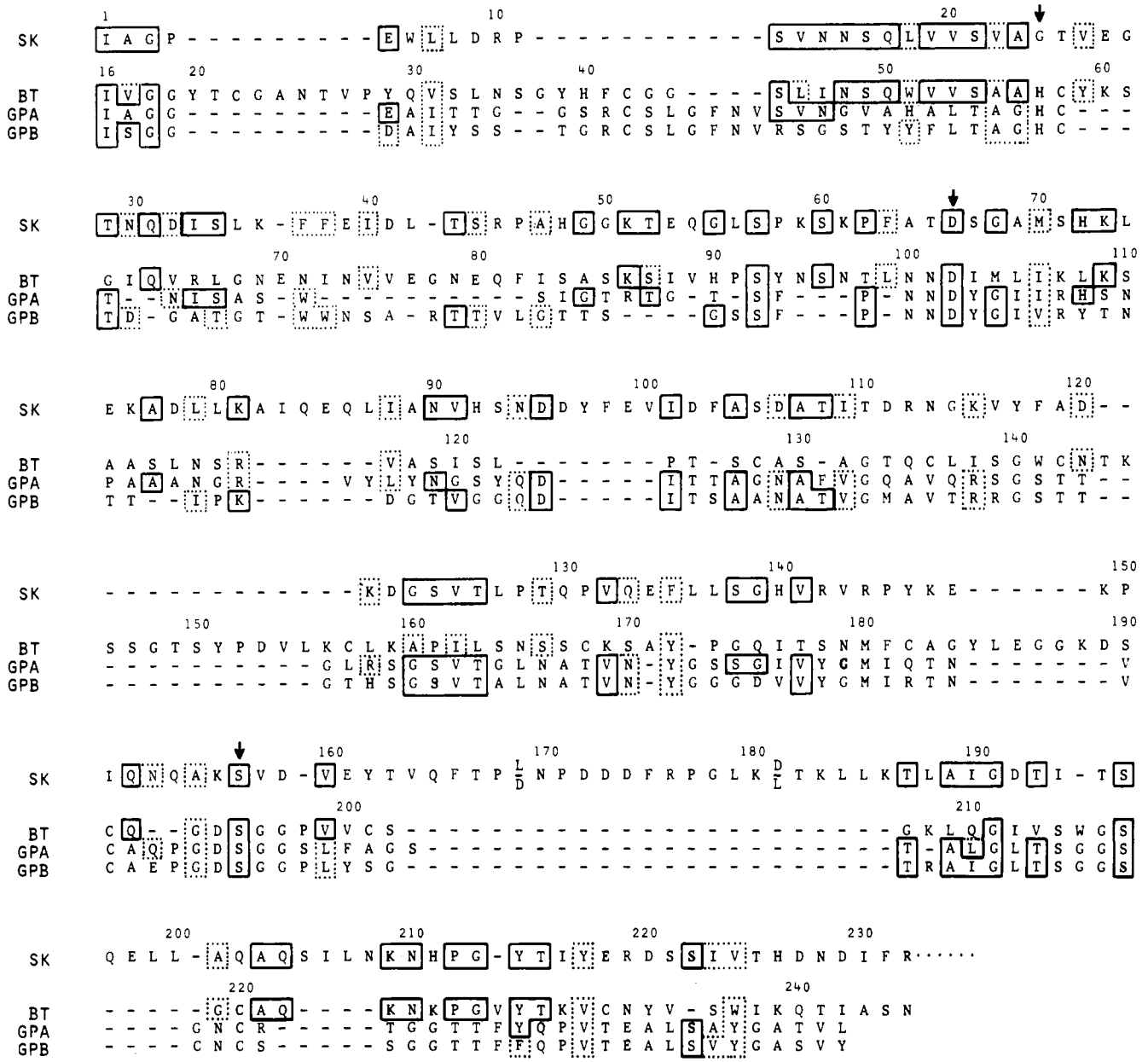

FIGURE 3: Homology in amino acid sequences of the NH_2-terminal region of streptokinase (residues 1–230) to serine proteases: trypsin and *S. griseus* proteases A and B. Boxes represent identical residues. Dotted lines represent structurally and functionally similar residues. The arrows locate positions of the three active-site residues of serine proteases. The alignment among the serine proteases is taken from the work of James et al. (1978).

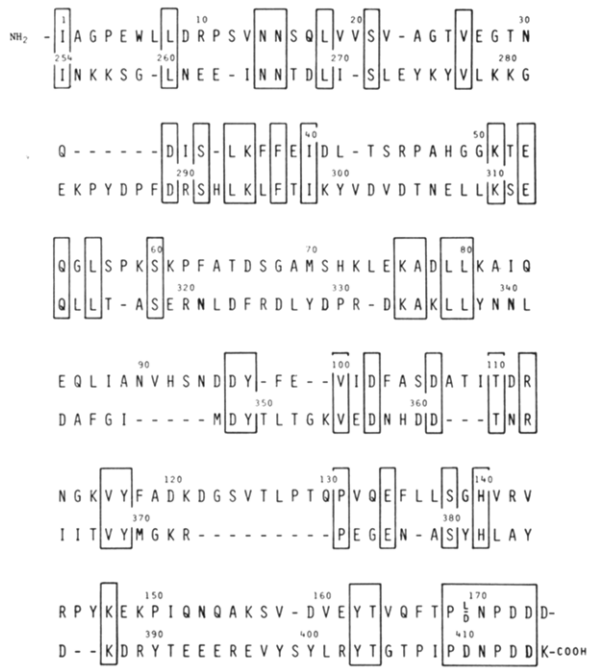

FIGURE 4: Internal structural homology of streptokinase. The observed homologous regions are between the NH_2-terminal 174 residues and the region between residues 254–415 (COOH terminus). They form nearly 2-fold similarity in the streptokinase structure.

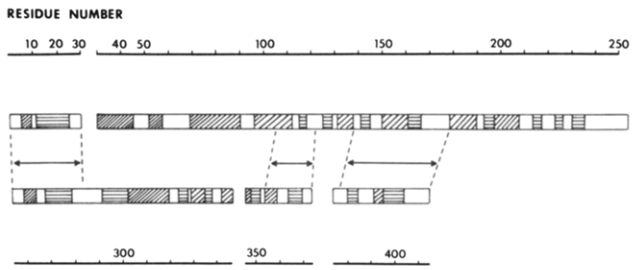

FIGURE 5: Comparison of the predicted secondary structures of the two regions with internal sequence homology from Figure 4. The secondary structure was predicted by using the method of Chou & Fasman (1974). Regions identified by double-headed arrows are those with apparent secondary structure homology. Diagonal hatch marks represent the α helix; horizontal hatch marks designate the β structure; open areas represent the random coil.

trypsin. These results accounted for the identification and placement of all 415 amino acid residues in the streptokinase molecule.

The molecular weight of the protein is 47 408. The amino acid composition of streptokinase, based on the complete sequence, is given in Table I. Streptokinase contains no cysteine or cystine. There are 69 carboxyl side chains, 61 basic side chains, and 170 (41%) hydrophobic residues. The distribution of these residues appears to be rather even throughout the linear structure of streptokinase.

Sequence Homology between Streptokinase and Serine Proteases. We have previously reported that the NH_2-terminal sequence of about 70 residues in streptokinase is homologous to the NH_2-terminal region of serine proteases such as trypsin and *Streptomyces griseus* proteases A and B (Jackson & Tang, 1978). With the complete sequence of streptokinase, it is now clear that serine proteases of the trypsin family are homologous with approximately the NH_2-terminal 230 residues of streptokinase. Figure 3 illustrates the alignment of the streptokinase sequence with that of bovine trypsin (Hartley, 1970) and *S. griseus* proteases A (Johnson & Smillie, 1974) and B (Jurašek et al., 1974). Since streptokinase is not a protease itself, the residues corresponding to the known serine protease active sites are of particular interest. As previously described (Jackson & Tang, 1978), the residue corresponding to the active-site His-57 is a glycine (residue 24) in streptokinase. The sequence that precedes this residue in streptokinase is highly homologous to that of the proteases (Figure 3); the His → Gly replacement is therefore quite certain. The alignment in Figure 3 also shows that the active-site residues corresponding to Asp-102 and Ser-195 are present in streptokinase (Asp-66 and Ser-157). It should be pointed out that in the alignment in Figure 3, major gaps in the sequence were taken from the published homology alignment of bovine trypsin and *S. griseus* proteases based on topological equivalences in the crystal structures (James et al., 1978). Streptokinase shares two major gaps with the *S. griseus* proteases. One gap is located between residues 19 and 29 of trypsin and the other gap is located between residues 143 and 156 of trypsin. The three-dimensional structure of the NH_2-terminal 230 residues of streptokinase is probably most similar to that of the *S. griseus* proteases. There is a good sequence homology between streptokinase and all three proteases; however, streptokinase shares more common gaps with the *S. griseus* proteases. Two major sequence gaps were introduced for all three serine proteases in the alignment of Figure 3. These gaps correspond to streptokinase regions of residues 96–100 and residues 165–186. These probably represent structures unique for streptokinase.

Internal Twofold Homology of Streptokinase and Its Evolutionary Origin. The streptokinase sequence contains internal homology between two large regions as shown in Figure 4. The 174-residue region at the NH_2 terminus is homologous with the region between residues 254 and 415 at the COOH terminus. Moderate homology can also be seen between the predicted secondary structures of these regions (Figure 5). This internal structural homology, which is almost between the two halves of the molecule, suggests that streptokinase contains two similar domains in its three-dimensional structure. This is supported by our previous finding that staphylokinase, another nonprotease plasminogen activator, has a subunit molecular weight of 23 000 (only about half of that for streptokinase, 47 408) and requires a dimer to be functionally active (Jackson et al., 1980, 1981). These observations predict close relationships in the structure and functional mechanisms of streptokinase and staphylokinase.

The amino acid sequences of these two regions of the streptokinase molecule appear to be reasonably homologous (Figure 5). There are 42 identical residues which account for about one-fourth of the region. Together with 18 functionally similar residues, they represent about 35% of the residues in these regions. On the other hand, the COOH-terminal region, residues 254–415, is only weakly homologous to the serine proteases (sequence comparisons not shown). In addition, the residues in this region that correspond to the active-site His-57 and Ser-195 of serine proteases are residues Lys-275 and Val-397, respectively. There is a possibility that Asp-323 may correspond to the active-site Asp-102 of serine proteases, but this cannot be certain due to the lack of similarity in the two sequences around these residues.

Evidence discussed above suggests that streptokinase was likely derived from a serine protease in evolution, as depicted in Figure 6. The original structural gene for the ancestral protease probably corresponded to the NH_2-terminal 230 residues of streptokinase. Mutational steps may have produced a nonprotease plasminogen activator that contained two subunits, as in the case of staphylokinase. The duplication and fusion of this gene could produce an ancestral streptokinase

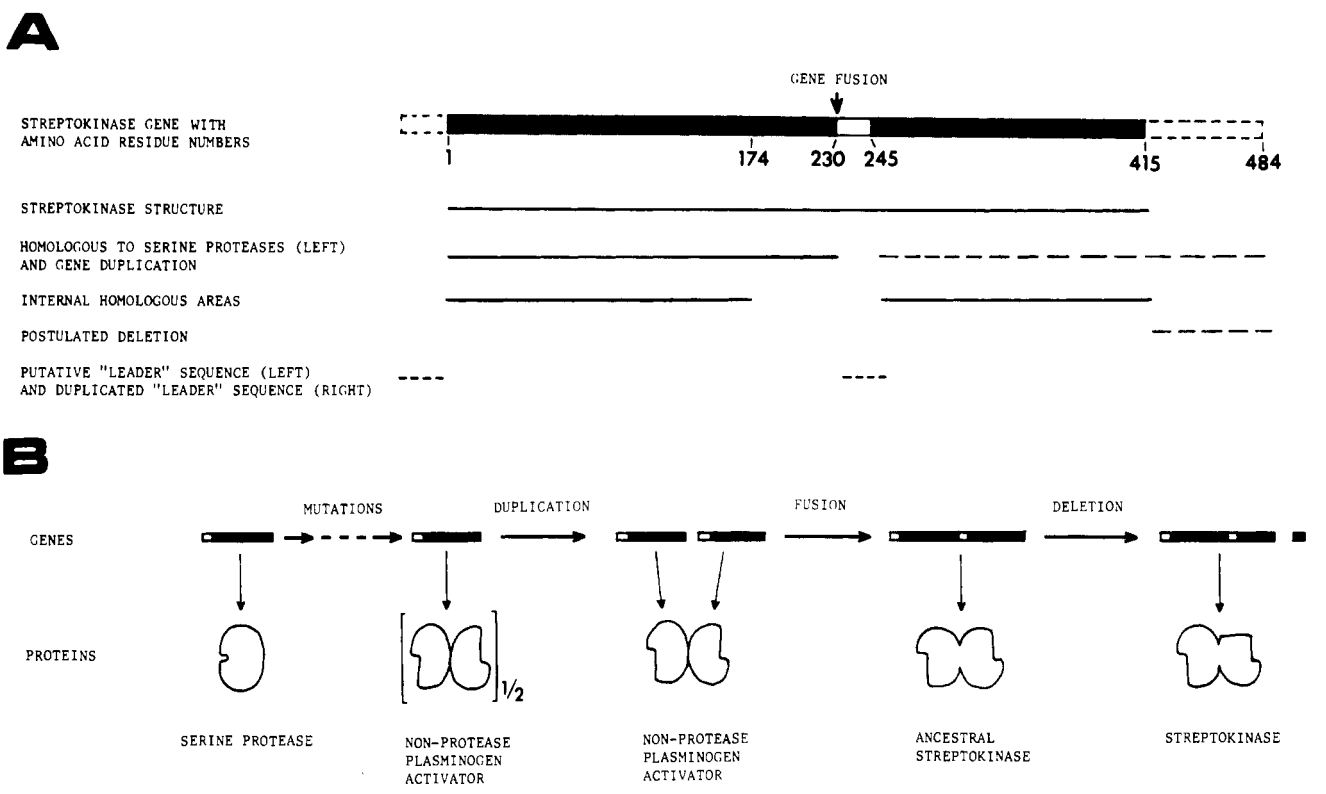

FIGURE 6: Postulated major steps in the evolution of the streptokinase gene and protein. Part A illustrates the homologous areas (solid lines) and postulated events (dashed lines) involving the streptokinase gene. Part B is a schematic presentation of the evolution of the streptokinase gene and protein. In both parts, the putative and duplicated "leader" sequence regions are blank areas.

near 484 amino acid residues. Eventual deletions, from the COOH-terminal half including about 55 residues from the COOH terminus and about 15 interior residues, would have produced a molecule similar to the current streptokinase. It may be noteworthy that the homology of the NH_2-terminal region of the streptokinase with serine proteases ends at about residue 230 (Figure 3) while the internal homology of the second half starts at residue 254 (Figure 4). This region of 24 residues (numbers 231–253) cannot be aligned in homology with the serine proteases nor with any other sequence within the streptokinase molecule. It is thus postulated that this region may have evolved from the leader sequence gene for the ancestral serine protease (Figure 6). Since many microbial serine proteases as well as streptokinse are secretory proteins, this explanation appears logical.

Possible Functional Implication. The extent of sequence homology between the NH_2-terminal 230 residues of streptokinase with serine proteases appears sufficient to anticipate a closer homology for the tertiary structures. The presence of the active-site serine and aspartic acid in the NH_2-terminal domain of streptokinase prompts an interesting question of whether serine-157 could have a functional role in the streptokinase activation of plasminogen. Evidence so far, however, does not support the hypothesis that streptokinase is capable of catalyzing proteolysis [for a review, see Castellino (1979)]. Several hypotheses have been put forward to explain the mechanism of binding activation of streptokinase (Jackson & Tang, 1978; Bode & Huber, 1976). Our current findings do not discriminate among these hypotheses. Not adequately accounted for in any of the current hypotheses is the apparent requirement for dimeric structures in the function of streptokinase and staphylokinase. The clarification in this area would be greatly assisted by the detailed knowledge of tertiary structures in the streptokinase–plasminogen interaction.

Acknowledgments

We thank Azar Fesmire for technical assistance in this work.

References

Atassi, M. Z., & Habeeb, A. F. S. A. (1972) *Methods Enzymol. 25*, 546–553.
Bajaj, S. P., & Castellino, F. J. (1977) *J. Biol. Chem. 252*, 492–498.
Bode, W., & Huber, R. (1976) *FEBS Lett. 68*, 231–236.
Brauer, A. W., Margolies, M. N., & Haber, E. (1975) *Biochemistry 14*, 3029–3035.
Buck, F. F., Hummel, B. C. W., & DeRenzo, E. C. (1968) *J. Biol. Chem. 243*, 3648–3654.
Castellino, F. J. (1979) *Trends Biochem. Sci. (Pers. Ed.) 4*, 1–5.
Chen, K. C. S., Tao, N., & Tang, J. (1975) *J. Biol. Chem. 275*, 5068–5075.
Chou, P. Y., & Fasman, G. D. (1974) *Biochemistry 13*, 222–245.
Davie, E. W., & Hanahan, D. J. (1977) in *The Plasma Proteins* (Putnam, F. W., Ed.) Vol. 3, pp 421–544, Academic Press, New York.
Hamker, H. C., Bas, B. M., & Muller, A. D. (1975) *Biochim. Biophys. Acta 379*, 180–188.
Hartley, B. S. (1970) *Philos. Trans. R. Soc. London, Ser. B 257*, 77–89.
Henderson, L. E., Copeland, T. D., & Oroszlan, S. (1980) *Anal. Biochem. 102*, 1–7.
Jackson, C. M., Gordon, J. G., & Hanahan, D. J. (1971) *Biochim. Biophys. Acta 252*, 255–261.
Jackson, K. W., & Tang, J. (1978) *Thromb. Res. 13*, 693–699.
Jackson, K. W., Esmon, N., & Tang, J. (1980) in *The Reg-*

ulation of Coagulation (Mann, K. G., & Taylor, F. B., Eds.) pp 515–520, Elsevier, New York.
Jackson, K. W., Esmon, N., & Tang, J. (1981) *Prog. Chem. Fibrinolysis Thrombolysis 5*, 49–52.
James, M. N. G., Delbaere, L. T. J., & Brayer, G. D. (1978) *Can. J. Biochem. 56*, 396–402.
Johnson, P., & Smillie, L. B. (1974) *FEBS Lett. 47*, 1–6.
Jurašek, L., Carpenter, M. R., Smillie, L. B., Gertter, A., Levy, S., & Ericsson, L. H. (1974) *Biochem. Biophys. Res. Commun. 61*, 1095–1100.
Lack, C. H., & Glanville, K. L. A. (1970) *Methods Enzymol. 19*, 706–714.
Milstone, H. (1941) *J. Immunol. 42*, 109–116.
Morgan, F. J., & Henschen, A. (1969) *Biochim. Biophys. Acta 181*, 93–104.
Pechère, J.-F., & Bertrand, R. (1977) *Methods Enzymol. 47*, 149–155.
Reid, K. B. M., & Porter, R. R. (1981) *Annu. Rev. Biochem. 50*, 433–464.
Saito, H., Ratnoff, O. D., & Donaldson, V. H. (1974) *Circ. Res. 34*, 641–651.
Schick, R. A., & Castellino, F. J. (1973) *Biochemistry 12*, 4315–4320.
Spackman, D. H., Stein, W. H., & Moore, S. (1958) *Anal. Chem. 30*, 1190–1206.
Steers, E., Craven, Z. R., Anfinson, C. B., & Bethune, J. L. (1965) *J. Biol. Chem. 240*, 2478–2484.
Summers, M. R., Smythers, G. W., & Oroszlan, S. (1973) *Anal. Biochem. 53*, 624–628.
Tarr, G. E. (1975) *Anal. Biochem. 63*, 361–370.
Tillet, W. S., & Garner, R. L. (1933) *J. Exp. Med. 58*, 485–502.
Werkheiser, W. C., & Markus, G. (1964) *J. Biol. Chem. 239*, 2644–2650.

A Homozygous Human Cell Line Contains Three Subsets of HLA-DR-like Antigens Distinguishable by Amino Acid Sequencing†

David W. Andrews, M. Rosa Bono, and Jack L. Strominger*

ABSTRACT: The major histocompatibility complex (HLA in humans) contains a relatively large set of gene loci in the HLA-D/DR region responsible for regulation of the immune response. The structural dissection of the protein products of these loci is a necessary accompaniment to understanding of this response. In this study, two subsets of HLA-DR-like molecules have been separated by using monoclonal antibodies, and their component α and β chains have been subjected to amino acid terminal sequencing. The results from this sequencing experiment show three differences in the first 14 residues of the β chains and no differences in the first 15 residues of the α chains. These data along with previous sequencing of the DC-1 antigen [Bono, M. R., & Strominger, J. L. (1982) *Nature (London) 299*, 836–838] demonstrate that three distinct subsets of HLA-DR-like antigens are expressed by a homozygous human lymphoblastoid cell line.

Initiation, intensity, and specificity of the immune response to extrinsic and intrinsic antigens are controlled by the products of genes in the major histocompatibility complex (MHC)[1] (Strominger et al., 1981; Benacerraf, 1981). This large set of genes was originally defined in mice by the rejection of grafts or the proliferation of lymphocytes between individuals. The loci in the MHC were divided into three classes on the basis of either structure or function of their products, or a combination of both (Klein et al., 1981). The class I genes encode a polypeptide chain of 44 000 daltons, which forms a heterodimer with β_2-microglobulin (12 000 daltons). This functional unit is distributed on virtually all nucleated cells in vertebrates. These genes are the H-2K, D, and L loci in the mouse and the HLA-A, B, and C loci in humans. The class II genes encode polypeptide chains of two different sizes (α and β chains). The regions that contain these loci are called HLA-D/DR in humans and Ia in mice. The glycoproteins encoded by the these loci have a restricted cellular distribution, indicative of their role in lymphocyte interactions. The apparent molecular weight of these chains varies from species to species and even within species. In humans, the HLA-DR glycoproteins are 34 000 (α chains) and 29 000 (β chains) daltons, but nevertheless they are remarkably similar to the class I antigens in structure. Class III genes encode complement components, whose functional and structural connection with class I and II gene products is problematic, although they also function in the immune system.

Serological studies have revealed 10 alleles in the HLA-D region, defined by cellular reactivity, and 12 in the HLA-DR region, defined by alloantisera (Terasaki, 1980). However, subdivision of the HLA-D/DR region has not been accomplished to the extent that it has in the murine system, where I-A, I-J, and I-E subregions are relatively well-defined and subregions I-B and I-C have been suggested. Only two products of the Ia region (encoded by four genes in the I-A and I-E subregions) have been identified, each with polypeptides of about 28 000 and 33 000 daltons. Amino-terminal sequence data have revealed homology between DR and I-E polypeptides (McMillan et al., 1977; Allison et al., 1978) and between the closely related specificity DC-1 and I-A polypeptides [Bono & Strominger, 1982; see also Goyert et al. (1982)]. Herein we report the amino-terminal sequences of two antigenically distinguishable subsets of HLA-DR antigens from an HLA-DRw6 homozygous cell line separated by using monoclonal antibodies. The β chains of these molecules show three differences in the first 14 residues, while their α chains are identical for the first 15 residues. Thus, there are at least three sets of HLA-DR-like antigens expressed in a homozygous

† From the Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138. *Received October 15, 1982.* Supported by National Institutes of Health Research Grant AI-10736.

[1] Abbreviations: MHC, major histocompatibility complex; NaDodSO$_4$, sodium dodecyl sulfate; DOC, deoxycholate; DTT, dithiothreitol; Tris, tris(hydroxymethyl)aminomethane; PMSF, phenylmethanesulfonyl fluoride.

0006-2960/82/0421-6625$01.25/0