



Office de la Propriété

Intellectuelle
du Canada

Un organisme
d'Industrie Canada

Canadian
Intellectual Property
Office

An agency of
Industry Canada

CA 2087342 C 2003/12/23

(11)(21) 2 087 342

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 1991/08/06
(87) Date publication PCT/PCT Publication Date: 1992/02/20
(45) Date de délivrance/Issue Date: 2003/12/23
(85) Entrée phase nationale/National Entry: 1993/01/14
(86) N° demande PCT/PCT Application No.: US 1991/005480
(87) N° publication PCT/PCT Publication No.: 1992/002615
(30) Priorité/Priority: 1990/08/09 (566,026) US

(51) Cl.Int.⁵/Int.Cl.⁵ C12N 15/57, C12N 9/54, C12P 21/00, C07K 2/00
(72) Inventeurs/Inventors:
ABRAHMSEN, LARS, SE;
BURNIER, JOHN, US;
WELLS, JAMES A., US
(73) Propriétaire/Owner:
GENENTECH, INC., US
(74) Agent: SMART & BIGGAR

(54) Titre : VARIANTES DE LA PROTEASE DE LA SERINE A ACTIVITE DE LIGASE PEPTIDIQUE
(54) Title: SERINE PROTEASE VARIANTS HAVING PEPTIDE LIGASE ACTIVITY

(57) Abrégé/Abstract:

The invention relates to serine protease variants derived from precursor serine proteases via recombinant and/or chemical methods to form protease variants having improved peptide ligase activity. The invention also includes novel ligation substrates which in combination with the serine protease variants and a second ligation substrate are capable of forming a ligation product. The invention also relates to methods for forming such ligation products and the products formed thereby.

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 9/56, 9/54, 15/00		2087342	A1	(11) International Publication Number: WO 92/02615 (43) International Publication Date: 20 February 1992 (20.02.92)
(21) International Application Number: PCT/US91/05480				(72) Inventors; and
(22) International Filing Date: 6 August 1991 (06.08.91)				(75) Inventors/Applicants (for US only) : ABRAHMSEN, Lars [SE/SE]; Sjalagardsg 10, S-111 31 Stockholm (SE). BURNIER, John [US/US]; 202 Sterling Avenue, Pacifica, CA 94044 (US). WELLS, James, A. [US/US]; 1341 Columbus Avenue, Burlingame, CA 94010 (US).
(30) Priority data: 566,026 9 August 1990 (09.08.90) US				(74) Agents: TRECARTIN, Richard, F. et al.; Flehr, Hohbach, Test, Albritton & Herbert, Four Embarcadero Center, Suite 3400, San Francisco, CA 94111-4187 (US).
(60) Parent Application or Grant (63) Related by Continuation US Filed on 566,026 (CIP) 9 August 1990 (09.08.90)				(81) Designated States: AT (European patent), BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent), US.
(71) Applicant (for all designated States except US): GENENTECH, INC. [US/US]; 460 Point San Bruno Boulevard, South San Francisco, CA 94080 (US).				Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>

(54) Title: SERINE PROTEASE VARIANTS HAVING PEPTIDE LIGASE ACTIVITY**(57) Abstract**

The invention relates to serine protease variants derived from precursor serine proteases via recombinant and/or chemical methods to form protease variants having improved peptide ligase activity. The invention also includes novel ligation substrates which in combination with the serine protease variants and a second ligation substrate are capable of forming a ligation product. The invention also relates to methods for forming such ligation products and the products formed thereby.

61051-2487

1

SERINE PROTEASE VARIANTS HAVING
PEPTIDE LIGASE ACTIVITY

5 TECHNICAL FIELD OF THE INVENTION

The invention relates to serine protease variants derived from precursor serine proteases via recombinant and/or chemical methods to form protease variants having improved peptide ligase activity. The invention also 10 includes novel ligation substrates which in combination with the serine protease variants and a second ligation substrate are capable of forming a ligation product. The invention also relates to methods for forming such ligation products and the products formed thereby.

15 BACKGROUND OF THE INVENTION

Chemical approaches for synthesis and engineering of proteins offer many advantages to recombinant methods in that one can incorporate non-natural or selectively labelled amino acids. However, peptide synthesis is 20 practically limited to small proteins (typically <50 residues) due to the accumulation of side-products and racemization that complicate product purification and decrease yields (for recent reviews see Kaiser, E.T. (1989) Acc. Chem. Res. 22, 47-54; Offord, R.E. (1987) 25 Prot. Eng. 1, 151-157).

2087312

-2-

Proteolytic enzymes, in particular serine proteases, have reportedly been used as alternatives to synthetic peptide chemistry because of their stereoselective properties and mild reaction conditions (for reviews see Kullman, 5 W. (1987) In: Enzymatic Peptide Synthesis, CRC Press, Florida U.S.; Chaiken, (1981) CRC Crit. Rev. Biochem. 11, 255-301). Such enzymes reportedly have been used to complement chemical coupling methods to produce larger peptides by blockwise enzymatic coupling of synthetic 10 fragments. Inouye et al. (1979), J. Am. Chem. Soc., 101, 751-752 (insulin fragments); Hommandberg and Laskowski, (1979) Biochemistry 18, 586-592 (ribonuclease fragments)). However, the narrow substrate specificities and intrinsic hydrolytic (peptidase) activity of serine 15 proteases have limited their use in peptide synthesis.

A central problem in the case of serine proteases in peptide synthesis is that hydrolysis of the acyl-enzyme intermediate is strongly favored over aminolysis (Fig. 1). Several laboratories have reported that the 20 equilibrium is shifted from hydrolysis toward aminolysis by use of mixed or pure organic solvents to carry out catalysis (Coletti-Previero et al., (1969) J. Mol. Biol. 39, 493-501; Barbas et al., (1988) J. Am. Chem. Soc. 110, 5162-5166). However, enzymes are generally less stable 25 and relatively insoluble in organic solvents (Wong et al., (1990) J. Am. Chem. Soc. 112, 945-953; Klibanov, (1986) Chemtech 16, 354-359). Further, kinetic activation barriers in organic solvents are higher for the charged transition-states involved leading to lower 30 enzymatic activity. In an attempt to avoid these problems, one laboratory reported that thiolsubtilisin, a derivative of the bacterial serine protease in which the active site Ser221 was chemically converted to a Cys (S221C), shifted the preference for aminolysis to 35 hydrolysis by >1000-fold for very small peptides. Nakasuta et al. (1987) J. Am. Chem. Soc. 109, 3808-3810.

-3- 2087342

This shift was attributed to the kinetic preference of thioesters to react with amines over water. Based upon similar principles, another laboratory reported that selenolsubtilisin had a 14,000-fold shift in preference 5 for aminolysis over hydrolysis. Wu and Hilvert (1989) J. Am. Chem. Soc. 111, 4513-4514. However the catalytic efficiencies for aminolysis of a chemically activated ester by either thiol- or selenolsubtilisin are about 10^3 - and 10^4 -fold, respectively, below the esterase 10 activity of wild-type subtilisin. Although chemically active esters have reportedly been used to increase the rates for acylation of thiol- or selenolsubtilisin (e.g. the acylation of thiolsubtilisin with a p-chlorophenyl ester of an 8-mer peptide for ligation with a 4-mer 15 peptide in >50% DMF), such activated esters present synthetic difficulties as well as creating substrates prone to spontaneous hydrolysis in aqueous solvents 20 (Nakatsuka et al. (1987) supra.).

The serine proteases comprise a diverse class of enzymes 25 having a wide range of specificities and biological functions. Stroud, R.M. (1974) Sci Amer. 131, 74-88. Despite their functional diversity, the catalytic machinery of serine proteases has been approached by at least two genetically distinct families of enzymes: the 30 Bacillus subtilisin-type serine proteases and the mammalian and homologous bacterial trypsin-type serine proteases (e.g., trypsin and S. gresius trypsin). These two families of serine proteases show remarkably similar mechanisms of catalysis. Kraut, J. (1977) Ann. Rev. Biochem., 46, 331-358. Furthermore, although the primary structure is unrelated, the tertiary structure of these two enzyme families bring together a conserved catalytic triad of amino acids consisting of serine, histidine and aspartate.

Subtilisin is a serine endoprotease (MW~27,500) which is secreted in large amounts from a wide variety of Bacillus species. The protein sequence of subtilisin has been determined from at least four different species 5 of Bacillus. Markland, F.S., et al. (1971) in The Enzymes, ed. Boyer, P.D., Acad. Press, New York, Vol. III, pp. 561-608; Nedkov, P. et al. (1983) Hoppe-Seyler's Z. Physiol. Chem., 364, 1537-1540. The three-dimensional crystallographic structure of subtilisin BPN' (from B. amyloliquefaciens) to 2.5Å resolution has also been reported. Bott, et al. (1988), J. Biol. Chem., 263, 10 7895-7906; McPhalen, et al. (1988), Biochemistry, 27, 6582-6598; Wright, C.S., et al. (1969), Nature, 221, 235-242; Drenth, J. et al. (1972) Eur. J. Biochem., 26, 177-15 181. These studies indicate that although subtilisin is genetically unrelated to the mammalian serine proteases, it has a similar active site structure. The x-ray crystal structures of subtilisin containing covalently bound peptide inhibitors (Robertus, J.D., et 20 al. (1972), Biochemistry, 11, 2439-2449), product complexes (Robertus, J.D., et al. (1972) Biochemistry 11, 4293-4303), and transition state analogs (Matthews, D.A., et al. (1975) J. Biol. Chem. 250, 7120-7126; Poulos, T.L., et al. (1976) J. Biol. Chem. 251, 1097-1103), which 25 have been reported have also provided information regarding the active site and putative substrate binding cleft of subtilisin. In addition, a large number of kinetic and chemical modification studies have been reported for subtilisin (Philipp, M., et al. (1983) Mol. Cell. Biochem. 51, 5-32; Svendsen, I.B. (1976) Carlsberg Res. Comm. 41, 237-291; Markland, F.S. Id.). Stauffer, D.C., et al. (1965) J. Biol. Chem. 244, 5333-5338; Polgar, L. et al. (1981) Biochem. Biophys. Acta 667, 351-354). 30 35 U.S. Patent 4,760,025 discloses subtilisin mutants wherein a different amino acid is substituted for the

-5-2087342

naturally-occurring amino acid residues of Bacillus amyloliquifaciens subtilisin at positions +32, +155, +104, +222, +166, +64, +33, +169, +189, +217, or +156.

The references discussed above are provided solely for
5 their disclosure prior to the filing date of the present
case, and nothing herein is to be construed as an
admission that the inventors are not entitled to antedate
such disclosure by virtue of prior invention or priority
based on earlier filed applications.

10 SUMMARY OF THE INVENTION

Based on the foregoing, it is apparent that the chemical
synthesis of large peptides and proteins is severely
limited by the available chemical synthesis techniques
and the lack of an efficient peptide ligase which is
15 capable of coupling block synthetic or recombinant
peptides.

Accordingly, it is an object herein to provide serine
protease variants which are capable of efficiently
ligating peptides and other substrates.

20 Further, it is an object herein to provide ligation
substrates which when used in combination with the
aforementioned serine protease variants are capable of
ligating a first ligation substrate with a second
ligation substrate to form a ligation product.

25 Further, it is an object herein to provide methods for
producing ligation products using serine protease
variants and ligation substrates.

Further, it is an object herein to provide ligation
products made by the aforementioned methods.

2087342

In accordance with the foregoing objects, the invention includes serine protease variants having an amino acid sequence not found in nature which are derived from a precursor serine protease by at least two changes in the 5 amino acid residues of the precursor serine protease. In particular, the precursor serine proteases are characterized by a catalytic serine residue which participates in the catalysis normally carried out by the precursor enzyme. This active site serine residue 10 is replaced with a different amino acid to substitute the nucleophilic oxygen of the serine side chain with a different nucleophile. Alternatively, the side chain of the active site serine may be directly modified chemically to substitute a different nucleophile for the 15 nucleophilic oxygen. The second change comprises replacement or modification of a second amino acid residue not consisting of the active site serine of the precursor enzyme. Such replacement or modification of the second amino acid residue, in combination with the 20 replacement or modification of the active site serine, produces a serine protease variant characterized by a peptide ligase activity measured in aqueous solution which is greater than that of a different serine protease variant containing only the substitution or modification 25 of the nucleophilic oxygen at the active site serine.

The invention also includes ligation substrates which are useful in combination with the aforementioned serine protease variants. In various precursor serine proteases, the enzyme catalyzes the cleavage of a peptide bond (the scissile peptide bond). The standard designation of protease hydrolysis substrate residues using the nomenclature of Schechter and Berger (1967) Biochem. Biophys. Res. Commun. 27, 157-162 and the scissile bond hydrolyzed is shown in Fig. 2A. Since 30 peptide ligation is essentially the reversal of hydrolysis, ligation substrates are defined similarly 35

as depicted in Figure 2B. Thus, in one aspect of the invention, first ligation substrates comprise at least an R1 amino acid residue with the carboxy terminus of the R1 residue esterified with an organic alcohol (e.g. 5 a 2-hydroxy carboxylic acid) or thiol. The R1 residue comprises those amino acid residues R1 which preferentially bind to the precursor serine hydrolase or which preferentially bind to those serine hydrolase variants which have been further modified to alter 10 substrate specificity at the P1 position. Such R1 residues also comprise non-naturally occurring amino acids for which the variant has specificity. In addition, the esterified 2-hydroxy carboxylic acid closely resembles the P1' residue in substrates for the 15 precursor serine protease or the residues preferred by those serine protease variants whose specificity has been modified at the P1' position.

The invention also includes ligation methods wherein the serine protease variant of the invention is contacted 20 with a first and a second ligation substrate to form a ligation product. The first ligation substrate comprises the aforementioned ligation substrate (Fig. 2B). The second ligation substrate comprises at least an R1' amino acid residue for which the serine protease variant has 25 specificity (Fig. 2C). It may also comprise non-naturally occurring amino acids for which the variant has specificity. The ligation product so formed by a ligation of the first and second ligation substrate contains the sequence R1-R1' (Fig. 2D).

30 In addition, the invention includes ligation products made by the aforementioned method. Such products are typically have a length greater than the length of about 17 amino acid residues. Such ligation products are also characterized by the ligation method which may be carried 35 out in, but not limited to, aqueous solution.

61051-2487

-7a-

In one aspect, the invention provides a subtilisin-type serine protease variant having an amino acid sequence not found in nature and which is derived from a precursor subtilisin-type serine protease having an α -helix containing proline at a residue equivalent to proline 225 in *Bacillus amyloliquefaciens* subtilisin and a catalytic serine equivalent to serine 221 in *Bacillus amyloliquefaciens* subtilisin, said derivation comprising: a) replacing said catalytic serine with a first amino acid having a different nucleophilic side chain or modifying the side chain of said catalytic serine, to substitute the nucleophilic oxygen of said side chain with a nucleophilic sulfur, thereby converting said catalytic serine to cysteine, and b) replacing said proline with a second different amino acid comprising a helix-forming amino acid or an amino acid with a side chain volume less than the side chain volume of proline; said subtilisin-type serine protease variant being characterized by having peptide ligase activity in aqueous solution which is greater than that of said precursor subtilisin-type serine protease variant containing only said substitution or modification of said nucleophilic oxygen.

In another aspect, the invention provides a ligation method comprising the step of contacting the subtilisin-type serine protease variant as described above with a first and a second ligation substrate, said first ligation substrate comprising at least a first amino acid residue (R1 residue), the carboxy-terminus of said R1 residue being esterified with an organic alcohols or thiols wherein the organic moiety is selected from the group consisting of C₆-C₁₂ aryl where the aryl group is unsubstituted or substituted by one or more of the groups nitro, hydroxy, halo (F, Cl, Br, I), C₁-C₈ alkyl, halo-C₁-C₈ alkyl, C₁-C₈-alkoxy, amino, phenoxy, phenyl, acetamido,

61051-2487

- 7b -

benzamido, di-C₁-C₈ alkylamino, C₁-C₈ alkylamino, C₆-C₁₂ aroyl, C₁-C₈ alkanoyl, and hydroxy-C₁-C₈ alkyl, C₁-C₁₂ alkyl either substituted or unsubstituted, branched, straight chain or cyclo where the substituents are selected from halo (F, Cl, 5 Br, I), C₁-C₈ alkoxy, C₆-C₁₂ aryloxy where the aryl group is unsubstituted or substituted by one or more of the groups nitro, hydroxy, halo (F, Cl, Br, I), C₁-C₈ alkyl, C₁-C₈-alkoxy, amino, phenoxy, acetamido, benzamido, di-C₁-C₈ alkylamino, C₁-C₈ alkylamino, C₆-C₁₂ aroyl, and C₁-C₈ alkanoyl, 10 isothioureido, C₃-C₇ cycloalkyl, ureido, amino, C₁-C₈ alkylamino, di-C₁-C₈ alkylamino, hydroxy, amino-C₂-C₈ alkylthio, amino-C₂-C₈ alkoxy, acetamido, benzamido wherein the phenyl ring is unsubstituted or substituted by one or more of the groups nitro, hydroxy, halo (F, Cl, Br, I), C₁-C₈ 15 alkyl, C₁-C₈-alkoxy, amino, phenoxy, acetamido, benzamido, di-C₁-C₈ alkylamino, C₁-C₈ alkylamino, C₆-C₁₂ aroyl, C₁-C₈ alkanoyl, C₆-C₁₂ arylamino wherein the aryl group is unsubstituted or substituted by one or more of the groups nitro, hydroxy, halo, C₁-C₈ alkyl, C₁-C₈-alkoxy, amino, 20 phenoxy, acetamido, benzamido, di-C₁-C₈ alkylamino, C₁-C₈ alkylamino, C₆-C₁₂ aroyl, and C₁-C₈ alkanoyl, guanidino, phthalimido, mercapto, C₁-C₈ alkylthio, C₆-C₁₂ arylthio, carboxy carboxamide, carbo-C₁-C₈ alkoxy, and aromatic heterocycle wherein the heterocyclic groups have 5-10 ring 25 atoms and contain up to two O, N, or S heteroatoms, and 2-hydroxy carboxylic acid and wherein said second ligation substrate comprises at least a second amino acid residue (R1' residue), wherein the ligation product formed by said ligation contains the sequence R1-R1'.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 depicts the kinetic aspects of peptide ligation. In Equation 1, the ester bond is cleaved by the enzyme, the alcohol group leaves and the acid becomes acylated to the active site nucleophile to form an acyl-enzyme intermediate (acyl-enzyme). Two possible reactions of the acyl-intermediate are hydrolysis (Equation 2) and aminolysis (Equation 3). Hydrolysis is nonproductive whereas aminolysis leads to adduct formation. The amidase activity of the enzyme (Equation 4) results in cleavage of the adduct and reformation of the acyl-enzyme intermediate.

Fig. 2A depicts the protease substrate residues designated using the nomenclature of Schechter and Berger (1967) Biochem. Biophys. Res. Commun., 27 157-162. Figs. 2B and 2C depict first and second ligation substrates used in peptide ligation. The various residues R2, R1, R1' and R2' comprise amino acid residues or analogues to such residues for which the serine protease variant or precursor enzyme has specificity, e.g. those amino acid residues comprising P2, P1, P1' and P2'. The carboxy terminus of the first ligation substrate contains Group X which is a 2-hydroxy carboxylic acid esterified to the C-terminal carboxyl group of the substrate used to form the first ligation substrate. The residues R2'', etc. comprise amino acid residues or analogous to such residues for which the protease variant or precursor enzyme has specificity, e.g. P2', etc. Fig. 2D depicts the ligation product formed by ligation of the first and second ligation substrates (Figs. 2B and 2C) by the serine protease variants of the invention.

Fig. 3 depicts the catalytic residues of Bacillus amyloliquefaciens subtilisin including the catalytic triad Ser221, His64 and Asp32. The tetrahedral intermediate transition state is shown as EST.



Fig. 4 depicts the tertiary structure of subtilisin from Bacillus amyloliquefaciens subtilisin. The α -helix of subtilisin associated with the catalytic residue Ser221 is highlighted.

5 Fig. 5 is a stereo view of the α -helix of Bacillus amyloliquefaciens subtilisin associated with the catalytic Ser 221.

Fig. 6 depicts the amino acid sequence for subtilisin from Bacillus amyloliquefaciens, Bacillus subtilis 10 VarI168 and Bacillus licheniformis using standard single letter designations for naturally occurring amino acids.

Fig. 7 is a stereo view overlay of wild-type (open bar) and P225A (solid bar) Bacillus amyloliquefaciens subtilisin in the region containing α -helix containing 15 Ser221 and position 225.

Fig. 8 depicts an active site view showing a peptide substrate in the transition state bound to the active site of wild-type subtilisin (residues P4 through P2'). The hydrolysis substrate is bound in an outstretched 20 β -sheet like conformation with the main chain carbonyl oxygens and amides involved in hydrogen bonds with corresponding groups of the enzyme. Specifically, the residues S1-S4 form the central strand of a three-stranded antiparallel β -sheet (McPhalen et al., 25 (1988) supra). Some of the more important side chain interactions on either side of the scissile bond are emphasized. Other residues affecting substrate specificity, e.g. Glu156 and Gly166 for P1 specificity, are not shown.

30 Fig. 9 depicts the general structure and specific 2-hydroxy carboxylic acids that are useful in making first ligation substrates that mimic a peptide substrate.

The structural relationship between specific 2-hydroxy and amino acids is also depicted.

Figs. 10A, 10B and 10C depict progress curves showing the consumption of peptide ester substrate (succinyl-Ala-
5 Ala-Pro-Phe-glycolate-Phe-amide at a starting concentration of 0.35 mM) and the concurrent appearance of hydrolysis and aminolysis products using the dipeptide Ala-Phe-amide (3.6 mM) as the nucleophile (acyl-acceptor) with different subtilisin variants. The same amount of
10 enzyme (10 μ M) was used in each case under identical conditions, pH 8.0, (25 \pm 0.2 °C). Aliquots were taken a various times and analyzed by RP-HPLC. The subtilisin variant was Pro225Ala in Fig. 10A, Ser221Cys in Fig. 10B, and Ser221Cys/Pro225Ala in Fig. 10C (Aminolysis (●),
15 Hydrolysis (○)).

Fig. 11 is a comparison of subtilisin variants having different P-1 specificities with each of four alternative first ligation substrates. Progress curves show the consumption of substrate turning into hydrolysis- and
20 aminolysis products in the presence of 1.5 mM of the nucleophile ligation peptide Ala-Phe-amide. The substrates are abbreviated to their tetra peptide in the one-letter form, i.e., AAPF is s-Ala-Ala-Pro-Phe-glycolate-Phe-amide. In each case, aminolysis
25 predominates over hydrolysis.

Fig. 12 shows reducing SDS-PAGE of time course aliquots of growth hormone ligations using either a first ligation substrate ester of the natural first eight amino acids or a first ligation peptide modified to become more
30 suited for the wild type specificity of subtilisin. The second ligation substrate was des-octa hGH. Lane 1, Low molecular weight standard; Lanes 2 and 9, non-ligated des-octa hGH; Lanes 3- 7, aliquots from the reaction using the G166E/S221C/P225A version of the peptide ligase

with the peptide ester of the natural N-terminus, after times 1, 10, 20, 40 and 80 minutes; Lane 8, wild type growth hormone; Lane 10 - 14, aliquots from the reactions using the S221C/P225A form of the ligase with the alternative peptide ester, after times 1, 10, 20, 40 and 80 minutes. In all lanes containing the des-octa mutant hGH, two smaller bands can be seen below the major band. These are the two chains resulting from a proteolytic cut during the expression and purification from E. coli. Furthermore, in all time points aliquots of the subtilisin variant may be seen at an apparent weight of about 30 kDa.

Fig. 13 depicts the method of construction of DNA encoding the replacement of proline at position 225 15 Bacillus amyloliquefaciens subtilisin with alanine.

Figure 14 shows reducing SDS-PAGE of time course aliquots of Protropin ligations. A nine residue peptide was ligated onto Protropin using the Ser221Cys/Pro225Ala version of the peptide ligase. Lane 1, low molecular weight standard; Lane 2, subtilisin; Lane 3, unligated Protropin; Lanes 4-9, hGH ligated with a peptide after 20 1,2,5,10,30 and 60 minutes.

DETAILED DESCRIPTION

As used herein, "serine protease" refers to a protease 25 which contains at least one catalytically active serine residue. Such serine proteases are ubiquitous being found in both prokaryotic and eukaryotic organisms. A common characteristic of many serine proteases, such as the subtilisin-type and trypsin-type serine protease, 30 is the presence of a common catalytic triad comprising aspartate, histidine and serine. In the subtilisin-type proteases, the relative order of these amino acids, reading from the amino to carboxy terminus is aspartate-histidine-serine. In the trypsin-type proteases, the

relative order, however, is histidine-aspartate-serine. Notwithstanding this relative sequence orientation of the catalytic residues, the secondary and tertiary structure of serine proteases bring these three catalytic 5 residues into close proximity to form the catalytically active site.

Figs. 3 and 5 depict the catalytic residues of Bacillus amyloliquefaciens subtilisin. The interactions of serine 221 in subtilisin to form a tetrahedral transition state 10 with the carbon of the scissile peptide bond is shown in Fig. 3. An acyl-enzyme intermediate between the catalytic serine and the carboxy portion of the substrate is formed after the carboxy terminal portion of the substrate leaves the active site. Hydrolysis of the 15 acyl-enzyme intermediate releases the amino terminal portion of the cleaved peptide from the enzyme and restores the serine alcohol group.

In subtilisin-type serine proteases, the OG group of the catalytic side chain of serine (serine-221 in subtilisin) 20 is located near the amino terminus of a long central α -helix which extends through the molecule. Bott, et al. (1988) J. Biol. Chem. 263, 7895-7906. In Bacillus amyloliquefaciens subtilisin this α -helix comprises methionine 222 through lysine 237. This helix is 25 conserved in evolutionarily related subtilisin-type serine proteases but is not found in the catalytic sites of trypsin-type serine proteases. McPhalen, et al. (1988) Biochemistry 27, 6582-6598. The α -helix associated with the active site of subtilisin-type serine proteases has 30 led to the suggestion that the dipole of this helix may have a functional role in catalysis. Hol, W.G.J. (1985) Prog. Biophys. Molec. Biol. 45, 149-195. The lack of α -helix at the active site of the trypsin-type serine proteases, however, has raised the unresolved question 35 of whether the active site helix of subtilisin-type

serine proteases is of any significance in catalysis. Hol (1985) supra.

In Fig. 4, the α -helix of Bacillus amyloliquefaciens subtilisin together with the catalytic serine 221 is 5 shown as it relates to the tertiary structure of the enzyme. A stereo view of this α -helix associated with the catalytic serine 221 is shown in Fig. 5.

The amino acid sequence for subtilisin from Bacillus amyloliquefaciens, Bacillus subtilis VarII68 and Bacillus licheniformis is shown in Fig. 6. The α -helix in Bacillus amyloliquefaciens subtilisin extends from Met222 10 Lys237. The corresponding (equivalent) α -helix in Bacillus subtilis is the same and in Bacillus licheniformis subtilisin comprises Met221 to Lys236. The 15 corresponding (equivalent) residue of Proline 225 in Bacillus amyloliquefaciens subtilisin is also Proline 225 in Bacillus subtilis and Proline 224 in Bacillus licheniformis subtilisin. The catalytic serine in Bacillus subtilis is at position 221 and at position 220 20 in Bacillus licheniformis subtilisin.

As used herein, a "precursor serine protease" refers to a naturally occurring serine protease and recombinant serine proteases. Examples of naturally occurring precursor serine proteases include the bacterial 25 subtilisins from Bacillus amyloliquefaciens, Bacillus licheniformis, Bacillus amylosaccaridicus, and homologous serine protease from fungi, plant and higher animal species, trypsin-type serine proteases from bacteria fungi, plant, animal and virus including trypsin, 30 chymotrypsin, α -lytic protease, elastase, plasminogen, thrombin, tissue plasminogen activators and homologs thereof.



"Recombinant serine proteases" refers to serine proteases in which the DNA sequence encoding the naturally occurring serine proteases modified to produce a variant DNA sequence which encodes the substitution, insertion 5 or deletion of one or more amino acids in the serine protease amino acid sequence. Suitable modification methods are disclosed herein and in EPO Publication No. 0 130 756 published January 9, 1985 and EPO Publication No. 0 251 446 published January 7, 1988. 10 When a particular serine protease is referred to, e.g. subtilisin, the term precursor subtilisin and recombinant subtilisin are used consistent with these definitions.

In addition to naturally occurring and recombinant serine proteases, the term "precursor serine protease" also 15 includes "synthetic serine proteases" which are naturally occurring or recombinant serine proteases which contain one or more amino acid residues which are not naturally occurring. Such synthetic serine proteases may be made by the methods described herein or by in vitro 20 transcription-translation methods as disclosed by Schultz, et al. (1989), Science, 244, 182-188.

As used herein a "serine protease variant" refers to a serine protease having an amino acid sequence which is derived from the amino acid sequence of a precursor 25 serine protease. The amino acid sequence of the serine hydrolase variant is "derived" from the precursor protease amino acid sequence by the substitution, deletion or insertion of one or more amino acids of the precursor amino acid sequence. Such modification is of 30 the "precursor DNA sequence" which encodes the amino acid sequence of the precursor serine protease. Further, in some instances, the serine protease variant may be derived from the precursor protease by the direct chemical modification of one or more side chains of the 35 amino acid residues of the precursor amino acid sequence.

-15- 2087342

For example, in one of the preferred embodiments the serine 221 of Bacillus amyloliquefaciens subtilisin is converted to cysteine. Such conversion may be obtained by modifying the DNA sequence of subtilisin to encode 5 cysteine thereby replacing serine with cysteine in the variant or by the direct modification of serine with an appropriate chemical agent to form the equivalent variant. See e.g. Neet, K.E., et al (1966), Proc. Natl. Acad. Sci. USA, 56, 1606-1611; Polgar, L., et al. (1966), 10 J. Amer. Chem. Soc., 88, 3153-3154, which describe the chemical modification of subtilisin to form thiolsubtilisin.

In the preferred embodiments, the nucleophilic oxygen of the side chain of an active site serine residue is 15 replaced or modified to substitute the nucleophilic oxygen of that side chain with a different nucleophile. Preferred nucleophiles include -SH, -SeH, -NH₂. The most preferred nucleophile is SH- representing the replacement of serine with cysteine.

20 In addition, the serine protease variants of the invention include the replacement or modification of a second amino acid residue in the precursor serine protease. The modification of this second amino acid residue is made to perturb the active site to accommodate 25 the changes induced therein by the substitution or modification of the side chain of the active site serine residue. When the active site serine has been replaced or modified to produce a side chain which is larger than the naturally occurring serine side chain (e.g. when 30 cysteine replaces serine or selenium replaces the nucleophilic oxygen), the second amino acid residue is a residue which upon replacement or modification causes the new nucleophilic catalytic side chain to be displaced in a way which effectively increases the size of the 35 active site. Thus, for example, in the case of the

subtilisin-type serine proteases, the nucleophile residing at position 221 in Bacillus amyloliquefaciens subtilisin may be moved by modifying one or more amino acid residues in the α -helix which is located at or near

5 the active site serine in the precursor serine protease. In the case of subtilisin, such modification or replacement is preferably of proline 225 (Caldwell, et al. (1989), J. Cell. Biochem. supp., 13A, 51). Modification of this residue in Bacillus

10 amyloliquefaciens subtilisin, or an equivalent residue in other subtilisins, is by replacement with a helix-forming amino acid. Chou, P.Y., et al. (1974), Biochemistry, 13, 211; Kyte, J., et al. (1982), J. Mol. Biol., 157, 105; Rose, G., et al. (1977), J. Mol. Biol., 113, 153. Examples of such amino acids include alanine, leucine, methionine, glutamine, valine and serine. Helix breaking amino acids such as glycine and proline are not preferred. Alternatively, replacement amino acids comprise those which have a smaller side chain volume

15 than the residue being replaced. Chothia (1984), Ann. Rev. Biochem., 53, 537. Replacement with the aforementioned amino acids or their analogs modifies the discontinuity (kink) preceding residues 223-237 in the α -helix.

20

25 A preferred replacement amino acid for position 225 is alanine. When such a variant is made, the α -helix containing Ser221 is known to move the -OH nucleophile away from the oxyanion hole. The effect of this particular replacement on naturally occurring subtilisin

30 is shown in Fig. 7 which is a stereo view of residues 220 through 230 of wild type and Pro225Ala subtilisin. The -OH nucleophile of Ser221 is moved away from the oxyanion hole and the catalytic histidine 64 by 0.2 to 0.4 \AA . When combined with the replacement of Ser221 with

35 cysteine, this displacement of the 221 residue effectively offsets the increase in nucleophile size of



0.65Å and 1.03Å, respectively. Pauling, L. (1960) in The Nature of the Chemical Bond, 3d ed., Cornell Univ. Press, Ithica, N.Y., pp. 246-260.

A characteristic of the serine protease variant of the 5 invention is that it has peptide ligase activity in aqueous solution which is greater than that of a different serine protease containing only the substitution or modification of the nucleophilic oxygen of the active site serine residue. Thus, variants such 10 as Ser221Cys/Pro225Ala have greater peptide ligase activity than the Ser221Cys variant.

As used herein, "peptide ligase activity" refers to the ability of an enzyme to ligate two or more substrates. Such substrates include the ligation substrates described 15 hereinafter as well as known activated peptides such as peptide thiobenzyl esters p-chlorophenyl esters (Nakatsuka, et al. (1987), J. Amer. Chem. Soc., 109, 3808-3810), p-nitrophenyl esters and other aryl esters. In addition, activated esters include alkyl esters such 20 as methyl, ethyl, glycolate, lactate, etc. and alkylthiol esters. Peptide ligase activity is measured by contacting the enzyme of interest with at least two peptides or substrates (one of which generally contains an activated carboxy terminus) under conditions favorable 25 for ligation. The k_{cat} , K_m and/or k_{cat}/K_m ratio is then determined for that enzyme. k_{cat} refers to the turn-over number of the enzyme, and gives the maximal number of substrate molecules that can be converted to product per unit time per enzyme molecule. The K_m is usually 30 inversely proportional to the affinity of substrate(s) for the enzyme. The catalytic efficiency which is the second order rate constant for the conversion of substrate to product is given as the k_{cat}/K_m ratio. Variant serine proteases having higher k_{cat}/K_m ratios 35 for the same peptide ligation have greater peptide ligase



activity. However, they can also be considered to have greater peptide ligase activity if they have improved k_{cat} assuming that the reactions can be run where the enzymes are saturated with substrates. When comparing 5 the peptide ligase activity of two enzymes, each enzyme is preferably contacted with the same peptides under the same conditions.

In addition to the above described modifications of the active site serine and a second amino acid residue to 10 form serine protease variants having ligase activity, various other replacement or modifications of amino acid side chains may be made to modify specificity for the ligation peptides used to form a desired ligation peptide product. For example, the subtilisin variant 15 Ser221Cys/Pro225Ala is derived from the wild type subtilisin from Bacillus amyloliquefaciens. This wild type subtilisin preferentially hydrolyzes peptides wherein the P1 residue for the hydrolysis peptide shown in Fig. 2A consists of phenylalanine, tyrosine, 20 tryptophan, leucine, methionine and lysine but not Thr, Val, Ile, Pro or Gly. In addition, preferred hydrolysis peptides should not contain Ile, Pro, Asp or Glu at the P1' position or small amino acids such as Gly, Ala, Pro, Ser, Asp or Thr at the P2' position. In the subtilisin 25 variant Ser221Cys/Pro225Ala, the amino acid residues forming the substrate binding cleft of subtilisin are not modified and accordingly the various enzyme subsites within this binding cleft which interact with one or more of the substrate residues, e.g. P2 through P2', are still 30 capable of binding the normal substrate for the wild type subtilisin. Of course, this variant demonstrates substantially reduced catalytic activity with such peptide substrate because of the modifications made to form the subtilisin variant.

-19- 2087342

Such a subtilisin variant, accordingly, is capable of binding ligation substrates containing amino acid residues or analogs corresponding to or closely related to those found in the normal peptide substrate. Thus, 5 the "first ligation substrate" as schematically represented in Fig. 2B contains residues Rn through R1 (reading from the amino to carboxy terminus) wherein R1 is a large hydrophobic amino acid or analog similar to that normally found in the P1 position of a subtilisin 10 hydrolysis substrate. Similarly, R2 corresponds to or is closely related to the amino acid residue (or its analog) normally found in position P2 of the hydrolysis substrate. The group X which is covalently linked to the carboxy terminus of the first ligation substrate and 15 which activates the first ligation peptide and the residues R2" through Rn will be discussed in more detail hereinafter.

The "second ligation substrate" represented schematically in Fig. 2C contains residues R1' through Rn' (reading 20 from the amino to carboxy terminus). When used with the above identified subtilisin variant modified at residues 221 and 225, R1' corresponds to the P1' amino acid residue (or its analog) in the normal subtilisin hydrolysis substrate. With regard to second ligation 25 substrate residue R2', this residue comprises a large hydrophobic amino acid residue (or its analog) which is similar to or corresponds closely to the P2' residue of the normal subtilisin hydrolysis substrate. For both the first and second ligation substrates, other amino 30 acid residues may be chosen to correspond to or be closely related to the amino acid residues (or their analogs) found in normal hydrolysis substrate. Appropriate R and R' residues for other serine protease 35 are chosen but not limited to the natural hydrolysis substrates for such proteases.

2087342

-20-

In addition, either the first or second ligation substrate may include other compounds that result in the site specific modification of the product. The specific compounds may be chosen to specifically target certain 5 sites within the product. For example, the first or second ligation substrate may contain heavy metal ions, that will result in heavy metal derivatives of the ligation product, useful in the X-ray crystallographic elucidation of the ligation product structure. The first 10 or second ligation substrate may also contain modified or isotopically labelled amino acids for biophysical studies.

Furthermore, the conditions under which the ligation reactions are carried out may be modified as needed to 15 allow ligation to occur. These modifications may be required to allow the ligation substrates to take the correct structural conformation for ligation to occur, without substantially destroying the ligase activity. For example, the experimental conditions may include the 20 addition of denaturing reagents, detergents, organic solvents or reducing agents, or alterations in pH or temperature, among others.

The the first and second ligation substrates, as well as the experimental conditions, may be selected as needed 25 to produce the desired level of specificity. Some applications of the present invention may require lower levels of ligation reaction specificity, while in other applications strict specificities may be desirable.

As indicated herein, the subtilisin variant 30 Ser221Cys/Pro225Ala is capable of ligating a first ligation substrate (containing an activation group) and a second ligation substrate in accordance with the above described preference for R1 and R2' amino acid residues. The peptide ligation activity for this variant is

2087342

-21-

substantially greater than that of the Ser221Cys variant for the same ligation substrates. This variant's preference for first and second ligation substrates having amino acids Phe, Tyr, Met, Leu, Trp and Lys at 5 position R1 and amino acids Phe, Tyr, Leu and Met at positions R2' provides substantial utility for ligating block substrates such as those made with known chemical synthetic techniques.

Synthetic ligation substrates are generally about 15-25 10 residues long and produce a ligation product having a length of about 50 residues. Such first ligation product may thereafter be ligated with another ligation substrate to build larger ligation product.

In order to provide broader specificity for substrate 15 ligation, other modifications may be made to the serine protease variant to change specificity for the first and/or second ligation substrate. As described herein, subtilisin variants containing the Ser221Cys/Pro225Ala modifications as well as modifications in the Glu156 and 20 Gly166 residues were made to modify the specificity of the variant for the R1 residue of the first ligation substrate. Three variants of the Ser221Cys/Pro225Ala variant were made. These specific variants included the modifications 221 and 225 and further included Gly166Glu, 25 Glu156Gln/Gly166Lys and Gly166Ile, Using the standard one-letter symbols for amino acid residues, these variants can be identified as G166E/S221C/P225A, E156Q/G166K/S221C/P225A and G166I/S221C/P225A respectively. By introducing these mutations into the 30 S221C/P225A variant, the peptide ligase specificity was substantially altered. See Fig. 11 wherein the ligation of various first ligation substrates (discussed in more detail hereinafter) and the second ligation substrate Ala-Phe-amide is shown.

20873A2

For a small R1 ester substrate (s-Ala-Ala-Pro-Ala-glc-Phe-amide) the G166I/S221C/P225A variant (ICA) is substantially better than the others (Fig. 11). The E156Q/G166K/S221C/P225A variant (QKCA) efficiently 5 aminolyses a Glu R1 first ligation substrate with a second ligation substrate Ala-Phe-amide (AF-amide). It also has greater peptide ligation activity than the other variants toward a Phe R1 first ligation substrate. For 10 a Lys R1 ester substrate the rates for three of the variants (including the complementary charged G166E/S221C/P225A variant (ECA)) are comparable and much 15 more active than for the like charged mutant, E156Q/G166K/S221C/P225A. For ligation of peptide substrates containing an R1 Arg ester, our preliminary data indicates G165E/S221C/P225A is substantially more active than the parent ligase. See Table I.

In general, the aminolysis rates for the optimal enzyme ligation peptide pair were comparable indicating it should be possible to efficiently ligate Lys, Ala, Phe, 20 and Glu R1 ligation substrates with the proper choice of S221C/P225A based ligase S221C/P225A (CA). Except for the Lys R1 ligation substrate, at least one of the three other specificity variants were significantly better than the parent peptide ligase. These additional 25 variant enzymes should provide added flexibility in design of ligation junctions.

These results are consistent with the demonstrated specificity of subtilisin variants G166E, E156Q/G166K and G166I for hydrolysis of peptides containing Lys or 30 Arg, Glu and Ala P1 substrates, respectively (Wells, et al. (1987) Proc. Natl. Acad. Sci. USA, 84, 1219-1223; Estell, et al. (1986), Science, 233, 659-663, and EPO Publication 0 251 446). It is expected that other modifications which are known to cause a change and/or 35 shift in substrate specificity for various P and P'

residues in wild type subtilisin can be effectively combined with the Ser221/Pro225 or equivalent modifications to further modify the specificity of the serine hydrolase variant for various first and second 5 ligation substrates.

TABLE I

Summary of preferred sequences for ligating peptides using variants of subtilisin.

	Residue						
	P4	P3	P2	P1	R1'	R2'	R3'
Avoid:				G,P,T,V,I	I,P,D,E ^a	P,G	
Preferred:	Small or large hydrophobics	flexible	flexible	M,Y,L(1) ^b	R,C,N	F,Y	flexible
				F(1,4)	T,K,H	L,M	
				K(1,2,3)	W,Q,Y	R,K	
				R(3)	A,V,S,G		
				A(2)			
				E(4)			

^(a)The deleterious effects of the Glu and Asp side-chains can be minimized in high salt (>1M; Carter, et al., (1989), Proteins, 6, 240-248).

^(b)These residues are preferred with the following variants of thiolsubtilisin: (1) = S221C/P225A; (2) = G166I/S221C/P225A; (3) = G166E/S221C/P225A; (4) = E156Q/G166K/S221C/P225A.

Many of these modifications to Bacillus amyloliquefaciens subtilisin are disclosed in EPO Publication 0 130 756 published January 9, 1988, and EPO Publication No. 0 251 446, published January 7, 1988. Such modifications

2087342

may be readily made to the Ser221/Pro225 variants described herein (as well as other variants within the scope of the invention) to form variants having a wide range of specificity for first and second ligation substrates. The methods disclosed in these EPO publications may be readily adapted to modify the DNA encoding the variants described herein.

In addition to combining modifications which affect substrate substrate specificity, it is also possible to 10 combine other modifications which affect other properties of the serine protease. In particular, modifications have been made to subtilisin which affect a variety of properties which may be desirable to combine with the variants of the present invention. For example, 15 substitution of methionine at position 50 Phe or Cys and at position 222 with Ala, Gly, Ser and Cys results in a subtilisin variant which is oxidatively stable as compared to wild type subtilisin. In addition, there are known modifications to subtilisin which result in 20 increased thermal stability, alkaline stability and changes in pH activity profile. See e.g. EPO Publication 0 251 446 published January 7, 1988. The invention contemplates combining these and other possible 25 modifications to form variants which in addition to having ligation activity are also characterized by changes in one or more other properties of the precursor enzyme. For example, serine protease variants which resist inactivation at higher temperatures than that of the variant not containing such modifications are useful 30 to ligate first and second ligation substrates wherein an increase in reaction temperature facilitates a partial or complete denaturation of one or more of the ligation peptides to increase the ligation yield. Similarly, variants which have optimal activity at a pH which also 35 facilitates denaturation of ligation substrate may prove useful in specific applications. Of course, various

other modifications not presently known but which are found to confer desirable properties upon the serine protease variants of the invention are contemplated to be within the scope of the invention.

5 Although activated aryl-ester substrates (such as thiobenzyl esters) are more efficient than corresponding alkyl-esters to acylate subtilisin, aryl-esters are more difficult to synthesize and inherently less stable. A series of alkyl-ester substrates was prepared to improve
10 upon their catalytic efficiencies as donor substrates for acyl-enzyme intermediate formation. Peptide substrates bind to subtilisin in an extended anti-parallel β -sheet conformation from residues P4 to P3' (McPhalen and James (1988), Biochemistry, 27, 6592-6598). Although the P4 and P1 residues dominate the substrate specificity of the enzyme (for review see
15 Philipp and Bender (1983), Mol. Cell. Biochem., 51, 5-32; Estell et al. (1986), Science, 233, 659-663), the catalytic efficiency for hydrolysis is enhanced
20 significantly when peptide substrates are extended from P1' to P3' (Morahara et al. (1970), Arch. Biochem. Biophys., 138, 515-525).

Referring to Figs. 2B and 2C, first and second ligation substrates are shown with R1 and R2' groups being as
25 previously described. The leaving group X in the ester in Fig. 2B may be any of the following organic alcohols or thiols: C_6-C_{12} aryl where the aryl group is unsubstituted or substituted by one or more of the groups nitro, hydroxy, halo (F, Cl, Br, I), C_1-C_8 alkyl, halo- C_1-C_8 alkyl, C_1-C_8 -alkoxy, amino, phenoxy, phenyl, acetamido, benzamido, di- C_1-C_8 alkylamino, C_1-C_8 alkylamino, C_6-C_{12} aroyl, C_1-C_8 alkanoyl, and hydroxy- C_1-C_8 alkyl, C_1-C_{12} alkyl either substituted or unsubstituted, branched, straight chain or cyclo where the substituents
30 are selected from halo (F, Cl, Br, I), C_1-C_8 alkoxy, C_6-C_{12}

2087342

aryloxy where the aryl group is unsubstituted or substituted by one or more of the groups nitro, hydroxy, halo (F, Cl, Br, I), C₁-C₈ alkyl, C₁-C₈-alkoxy, amino, phenoxy, acetamido, benzamido, di-C₁-C₈ alkylamino, 5 C₁-C₈ alkylamino, C₆-C₁₂ aroyl, and C₁-C₈ alkanoyl, isothioureido, C₃-C₇ cycloalkyl, ureido, amino, C₁-C₈ alkylamino, di-C₁-C₈ alkylamino, hydroxy, amino-C₂-C₈ alkylthio, amino-C₂-C₈ alkoxy, acetamido, benzamido wherein the phenyl ring is unsubstituted or substituted 10 by one or more of the groups nitro, hydroxy, halo (F, Cl, Br, I), C₁-C₈ alkyl, C₁-C₈-alkoxy, amino, phenoxy, acetamido, benzamido, di-C₁-C₈ alkylamino, C₁-C₈ alkylamino, C₆-C₁₂ aroyl, C₁-C₈ alkanoyl, C₆-C₁₂ arylamino wherein the aryl group is unsubstituted or substituted 15 by one or more of the groups nitro, hydroxy, halo, C₁-C₈ alkyl, C₁-C₈-alkoxy, amino, phenoxy, acetamido, benzamido, di-C₁-C₈ alkylamino, C₁-C₈ alkylamino, C₆-C₁₂ aroyl, and C₁-C₈ alkanoyl, guanidino, phthalimido, mercapto, C₁-C₈ alkylthio, C₆-C₁₂ arylthio, carboxy, 20 carboxamide, carbo-C₁-C₈ alkoxy, C₆-C aryl wherein the aryl group is unsubstituted or substituted by one or more of the groups nitro, hydroxy, halo, C₁-C₈ alkyl, C₁-C₈-alkoxy, amino, phenoxy, acetamido, benzamido, di-C₁-C₈ alkylamino, C₁-C₈ alkylamino, hydroxy-C₁-C₈ alkyl, C₆-C₁₂ aroyl, and C₁-C₈ alkanoyl, and aromatic heterocycle 25 wherein the heterocyclic groups have 5-10 ring atoms and contain up to two O, N, or S heteroatoms.

In one aspect of the invention, X is preferably a 2-hydroxy carboxylic acid. The general formula for a 2-hydroxy carboxylic acid is shown in Fig. 9. As can be seen therein, the core structure of the 2-hydroxy carboxylic acid is similar to the core structure for an amino acid except for the replacement of the amino group for the 2-hydroxy group. Accordingly, an appropriate side chain R group may be chosen for the 2-hydroxy acid to correspond to the side chain R groups found on 30 35



naturally occurring amino acids. Thus, the various 2-hydroxy carboxylic acids e.g. glycolate corresponding to the amino acid glycine and lactate corresponding to the amino acid alanine, etc. may be esterified with the 5 carboxy terminus of a substrate to form a first ligation substrate.

First ligation substrates were constructed wherein the 2-hydroxy carboxylic acid was glycolate or lactate. In essence, the 2-hydroxy carboxylic acid acts as an amino 10 acid residue which may bind to that portion of the enzyme binding cleft which interacts with the P1' residue of a hydrolysis substrate. Further, the free-carboxyl group of the 2-hydroxy carboxylic acid may be amidated to form an amide or amidated with an amino acid residues (or 15 analog) or peptide (or peptide analog) represented by R2" or R2" through Rn", respectively, as shown in Fig. 2B. As indicated therein, the R2" is preferably chosen to optimize the interaction with that portion of the binding cleft of the enzyme which normally interacts with 20 the hydrolysis substrate residue P2'. Similar analogies exist for other R" residues. The leaving group thus obtained activates the first ligation substrate such that the activation energy for ester cleavage is lowered by optimizing binding to the serine protease variant.

25 As used herein, a "ligation product" is formed by ligation of first and second ligation substrate by the serine hydrolase variants of the invention. It is to be understood, however, that ligation products and the first and second ligation peptides need not be made 30 entirely of naturally occurring amino acids and further need not be entirely proteinaceous. In this regard the only requirement of first and second ligation substrates is that they contain at least an R1 amino acid or functional analog thereof (at the carboxy terminus of 35 the first ligation peptide) and an amino acid or



functional analog at the R1' position of the second ligation peptide. Such R1 and R1' residues are capable of binding in the appropriate portion of the substrate binding cleft of the serine hydrolase variant such that 5 ligation occurs. Specifically included in R1 and R1' are those amino acid residue analogs which are functional with the serine protease variants of the invention. Such analogs include L-selenocysteine, L-selenomethionine and L-propargylglycine (sigma).

10 To the extent that such binding and ligation also requires R2 and R2' amino acids (or their analogs) or additional amino acids R3 or R3', etc. (or their analogs) the first and second ligation substrates peptides will contain such structure. However, first and/or second 15 ligation substrates may contain non-naturally occurring amino acids at positions outside of the region required for binding and ligation. Further, since it is only necessary for the first and second ligation substrates to have sufficient binding to bring about ligation, the 20 first and second ligation substrates and the corresponding ligation product formed therefrom may contain virtually any chemical structure outside of the necessary binding and ligation region. Accordingly, the invention is not limited to ligation substrates and 25 ligation products which correspond to a polypeptide or protein containing naturally occurring amino acids.

The efficiency of peptide ligation using a series of glycolate and lactate-esters with S221C/P225A subtilisin was analyzed (Table II). As indicated, there is a 30 systematic increase in k_{cat}/K_M of about 10-fold in extending esters from -glc-amide through -glc-Phe-Gly-amide. Most of this increase is the result of lower K_M values. There is a similar progression starting from -lac-amide that further illustrates the advantage of 35 extending the ester chain length. The lactate-ester

series is generally 4- to 5-fold less reactive than the glycol-*ce*-ester series and contains an additional chiral center. Therefore, because of the improved catalytic efficiency and ease to synthesize, the -glc-Phe-amide ester substrate was further studied.

TABLE II

Substrate leaving group comparison. The acyl-donor part is in all cases *s*-Ala-Ala-Pro-Phe- and is followed by the alternative leaving groups. The initial rate of ligation using the dipeptide -Ala-Phe-amide (3.1 mM) was measured. ^aThe values of k_{cat}/K_M , the aminolysis fraction of the apparent second order rate constant for the reaction between the substrate and the enzyme, are the most reliable (error of about 20%).

	leaving group ^a	$k_{cat}(s^{-1})$	$K_M (mM)$	$\frac{k_{cat}}{K_M} (s^{-1}mM^{-1})^a$
	-glc-amide	9.5	3.0	3.20×10^3
20	-glc-Phe-amide	20.0	1.3	15×10^3
	-glc-Phe-Gly-amide	20.0	0.6	34×10^3
	-lac-amide	3.8	4.5	0.86×10^3
	-lac-Leu-amide	2.3	1.9	1.2×10^3
	-lac-Phe-amide	2.9	1.0	2.9×10^3
25	-Sbz	230.0	0.4	650×10^3

The P225A variant rapidly and quantitatively hydrolyzes the -glc-Phe-amide-ester with very little aminolysis (Fig. 10A). The S221C variant aminolyzes the substrate slowly and hydrolyzes the adduct (ligation product) so that during this time almost one-third of the substrate becomes hydrolyzed (Fig. 10B). However, the S221C/P225A variant gives rapid and almost quantitative aminolysis (>90%; Fig. 10C). Moreover, as expected from the very

-30-

low amidase activity of S221C/P225A (Table III), the aminolysis product (ligation product) was not detectably hydrolyzed unlike the result using the S221C or P225A variants (data not shown). On the basis of rapid 5 aminolysis and slow ligation product hydrolysis, S221C/P225A is a more useful enzyme than either of its parent single mutants for ligation of substrates using the -glc-Phe-amide-ester donor substrate.

Table III: Kinetic constants for the hydrolysis of an 10 amide substrate (s-Ala-Ala-Pro-Phe-pNA) and for an activated ester substrate (s-Ala-Ala-Pro-Phe-Sbz). The ratio of esterase to amidase activities is the ratio of the apparent second order rate constants (kcat/KM). The 15 aminolysis to hydrolysis ratio was investigated using the thiobenzyl ester substrate with 3.6 mM of the dipeptide Ala-Phe-amide as the nucleophile.

TABLE III

enzyme	S-Ala-Ala-Pro-Phe-pNA		S-Ala-Ala-Pro-Phe-Sbz		esterase/ amidase	aminolysis/ hydrolase
	kcat	KM	kcat	KM		
wild type ^a	(4.4±0.01)x10 ¹	(1.8±0.1)x10 ⁻⁴	(2.5±0.1)x10 ⁵	(2.3±0.1)x10 ³	(1.9±0.1)x10 ⁻⁴	(1.2±0.1)x10 ⁷
Pro225Ala	4.1±0.04	(7.8±0.2)x10 ⁻⁴	(5.2±0.1)x10 ³	(2.3±0.03)x10 ³	(3.8±0.1)x10 ⁻⁴	(6.2±0.1)x10 ⁶
Ser221Cys	(1.3±0.04)x10 ⁻³	(4.9±0.4)x10 ⁻⁴	(2.7±0.1)	1.4±0.1	(5.5±0.1)x10 ⁻⁵	(2.5±0.3)x10 ⁴
221Cys/225Ala	<3x10 ⁻⁵	ND	(4.1±0.2)x10 ¹	(1.9±0.2)x10 ⁻⁴	(2.1±0.2)x10 ⁵	ND
					2.5x10 ¹	

(a) Data for the pNA substrate and for the Sbz substrate from Carter and Wells (1988), Nature, 332, 564-568; and Wells, et al. (1986), Phil. Trans. R. Soc. Lond., A317, 415-423, respectively.

Sequence requirements for the nucleophilic acceptor peptide (second ligation substrate) were investigated by determining the ligation efficiency of a series of acceptor dipeptides having the form $\text{NH}_2\text{-R1'-Phe-amide}$, 5 where R1' corresponds to the amino-terminal residue of the second ligation (acceptor) substrate that resides in the R1' binding site during attack of the thioacyl-enzyme intermediate. As the R1' residue is varied in size or charge (Gly, Ala, Leu, Arg) the 10 apparent second order rate constant for aminolysis of s-Ala-Ala-Pro-Pro-Phe-glc-Phe-amide varies less than 7-fold (Table IV). This is consistent with the relatively broad specificity for hydrolyzing various P1' peptide substrates (Carter et al. (1989), Proteins, 6, 240-248).

15

TABLE IV

Comparison of efficiency in the aminolysis reaction of different di- and tri-amino acid peptides. The apparent second order rate constants for the reaction of the nucleophile with the peptide ligase acylated by s-Ala-20 Ala-Pro-Phe are compared.

	peptide ^a		peptide	
	GF	0.8	GA	0.1
	AF	1.0	GL	0.3
25	LF	0.3	FG	0.006
	RF	2.0	AFA	2.0
	RG	0.04	LFD	0.3

30 ^(a)The one-letter codes are used. All peptides were amidated at the carboxy terminus.

The R2' site was probed with a series of dipeptides having the form $\text{NH}_2\text{-Gly-R2'-amide}$ (Table IV). Although

there is a preference for larger hydrophobic amino acids, the rate constant for ligation varies only 8-fold in going from Ala to Leu to Phe for S221C/P225A. For some dipeptide combinations the difference can be much larger.

5 For example, Arg-Phe is nearly 100-fold faster to aminolyze than Arg-Gly. Moreover, bad combinations, such as a large hydrophobic amino acid at R1' and Gly at R2' can make for extremely poor ligation substrates (compare NH₂-Phe-Gly-amide with NH₂-Gly-Phe-amide; Table IV).
10 Extending the nucleophilic peptide can enhance the catalytic efficiency of ligation 2- to 3-fold (compare NH₂-Ala-Phe-amide with NH₂-Ala-Phe-Ala-amide; Table IV).

As a test for ligation of two large ligation substrates, a peptide ester was synthesized containing the first 15 eight amino acids of hGH (FPTIPLSR) esterified to glycolate-Phe-amide (first ligation peptide). The acceptor peptide fragment was des-octa hGH, that contained residues 9-191 of hGH (second ligation peptide). The G166E/S221C/P225A variant produced 20 ligation product after 80 min. having the expected molecular weight (Fig. 12). Amino-terminal sequencing of the first 10 residues of the product showed that a single FPTIPLSR fragment was ligated properly to the N-terminus of des-octa hGH beginning with sequence Leu-Phe- 25 Asp to produce the full-length hormone. The parent peptide ligase (S221C/P225A) was significantly less efficient at ligating the two hGH peptide fragments. The improved efficiency of the G166E/S221C/P225A enzyme over the parent ligase is attributed to its increased 30 activity for Arg R1 substrates due to the G166E substitution in the P1 binding site. Polymerization of the unprotected peptide ester was not observed. This is most likely the result of the Pro residue at the second amino acid residue of the first ligation peptide 35 which is a very poor P2' residue (Carter and Wells, (1989), supra).

2087342

To demonstrate the utility of the ligase (S2221C/P225A), a nonomer ester first ligation substrate (FPTIPAAPF) was constructed that mimicked the optimal enzyme substrate (s-Ala-Ala-Pro-Phe-glc-Phe-amide). This peptide ester 5 was ligated onto des octa hGH (second ligation substrate) by S2221C/P225A (Fig. 12). Protein sequencing of the ligation product only gave the expected amino-terminal sequence, not the unreacted amino-terminus of des-octa hGH. This suggests that ligation only occurred to the 10 α -amino-group of hGH and not to ϵ -amino-groups of lysine.

This semi-synthesis of hGH is the first example of ligation of peptide fragments in aqueous to form a large polypeptide. Previous ligation with thiolsubtilisin in solutions containing greater than 50% DMF 15 (dimethylformamide) reportedly produced a peptide having a length of no longer than 17 amino acid residues. Nakatsuka, et al. (1987), J. Am. Chem. Soc., 109, 3808-3810. In the ligation described herein, first and second ligation peptides were ligated with a serine protease 20 variant of the invention to form a ligation product having 191 amino acid residues. Further, such ligation was accomplished in aqueous solution containing less than 2% by volume of nonaqueous solvent and did not require that peptide side-chains contain protector groups.

25 As used herein, the term "aqueous solution" refers to any solution that contains water. In some instances, an aqueous solution may comprise as little as 1% to approximately 5% water. However, an aqueous solution typically comprises greater than about 50% to 100% 30 aqueous solution (excluding solutes). Accordingly, solutions containing a small percentage of nonaqueous solvent are considered to be within the scope of the definition of aqueous solution.

It is to be understood, however, that to the extent that the variants of the invention are only defined with regard to the peptide ligase activity of the serine protease variant in aqueous solution as compared to a 5 variant modified only at the catalytic serine. This definition does not preclude use of the serine protease variants of the invention in solutions containing little or no water.

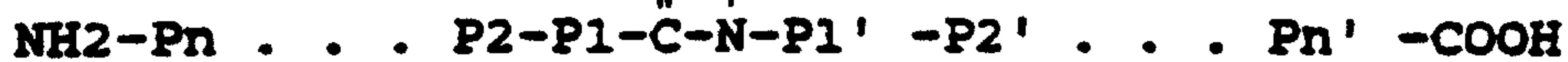
The following is presented by way of example and is not 10 to be construed as a limitation of the scope of the claims.

Materials and Methods

Abbreviations: DMA, dimethylacetamide; DMSO, dimethylsulfoxide; DTNB, 5,5'-dithiobis(2-nitrobenzoic acid); DTT, DL-dithiothreitol; hGH, human growth hormone; NEM, N-ethyl maleimide; PAGE, polyacrylamide gel electrophoresis; SDS, sodium dodecyl sulfate; s-Ala-Ala-Pro-Phe-pNA, N-succinyl-L-Ala-L-Ala-L-Pro-L-Phe-para-nitroanilide; s-Ala-Ala-Pro-Phe-Sbz, the thiobenzyl ester of the same succinylated peptide; TFA, trifluoroacetic acid; Tricine, N-tris(hydroxymethyl) methylglycine; Tris, tris(hydroxymethyl) aminomethane; E-Ac, acyl- or thioacyl-enzyme intermediate. Mutant proteins are designated by the wild-type residue 15 (single-letter amino acid code) followed by their position and the mutant residue. Multiple mutants are separated by a slash. For example, S221C/P225A indicates that the serine at position 221 and the proline at position 225 have been replaced by cysteine and alanine, 20 respectively. Protease substrate residues are designated 25 using the nomenclature of Schechter and Berger (1967), 30

61051-2487

-36-



where the scissile peptide bond is between the P1 and
5 P1' residues.

Materials: Enzymes for DNA manipulations were from New England Biolabs or Bethesda Research Labs. Oligonucleotides were synthesized by the Organic Chemistry Department at Genentech. All peptides contain 10 L-amino acids unless otherwise indicated, and were synthesized by standard methods (Barany and Merrifield, 1979). DL-Dithiothreitol (DTT)¹, DTNB, 2-mercapto ethanol, NEM, TFA, Tween[®]80, Tricine, dimethylsulfoxide, dimethylacetamide, and the substrates s-Ala-Ala-Pro-Phe- 15 pNA and s-Ala-Ala-Pro-Phe-Sbz were from Sigma. The solvents ethanol and acetonitrile were from J. T. Baker Inc. and ammonium sulfate from ICN Biochemicals Inc. Dipeptides Ala-Phe-amide, Arg-Gly-amide, Arg-Phe-amide, Gly-Ala-amide, Gly-Leu-amide, Gly-Phe-amide, Leu-Phe- 20 amide, Phe-Gly-amide and the tripeptide Ala-Phe-Ala-amide were obtained from BACHEM Feinchemikalien AG. The tripeptide Leu-Phe-Asp-amide was synthesized according to the general methods described in G. Barany and R.B. Merrifield (1979), in "Solid-Phase Peptide Synthesis" 25 in the Peptides, Analysis, Synthesis Biology Special Methods i Peptide Synthesis, Part A, Vol. 2 (E. Grow, J. Meienhatter (eds.), N.Y., Academic Press), pp. 3-254. Activated thiol Sepharose as well as G-25 and G-75 Sepharose were obtained from Pharmacia LKB Technology 30 AB.

Expression and Purification of Variant Subtilisins: The subtilisin gene in the M13-E. coli-B. subtilis shuttle plasmid, pSS5 (Carter and Wells (1987), *Science*, 237, 398-399), was expressed in the *B. subtilis* host strain

*Trade-mark

(BG2036) that is lacking its endogenous subtilisin and neutral protease genes (Yang, M.Y., et al. (1984), J. Bacterial, 160, 15-21). Since maturation of subtilisin involves proteolytic removal of the prosequence (Power et al. (1986), Proc. Natl. Acad. Sci. USA, 83, 3096-3100), the variants with reduced protease activity were expressed in the presence of active subtilisin. This was done either by adding a small amount of purified subtilisin (to a final concentration of 500 µg/L) late in the logarithmic growth phase, or by co-culturing from an inoculum of BG2036 containing 0.1% wild type subtilisin expressing cells (Carter and Wells, (1987), supra).

The purification of inactive subtilisin variants was essentially as described (Carter and Wells (1987), supra) except that an equal amount of cold ethanol was added to the supernatant to precipitate impurities prior to the precipitation of subtilisin by the addition of two additional volumes of cold ethanol. Furthermore, the CM-Trisacryl was substituted by SP-Trisacryl M in the ionexchange chromatography step. For the S221C mutants, the active site cysteine was utilized for purification on activated thiol Sepharose. This latter step is essential in order to separate the variant proteins from any traces of wild-type "helper" subtilisin. The equivalent step in the original procedure used a cysteine residue introduced on the surface of the protein that results in efficient removal of wild-type activity (Carter and Wells (1987), supra; Carter and Wells (1988), Nature, 332, 564-568). The mutant P225A is capable of autoproteolytic processing, and therefore was cultured without helper subtilisin and was purified by standard procedures (Estell et al. (1985), J. Biol. Chem., 260, 6518-6521).



Kinetic Assays. The esterase and amidase activities were obtained from initial rate measurements using a Kontron Uvikon 860 spectrophotometer. The assay for esterase activity utilized the substrate s-Ala-Ala-Pro-Phe-Sbz 5 at (25 ± 0.2 °C) in 100 mM Tris-HCl (pH 8.60), 4% (v/v) dimethylsulfoxide, 0.005% (v/v) Tween 80. With the non-cysteine containing proteases, DTNB (Ellman (1959), Arch. Biochem. Biophys., 82, 70-77) was added to a concentration of 37.5 μM to visualize the release of 10 thiobenzoate upon hydrolysis of the substrate. With the S221C derivative proteases the difference in absorbance at 250 nm between the substrate and the hydrolyzed product was used to monitor the reaction directly. The amidase activities were measured under 15 identical conditions by following the increase in absorbance at 412 nm upon hydrolysis of p-nitroanilide from s-Ala-Ala-Pro-Phe-pNA.

Enzymatic ligation of peptides were performed at (25 ± 0.2) °C in 90 mM Tricine (pH 8.0), 2% (v/v) 20 dimethylacetamide, 0.005% (v/v) Tween 80. The comparison between substrates having different leaving groups was simplified by measuring initial reaction rates at a low substrate concentrations (70-75 μM) and at a higher concentration (1.33 mM) where rates are proportional to 25 kcat/KM and the kcat, respectively (Fersht (1977) in Enzyme Structure and Mechanism, W.H. Freeman & Co. USA). The higher substrate concentration may still be below the KM for some of the substrates, so that values for kcat are less accurate. The aminolysis with di- and 30 tripeptides was performed at low peptide concentrations. The aminolysis rate $V = k_{\text{aminolysis}} [N] [E-\text{Ac}] / KN$, where [N] is the nucleophile concentration, [E-Ac] is the concentration of acyl-enzyme intermediate and KN is the dissociation constant for the binding of the nucleophile 35 to the acyl-enzyme intermediate (Riechmann and Kasche (1985), Biochem. Biophys. Acta., 830, 164-172). Since

$KN = [N] [E-Ac] / [N \cdot E-Ac]$ a change in [N] will result in a change in [E-Ac] but at low [N], $[E-Ac] \gg [N \cdot E-Ac]$ and $k_{aminolysis} / KN$ (the apparent second order rate constants for the reaction between the different 5 nucleophile and the acyl-enzyme intermediate) may be compared. The concentrations of the peptide nucleophiles, and the calibration of the absorbance data for the different ligation products was obtained from amino acid composition analysis. The rates of peptide 10 ligation were measured at four or five different concentrations for each nucleophile.

Ligation reactions were analyzed by taking aliquots at different times and analyzing the peptide products by C-18 reversed phase HPLC. Peptides were eluted in a 15 gradient of acetonitrile/0.1% TFA in water/0.1% TFA and the absorbance at 214 nm monitored. Amino acid composition analysis was used to confirm both the hydrolysis and aminolysis products and to calibrate the absorbance values. The structures of the hydrolysis and 20 aminolysis products (using the dipeptide Ala-Phe-amide as a nucleophile) were confirmed by mass spectrometrical analysis.

Example 1

Production of Subtilisin Variants

25 Molecular modeling was performed on an Evans and Sutherland PS300 using the program FRODO (Jones (1978), J. App. Crystallogr., 11, 268-272) and coordinates from a 1.8 Å resolution structure of subtilisin BPN' from Bacillus amyloliquefaciens (Bott et al. (1988), J. Biol. Chem., 263, 7895-7906). The S221C mutation was introduced (Carter et al. (1986), Nucl. Acids Res., 13, 4431-4443)) into the wild-type subtilisin gene (Wells et al. (1983), Nucl. Acids Res., 11, 7911-7925) using the oligonucleotide 5'-ACAACGGTACCT~~G~~GCATGGCATCTCC 30 (asterisks indicate the positions of altered nucleotides 35

and underlined is a unique KpnI site). The S221C/P225A mutations were introduced into the S221C template using the oligonucleotide 5'- GCGTACAACGGTAC^{*}TGCATGGCA TCTGCGCACGTTGCC (asterisks indicate altered nucleotides 5 and underlined is a new FspI site) by restriction-selection against the KpnI site (Wells et al. (1986), Phil. Trans. R. Soc. Lond., A317, 415-423). The construction of the mutants G166E and E156Q/G166K was described by Wells et al. (1987), Proc. Natl. Acad. Sci. USA, 84, 1219-1223, and the mutant G166I by Estell et al. (1986), Science, 223, 659-663. Combinations of the mutations around the active site (positions 221 and 225) and around the P1 binding pocket (156 and 166) were obtained by ligation of mutated restriction fragments 15 split by the enzyme PpuMI. See EPO Publication No. 0 251 446, published January 9, 1988. All mutants were verified by dideoxy sequencing (Sanger et al. (1977), Proc. Natl. Acad. Sci. USA, 83, 3096-3100). The mutated gene encoding P225A mutant was a kind gift from 20 T. Graycar (Genencor, S. San Francisco, CA). It was synthesized by primer extension mutagenesis on a single stranded M13 subclone of Bacillus amyloliquefaciens subtilisin using the mutagenic oligonucleotides depicted in Fig. 13.

25 The above identified subtilisin variants were used in conjunction with the ligation substrates disclosed in Example 2 and Example 3 to provide the previously discussed results.

Example 2

Synthesis of FPTIPAAPF-glycolate-F-amide

The synthesis of a C-terminal amide ligation peptide is accomplished by attachment of the first Boc protected amino acid (Boc-Phenylalanine) to 4-methyl-benzhydrylamine resin (Bachem L.A.) using 35 diisopropylcarbodiimide in methylenechloride. Standard

Boc synthetic protocols for the synthesis of the protected peptide are followed. Barany, G., et al. (1979), in "Solid-Phase Peptide Synthesis", in The Peptides, Analysis, Synthesis, Biology. Special Methods in Peptide Synthesis, Part A. Vol. 2 (E. Gross, J. Meienhofer (eds), New York Academic Press), pp. 3-254. The glycolic acid residue was incorporated as the corresponding t-butyl ether. Removal of the t-butyl ether with 50% trifluoroacetic acid in methylene chloride and coupling of the subsequent amino acid with diisopropylcarbodiimide and 10 mol% dimethylaminopyridine in 90% methylene chloride, 10% dimethylacetamide, afforded the ester linkage. Subsequent amino acids were incorporated again using standard Boc protocols. The crude peptide was deprotected and removed from the resin with hydrogen fluoride. The crude peptide was then purified by reverse phase HPLC. Purity was determined by mass spectral analysis. $M+1$ calc. 1164.5, $M+1$ obs. 1164.6. Similar methods are used to produce other first ligation peptides.

Example 3

Growth Hormone Expression and Semi-Synthesis

The truncated form of human growth hormone (des-octa hGH containing residues 9-191) was expressed in E. coli W3110 (tonA; ATCC27325) using the E. coli alkaline phosphatase promoter and the signal peptide from E. coli heat-stable enterotoxin II (Chang et al. (1987), Gene, 55, 189-196). Cell paste was resuspended in four volumes of 10 mM Tris HCl, pH 8.0, to release the hGH from the periplasmic space. The cells were pelleted and the hGH was purified from the supernatant (Olson et al. (1981), Nature, 293, 408-411).

Peptides derived from the amino-terminal sequence of hGH having the sequence FPTIPLSR or FPTIPAAPF on their carboxyl-termini were esterified with the leaving group glycolate-Phe-amide. The reaction between des-octa hGH 5 (0.5 mM) and either of the two peptide substrates (2.4 mM final) was performed in Tricine (pH 8.0) at 20°C using either S221C/P225A or G166E/S221C/P225A subtilisin at final concentrations of 3.0 μ M or 3.4 μ M, respectively. The reaction was stopped by mixing an aliquot with an 10 equal volume of 100 mM NEM (which alkylates the active site cysteine). Loading buffer (5% 2-mercaptoethanol, 5% glycerol, 10 mM Tris HCl pH 8.0, 1 mM EDTA, 0.25% SDS, final concentrations) was added and the samples were boiled and analyzed by SDS-PAGE (Laemmli (1970), Nature, 15 227, 680-685). See Fig. 12. The ligation products were blotted onto polyvinylidene difluoride membranes (Matsudaira et al. (1987), J. Biol. Chem., 262, 10035-10038) and the amino-terminal sequences were determined.

Example 4

20 Modification of Protoprin with peptides

In addition to ligating peptides to form larger proteins, Protoprin (met-hGH) has also been modified with the S221C/P225A peptide ligase. One of the peptides used in the growth hormone semi-synthesis, FPTIPAAPF-glycolyl-F-amide, was used in these ligations. Protoprin (75 μ M) 25 was reacted with the peptide (350 μ M to 7 mM) in the presence of 3 μ M of the ligase in 115 μ l of the reaction buffer (10 mM Tricine, pH 8.0) at 25°C. The progress of the ligation reaction was monitored by mixing an 30 aliquot (15 μ l) with an equal volume of 100 mM NEM at time intervals between 1 minute and 1 hour. The samples were then run on a SDS-PAGE gel (Laemmli (1970), Nature 227, 680-685). Figure 14 shows that most of the substrate is ligated after an hour, as shown by the 35 appearance of a product of expected molecular weight.

Example 5Confirmation of ligase substrate specificity

To demonstrate that the ligase substrate specificity is similar to the substrate specificity for the hydrolysis of peptides for subtilisin, several different ligations were tried. The substrates used and the results obtained confirm that ligation is restricted to ligation substrates that have amino acid sequences similar to the sequence of subtilisin hydrolysis substrates.

10 The presence of proline at either the P1' or P2' position has been shown to inhibit subtilisin hydrolase activity. Carter et al. (1989), Proteins: Structure, Function and Genetics, Vol. 6., 240-248 (1989). hGH has a proline at the P2' position while the des-1 hGH has a proline 15 at the P1' position. Accordingly, substitution of Protropin with either hGH or des-1 hGH in the experimental conditions of Example 4 resulted in no ligation (data not shown).

This result was verified using insulin-like growth 20 factor-1 (IGF-1). The N-terminal sequence of IGF-1 is GPETLC. This proline at position P2' was expected to prevent ligation, similar to hGH. The substitution of IGF-1 for Protropin in the Example 4 experimental conditions resulted in no ligation (data not shown).

25 These results indicate that the N-terminal sequences of the peptides to be ligated are important, just as they are for the corresponding sequence in hydrolysis reactions with the wild-type subtilisin.

2087342

-44-

Example 6Specificity is further defined by the N-terminus

A further level of substrate specificity is obtained due to the importance of conformation of the N-terminus of the substrate. In addition to the amino acid sequence, the N-terminal structure of one of the pair of a prospective substrates is important. Subtilisin has been shown to prefer an extended conformation of the amino acids at the cleavage site. Several acceptor proteins that possess favorable amino acids sequence failed to be ligated or were ligated in only modest amounts, or required special treatment for ligation to occur, suggesting that these substrates lack the necessary extended conformation at the N-terminus.

The potential importance of conformation and N-terminal flexibility was demonstrated using IGF-1. Two modified forms of IGF-1 were used. The IGF-1 from brain has 3 residues from the amino terminus removed. This form of the protein, des-3 IGF-1, no longer has a proline at the P2' position. Under Example 4 conditions, no ligation of either peptide was seen to des-3 IGF-1 under these non-reducing conditions. However, by pretreating the substrate with reducing agent (10 mM DTT) for 1 hour, ligation was observed (5%). Another peptide, Long Arg-3 IGF-1, has the first 13 amino acid residues of porcine growth hormone (MFPAMPLSSLFVN) attached to the Glu3-Arg variant of IGF-1. The high N-terminal sequence homology of this protein to Protropin would suggest it can be used as a substrate. Under Example 4 conditions with the addition of pretreatment with 10 mM DTT for an hour, some ligation (\approx 5%) was observed. By adding 0.1% SDS and 10% DMSO, and carrying out the reaction at 50°C, the amount of product was increased 2 to 3-fold (data not shown).

61051-2487

-45-

Using the hGH receptor in place of Protoprin under Example 4 experimental conditions resulted in no ligation, even after 5 hours with a 4 fold higher concentration of the ligase (data not shown). While the 5 hGH receptor has the N-terminal sequence PSGSEAT and thus the sequence conforms to acceptable subtilisin substrate primary sequence, the lack of ligation is attributed to either the lack of accessibility of the N-terminus or a highly structured N-terminus.

10 A similar result was obtained for Relaxin. Relaxin consists of two chains, A and B. The A chain has a pyroglutamate at the N-terminus, so the α -amine is incapable of acting as a nucleophile. In contrast, the B-chain has the sequence DSWM and should be a suitable 15 substrate in the presence of high salt. However, modification of the Example 4 experimental conditions to include 2 M NaCl in the buffer solution resulted in no ligation after an hour (data not shown).

These results demonstrate that by judicious selection 20 of reaction conditions (including temperature, detergents, pH and non-aqueous solvents), it should be possible to selectively alter the conformation of the N-terminal segment of the substrate protein while retaining the ligase activity, thus allowing highly 25 specific ligations to occur.

Having described the preferred embodiments of the present invention, it will appear to those of ordinary skill in the art that various modifications may be made and that such modifications are intended to be within the scope 30 of the present invention.

61051-2487

-46-

CLAIMS:

1. A subtilisin-type serine protease variant having an amino acid sequence not found in nature and which is derived from a precursor subtilisin-type serine protease 5 having an α -helix containing proline at a residue equivalent to proline 225 in *Bacillus amyloliquefaciens* subtilisin and a catalytic serine equivalent to serine 221 in *Bacillus amyloliquefaciens* subtilisin, said derivation comprising:

10 a) replacing said catalytic serine with a first amino acid having a different nucleophilic side chain or modifying the side chain of said catalytic serine, to substitute the nucleophilic oxygen of said side chain with a nucleophilic sulfur, thereby converting said catalytic serine to cysteine, and

15 b) replacing said proline with a second different amino acid comprising a helix-forming amino acid or an amino acid with a side chain volume less than the side chain volume of proline;

20 said subtilisin-type serine protease variant being characterized by having peptide ligase activity in aqueous solution which is greater than that of said precursor subtilisin-type serine protease variant containing only said substitution or modification of said nucleophilic oxygen.

25 2. The subtilisin-type serine protease variant of Claim 1 wherein said precursor subtilisin-type serine protease is subtilisin.

3. The subtilisin-type serine protease variant of Claim 1 wherein said catalytic serine is replaced by cysteine.

61051-2487

-47-

4. The subtilisin-type serine protease variant of
Claim 1 further comprising replacement of at least one other
amino acid residue in said amino acid sequence.
5. The subtilisin-type serine protease variant of
Claim 4 wherein said replacement of at least one other amino
acid residue modifies a property selected from the group
consisting of substrate specificity, oxidative stability,
thermal stability and pH activity profile.
6. The subtilisin-type serine protease variant of
Claim 5 wherein said property is substrate specificity.
7. The subtilisin-type serine protease variant of
Claim 6 wherein said at least one other amino acid residue
comprises glutamate 156 or glycine 166 of the amino acid
sequence of *Bacillus amyloliquefaciens* subtilisin or
equivalent amino acid residues in other precursor
subtilisin-type serine proteases.
8. The subtilisin-type serine protease variant of
Claim 7 wherein said glutamate 156 is replaced by glutamine
and said glycine 166 is replaced by lysine.
- 20 9. The subtilisin-type serine protease variant of
Claim 7 wherein said glycine 166 is replaced by glutamate.
10. The subtilisin-type serine protease variant of
Claim 7 wherein said glycine 166 is replaced by isoleucine.
11. A ligation method comprising the step of
25 contacting the subtilisin-type serine protease variant of
Claim 1 with a first and a second ligation substrate, said
first ligation substrate comprising at least a first amino
acid residue (R1 residue), the carboxy-terminus of said R1
residue being esterified with an organic alcohols or thiols

61051-2487

-48-

wherein the organic moiety is selected from the group consisting of

C_6-C_{12} aryl where the aryl group is unsubstituted or substituted by one or more of the groups nitro, hydroxy, 5 halo (F, Cl, Br, I), C_1-C_8 alkyl, halo- C_1-C_8 alkyl, C_1-C_8 -alkoxy, amino, phenoxy, phenyl, acetamido, benzamido, di- C_1-C_8 alkylamino, C_1-C_8 alkylamino, C_6-C_{12} aroyl, C_1-C_8 alkanoyl, and hydroxy- C_1-C_8 alkyl,

C_1-C_{12} alkyl either substituted or unsubstituted, 10 branched, straight chain or cyclo where the substituents are selected from halo (F, Cl, Br, I),

C_1-C_8 alkoxy,

C_6-C_{12} aryloxy where the aryl group is unsubstituted or substituted by one or more of the groups 15 nitro, hydroxy, halo (F, Cl, Br, I), C_1-C_8 alkyl, C_1-C_8 -alkoxy, amino, phenoxy, acetamido, benzamido, di- C_1-C_8 alkylamino, C_1-C_8 alkylamino, C_6-C_{12} aroyl, and C_1-C_8 alkanoyl,

isothioureido,

C_3-C_7 cycloalkyl,

20 ureido,

amino,

C_1-C_8 alkylamino,

di- C_1-C_8 alkylamino,

hydroxy,

25 amino- C_2-C_8 alkylthio,

amino- C_2-C_8 alkoxy,

61051-2487

-49-

acetamido,

benzamido wherein the phenyl ring is unsubstituted or substituted by one or more of the groups nitro, hydroxy, halo (F, Cl, Br, I), C₁-C₈ alkyl, C₁-C₈-alkoxy, amino, 5 phenoxy, acetamido, benzamido, di-C₁-C₈ alkylamino, C₁-C₈ alkylamino, C₆-C₁₂ aroyl, C₁-C₈ alkanoyl,

10 C₆-C₁₂ arylamino wherein the aryl group is unsubstituted or substituted by one or more of the groups nitro, hydroxy, halo, C₁-C₈ alkyl, C₁-C₈-alkoxy, amino, phenoxy, acetamido, benzamido, di-C₁-C₈ alkylamino, C₁-C₈ alkylamino, C₆-C₁₂ aroyl, and C₁-C₈ alkanoyl,

guanidino,

phthalimido,

mercapto,

15 C₁-C₈ alkylthio,C₆-C₁₂ arylthio,

carboxy

carboxamide,

carbo-C₁-C₈ alkoxy, and

20 aromatic heterocycle wherein the heterocyclic groups have 5-10 ring atoms and contain up to two O, N, or S heteroatoms, and 2-hydroxy carboxylic acid and

25 wherein said second ligation substrate comprises at least a second amino acid residue (R1' residue), wherein the ligation product formed by said ligation contains the sequence R1-R1'.

61051-2487

-50-

12. The method of Claim 11 wherein said 2-hydroxy carboxylic acid is glycolate or lactate.

13. The method of Claim 12 wherein said 2-hydroxy carboxylic acid is glycolate.

5 14. The method of Claim 13 wherein said 2-hydroxy carboxylic acid is amidated.

15. The method of Claim 14 wherein said 2-hydroxy carboxylic acid is amidated with the amino-terminus of an amino acid or third peptide.

10 16. The method of Claim 11 wherein the R1 and R1' residues of each of said first and said second ligation substrates respectively has a length of between 1 and 25 residues and said ligation product contains about 50 residues.

15 17. The method of Claim 11 wherein said contacting is with the subtilisin variant of Claim 1 and said R1 amino acid residue of said first ligation peptide is a large hydrophobic amino acid selected from the group consisting of phenylalanine, tyrosine, tryptophan, leucine, and 20 methionine.

18. The method of Claim 17 wherein said contacting is with the subtilisin variant of Claim 1 and said R1' residue is selected from the group consisting of amino acid residues not consisting of isoleucine and proline.

25 19. The method of Claim 11 wherein said second ligation substrate comprises at least R1' and R2' residues and said R2' residue is selected from the group consisting of phenylalanine, tyrosine, tryptophan, leucine, methionine, arginine and lysine.

61051-2487

-51-

20. The method of Claim 11 wherein said contacting is under conditions that allow the C-terminus of said first ligation substrate and the N-terminus of said second ligation substrate to be in the correct conformation for 5 ligation to occur, without substantially destroying the ligation activity of said serine protease variant.

21. Composition comprising the subtilisin-type serine protease variant of Claim 1 and a ligation substrate comprising a peptide wherein the carboxy-terminus of said 10 peptide is esterified with a 2-hydroxy carboxylic acid.

22. DNA encoding the serine protease variant of Claim 1.

23. Expression vector containing the DNA of Claim 22.

24. Host cell transformed with the expression vector 15 of Claim 23.

SMART & BIGGAR

OTTAWA, CANADA

PATENT AGENTS

1/15

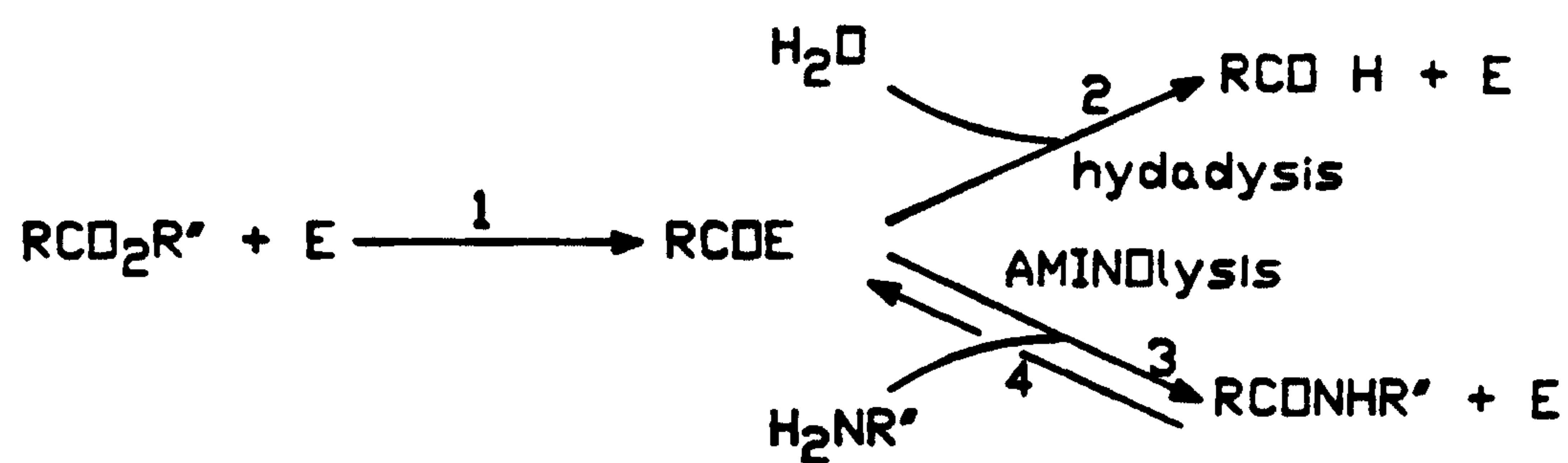


FIG.-1

2/15

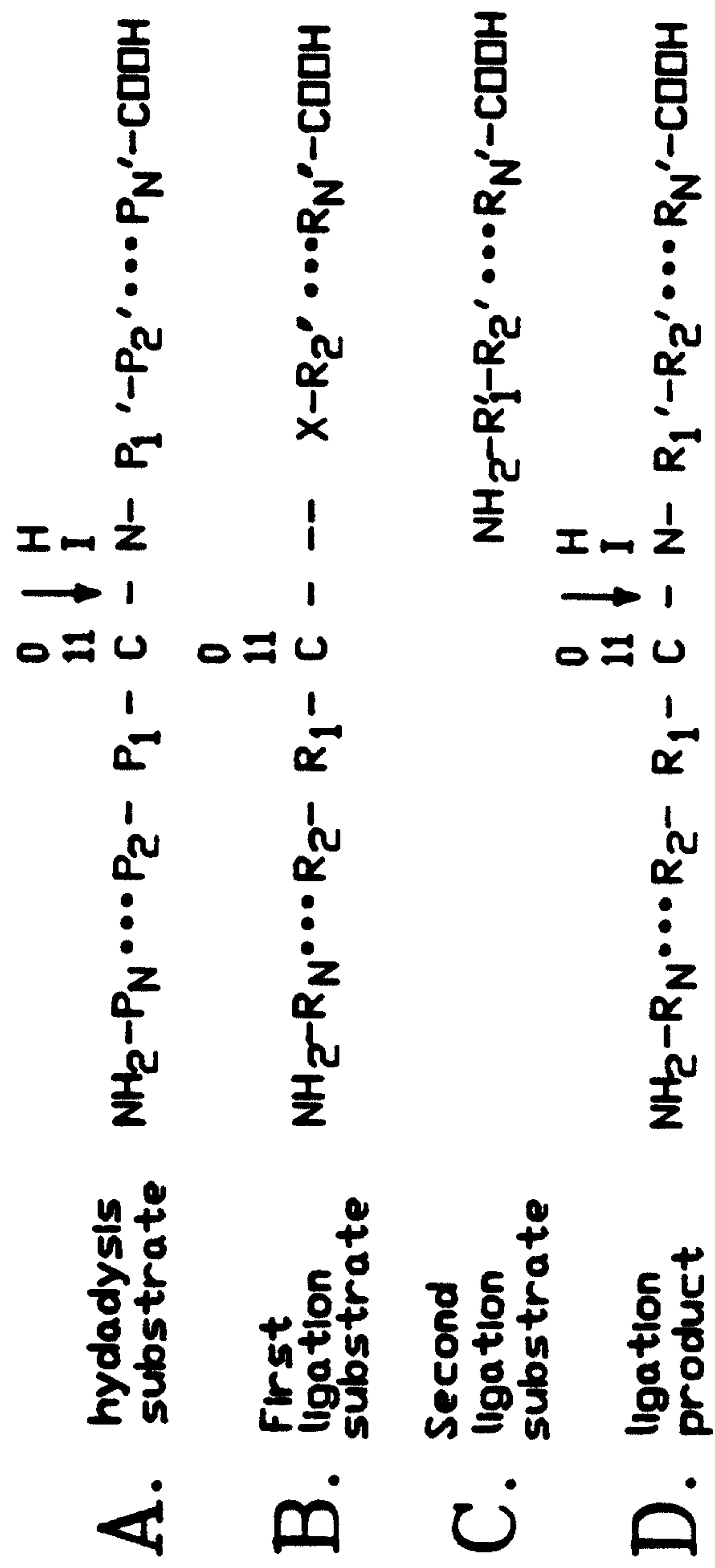


FIG. - 2

3/15

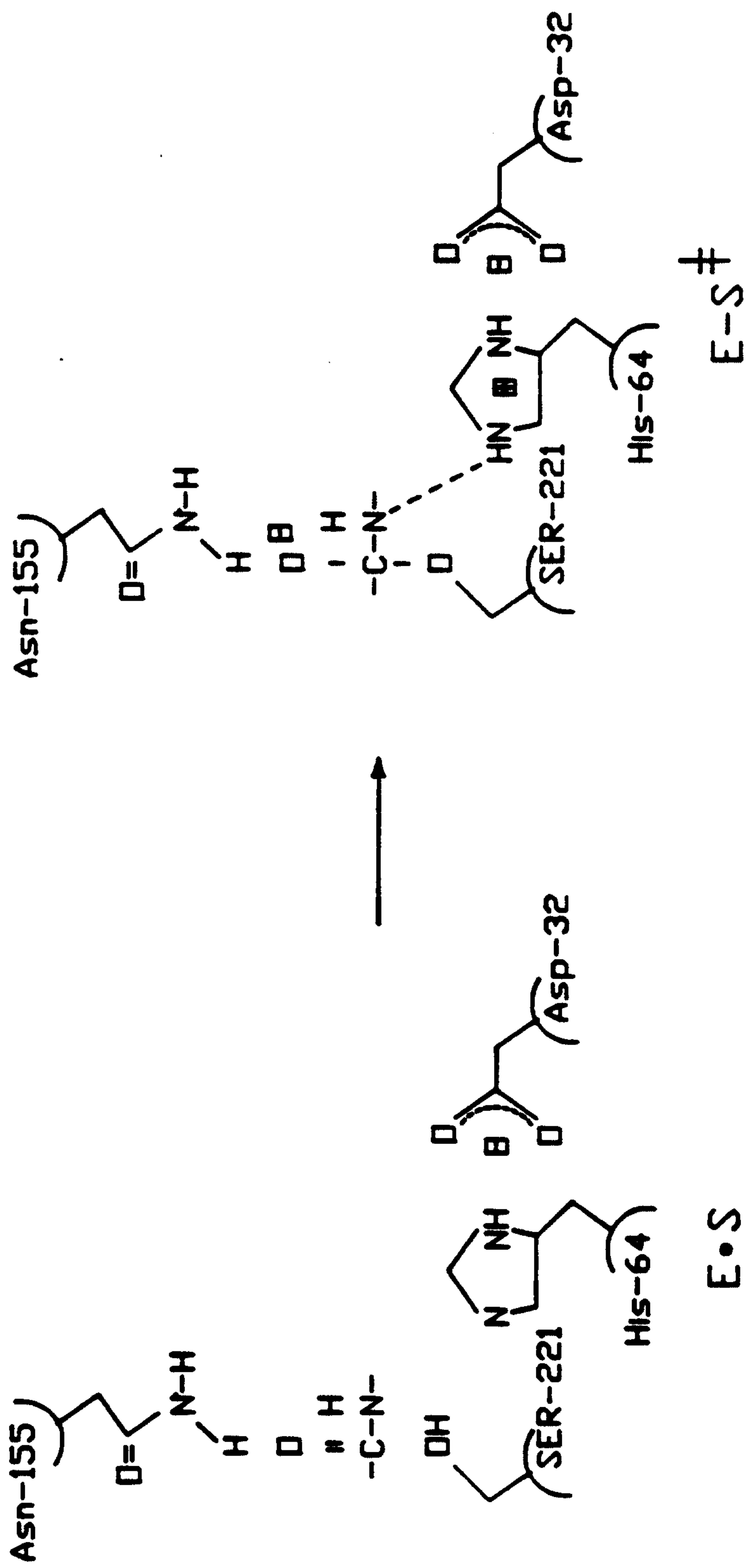


FIG. - 3

4/15

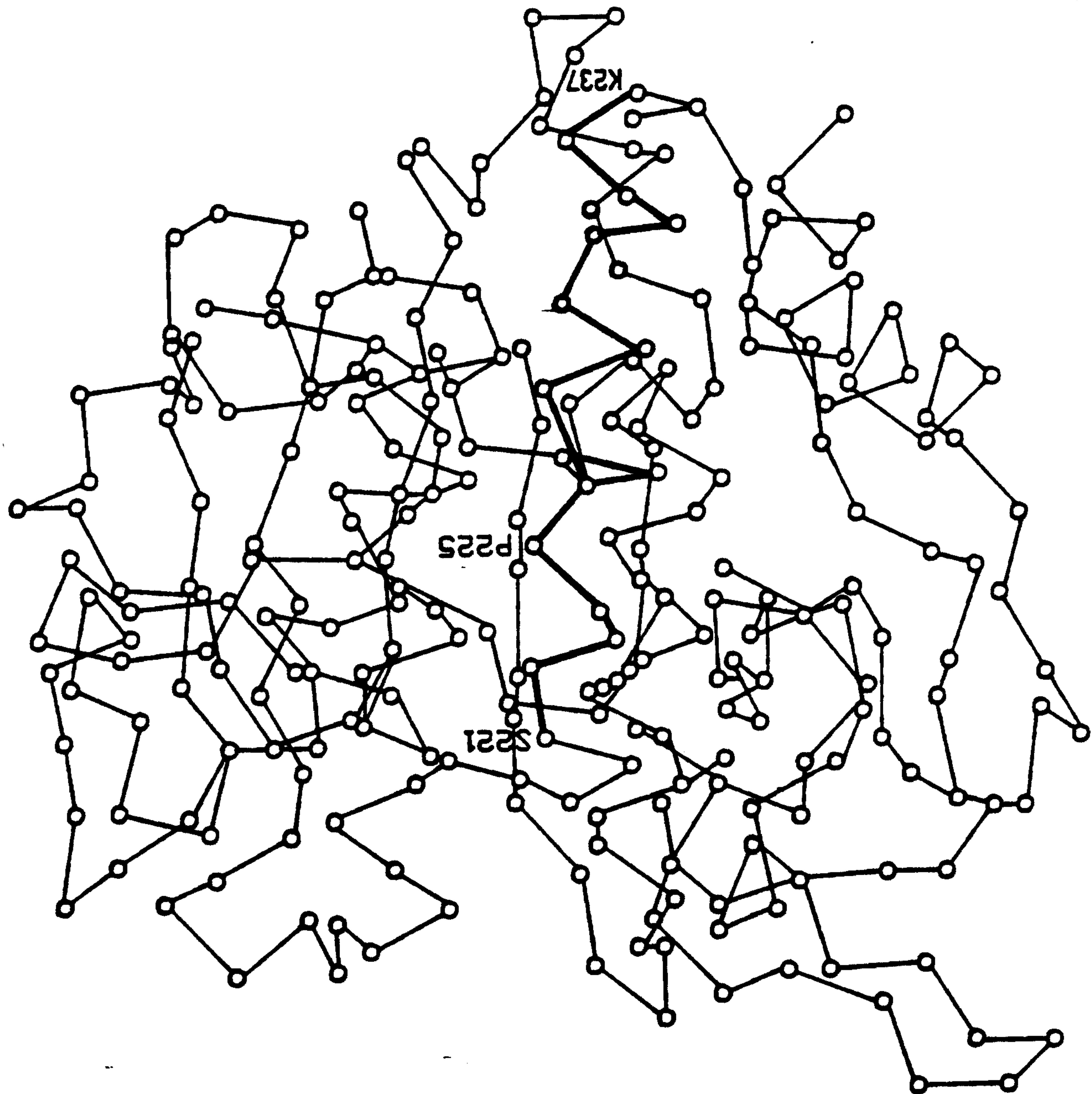
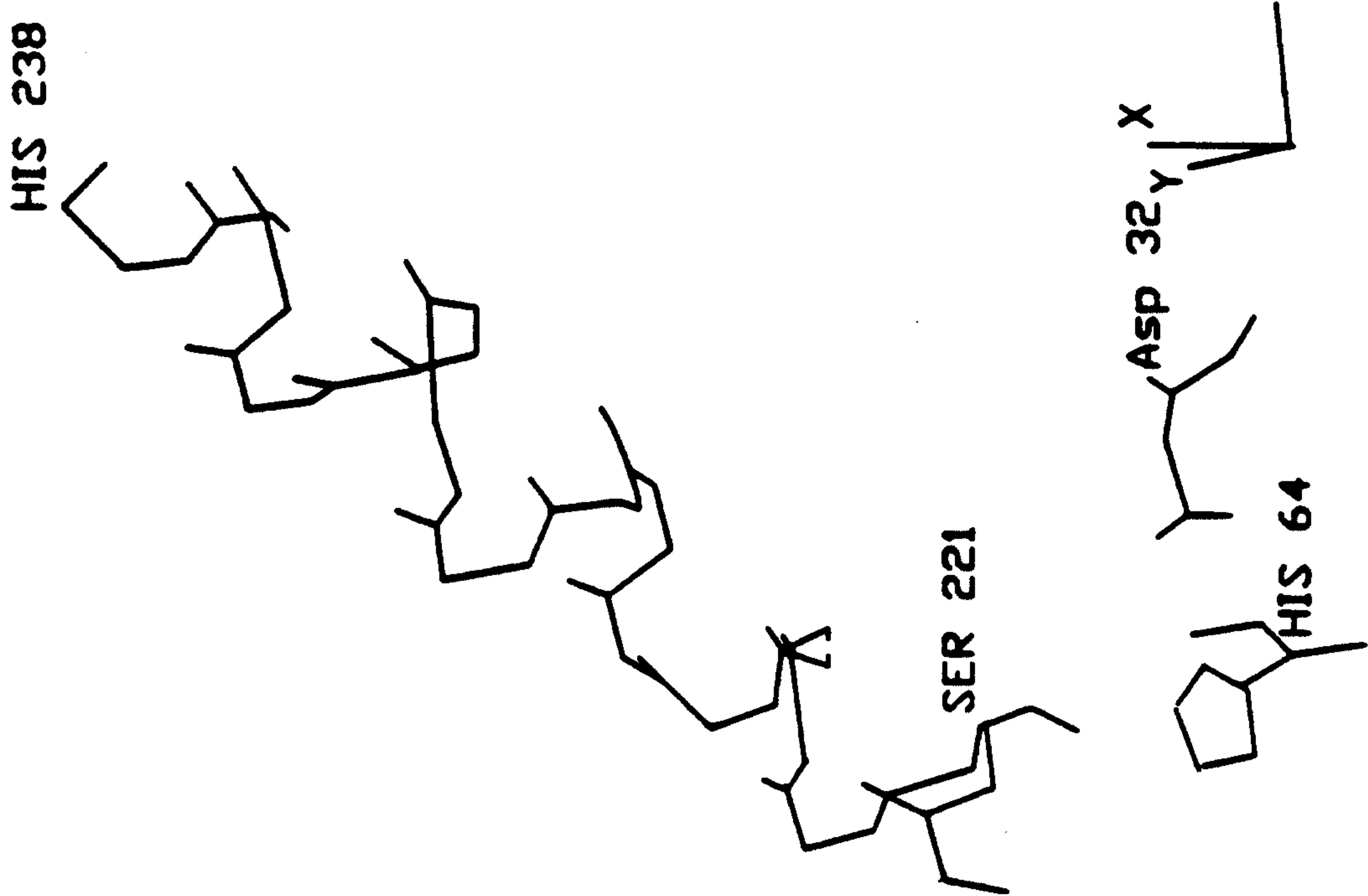
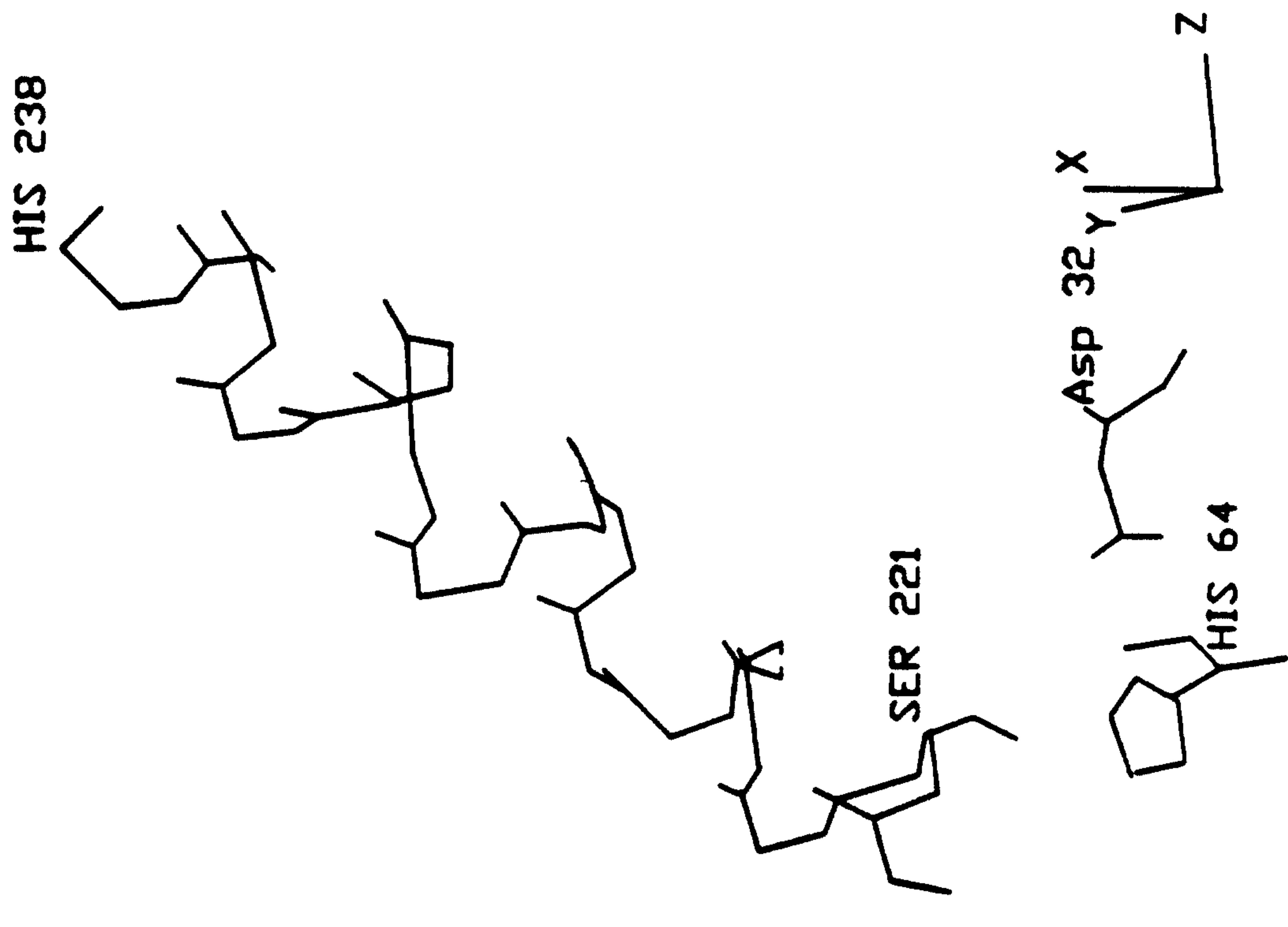


FIG. - 4

5/15



କାନ୍ତିକା

6/15

HOMOLOGY OF BACILLUS PROTEASES

1. BACILLUS AMYLOLIQUIFACIENS

2. BACILLUS SUBLILINARI 168

3. BACILLUS LICHENIFORMIS (CARLSBERGENSIS)

FIG.-6A

1	10	20
A	G S V P Y S V S Q I K A P A L N S Q G	
A	G S V P Y S I S Q I K A P A L N S Q G	
A	G T V P Y S I P L I K A D K V Q A Q G	
21	30	40
Y	T G S N V K V A V I D S G I D S S N P	
Y	T G S N V K V A V I D S G I D S S N P	
F	K G A N V K V A V L D T G I D A S N P	
41	50	60
D	L N V A G G G A S R V P S E T N P F Q D	
D	L N V R G G G A S F V P S E T N P Y Q D	
D	L N V V G G G A S F V A G E A Y N T Q D	
61	70	80
N	N S N G T N V A G T V A A L N N S I G	
G	S S N G T N V A G T I A A L N N S I G	
G	N G N G T N V A G T V A A L D N T T G	
81	90	100
V	L G V A P S A S L Y A V E V L G A D G	
V	L G V S P S A S L Y A V E V L D S T G	
V	L G V A P S V S L Y A V E V L N S S G	
101	110	120
S	G Q Y S N I I N G I E N A I A N N R D	
S	G Q Y S N I I N G I E N A I S N N R D	
S	G S Y S G I V S G I E N A T T N G R D	
121	130	140
V	I N R S L G G P S G S A A L K A A V D	
V	I N R S L G G P T G S T A L K T V V D	
V	I N R S L G G A S G S T A R K Q A V D	

7/15

141

150

160

K A V A S G V V V V A A A A G N E G T S G
 K A V S S G I V V V A A A A G N E G S S G
 N A Y A R G V V V V A A A A G N S G N S G

161

170

180

S S S T V G Y P G K Y P S V I A V G A V
 S T S T V G Y P A K Y P S T I A V G A V
 S T N T I G Y P A K Y D S V I A V G A V

181

190

200

D S S N Q R A S F S S V G P E L D V R A
 N S S N Q R A S F S S A G S E L D V R A
 D S N S N R A S F S S V G A E L E V R A

201

210

220

P G V S I Q S T L P G N K Y G A Y N G T
 P G V S I Q S T L P G G T Y G A Y N G T
 P G A S V Y Y S T Y P T N T Y A T L N G T

221

230

240

S R A S P N V A G A A A A L I L S K N P N
 S R A T P N V A G A A A A L I L S K N P T
 S R A S P N V A G A A A A L I L S K N P N

241

250

260

M T N T Q V R S S L E N T T T K L G D S
 M T N A Q V R D R L E S T A T Y L G N S S
 L S A G Q V R K R L S S T A T T L G S S

261

270

F Y Y G K G L I N V Q A A A A Q
 F Y Y G K G L I N V Q A A A A Q
 F Y Y G K G L I N V E A A A A Q

FIG.-6B

8/15

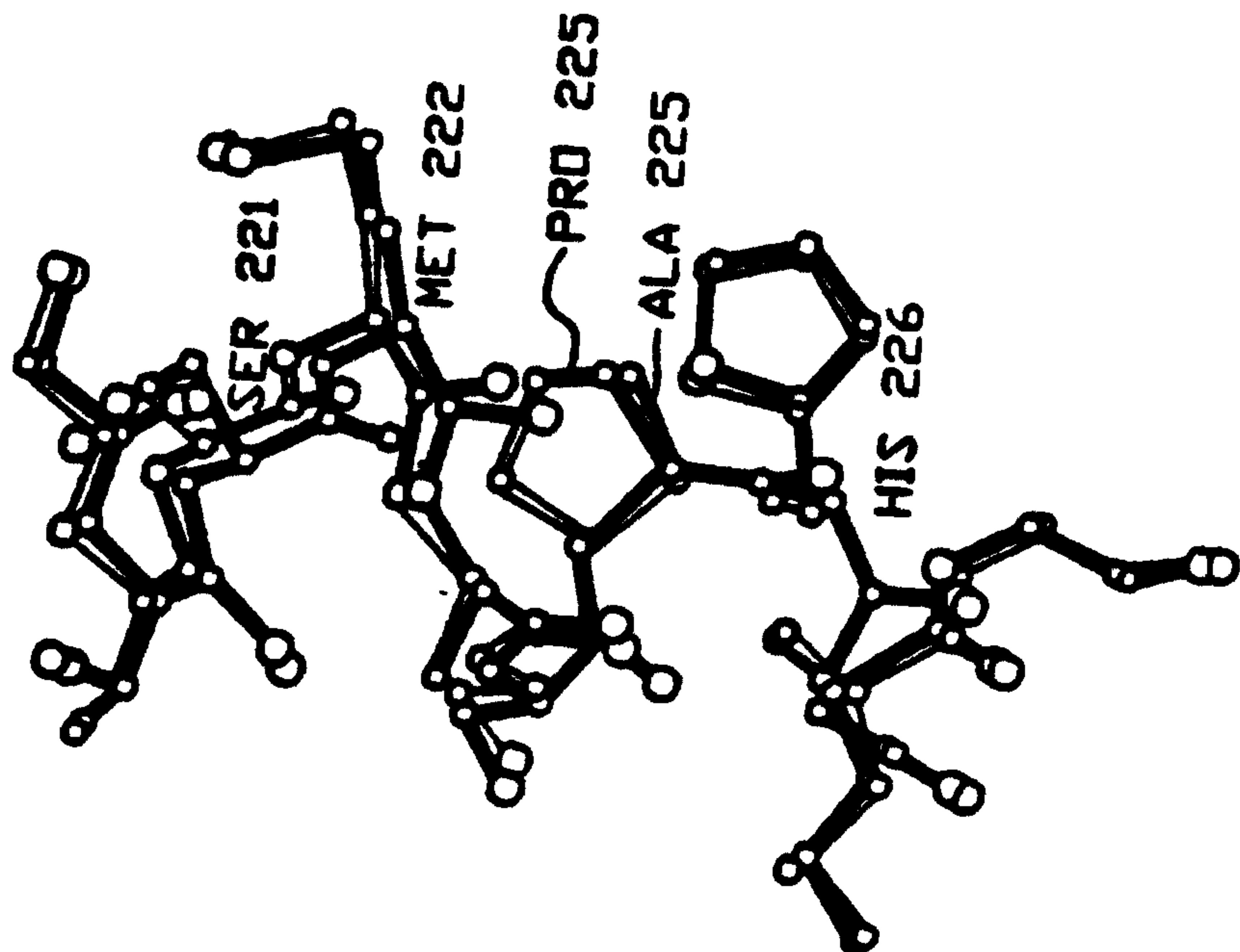
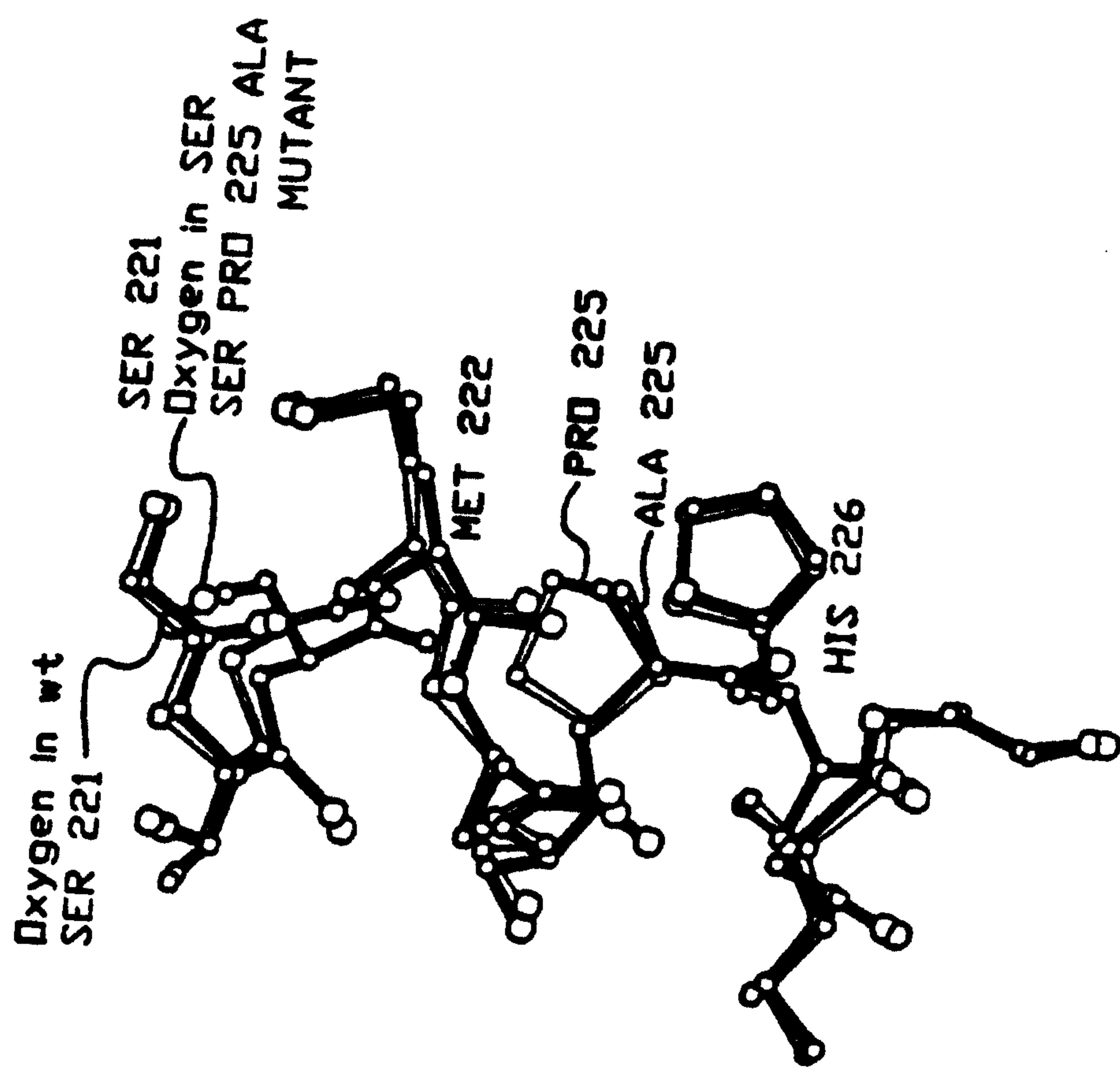


FIG. - 7

9/15

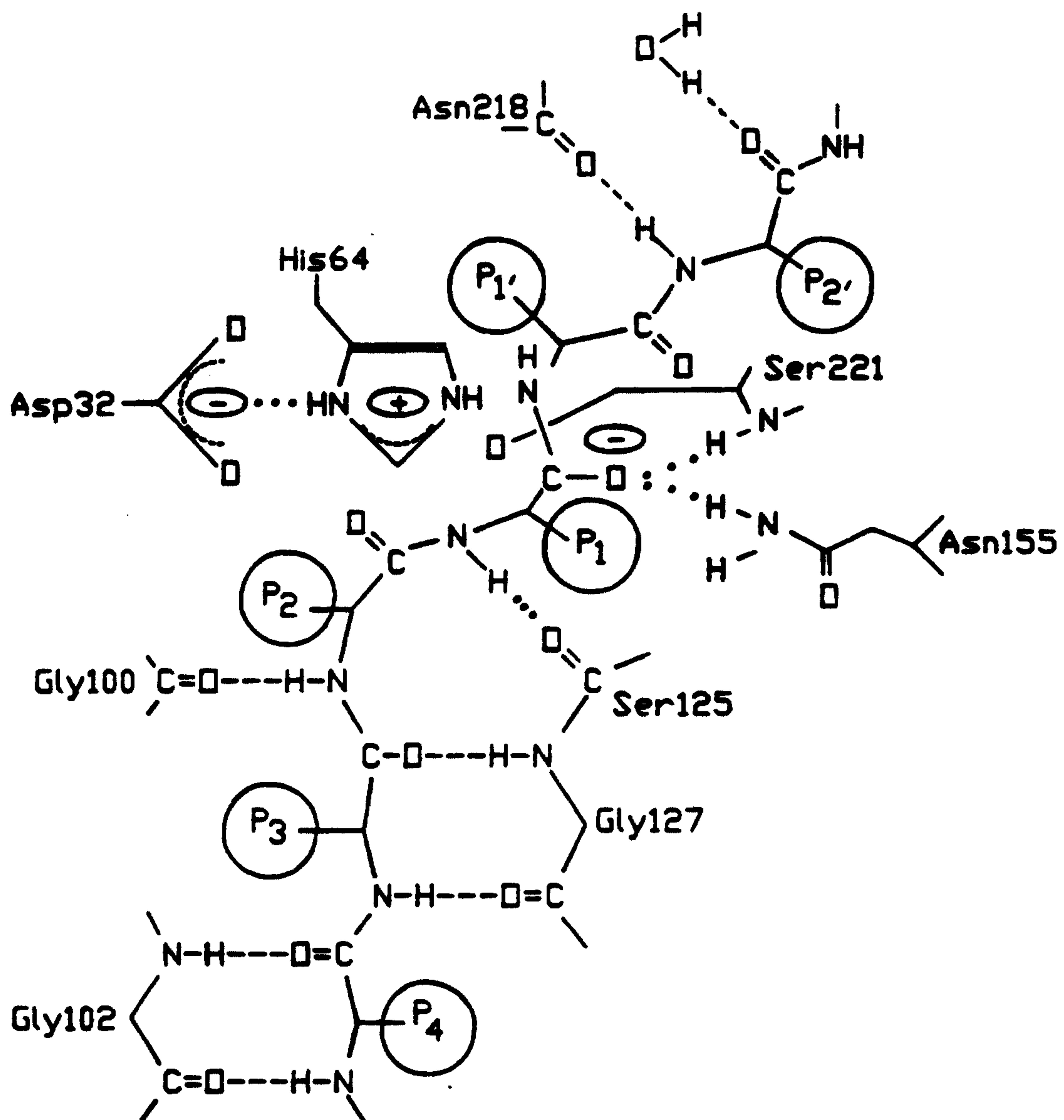


FIG. - 8

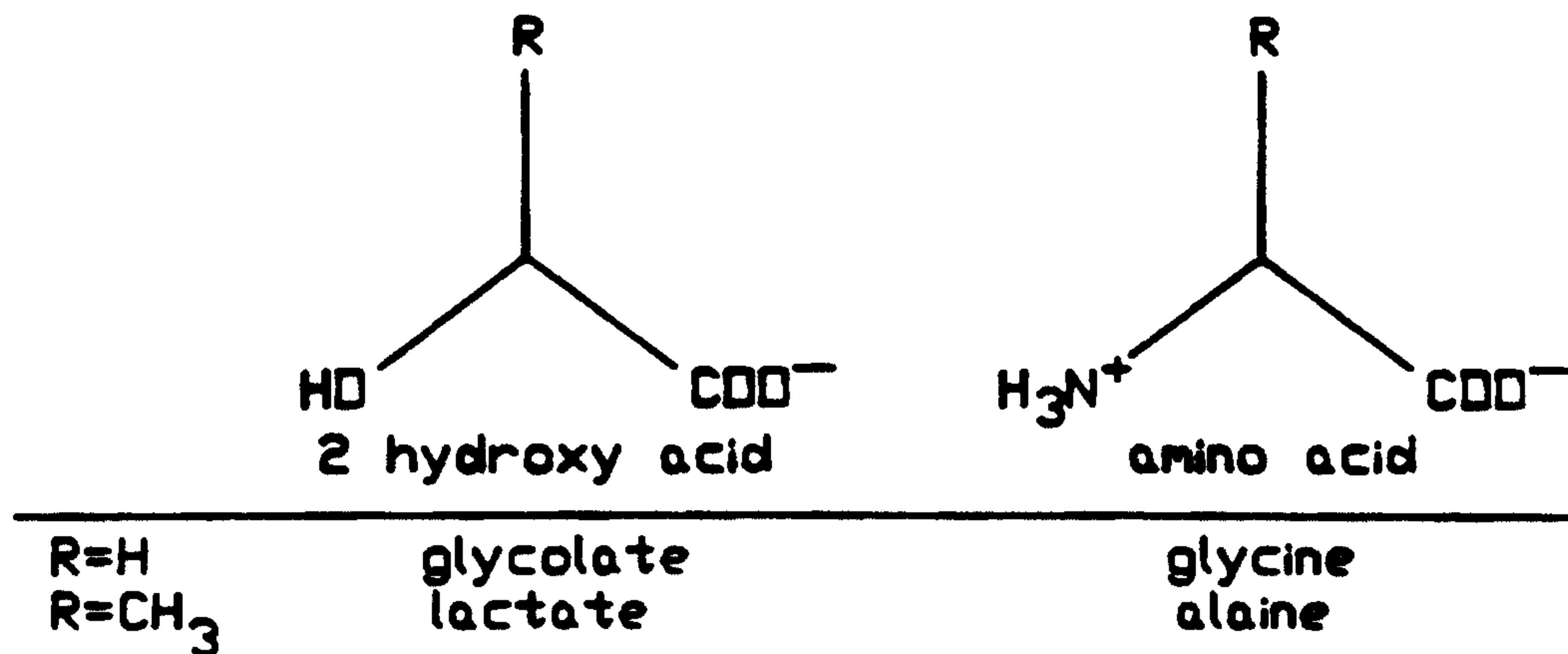


FIG. - 9

FIG.-10A

10/15

PRODUCTS

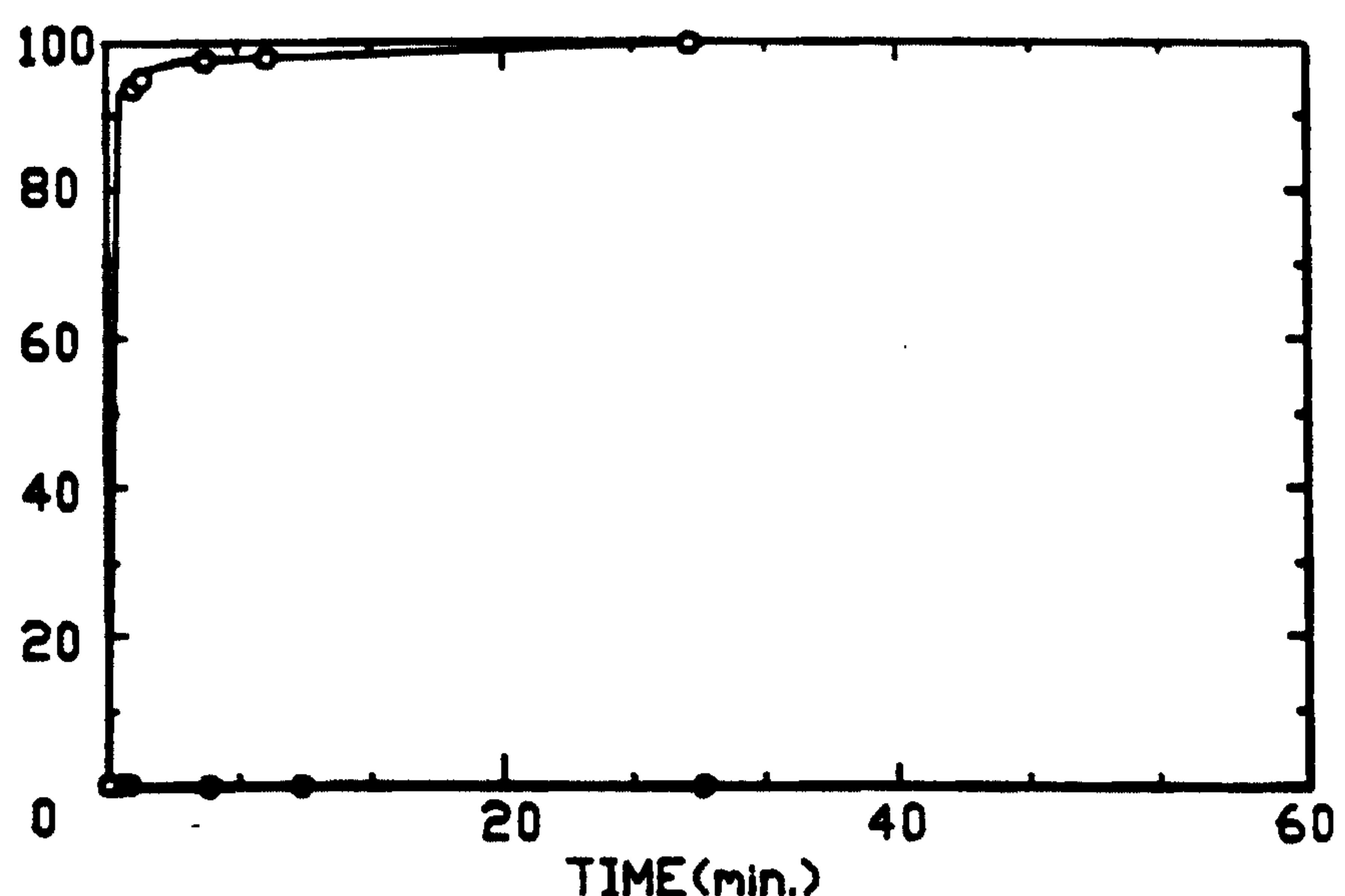


FIG.-10B

PRODUCTS

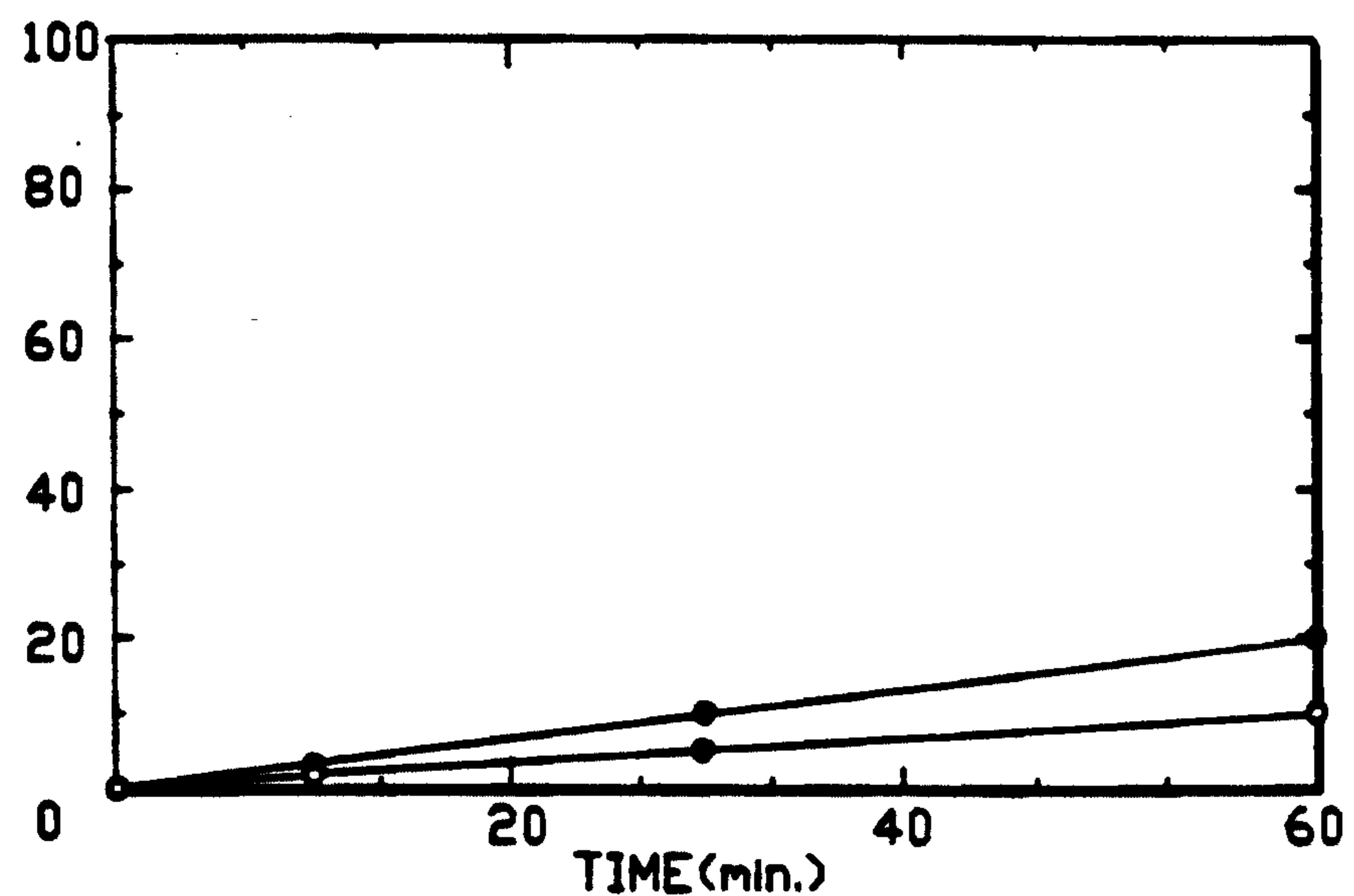
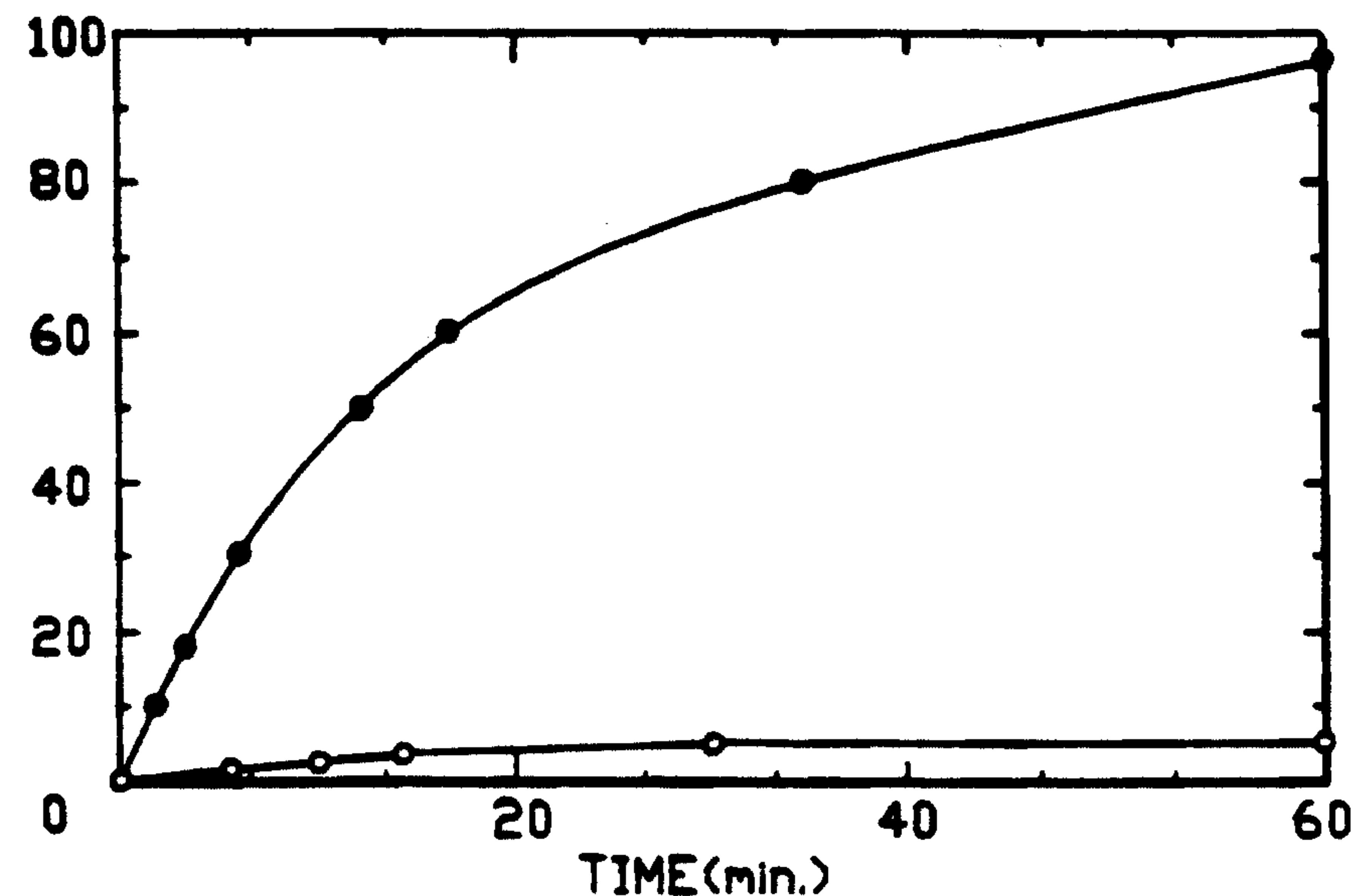


FIG.-10C

PRODUCTS



11/15

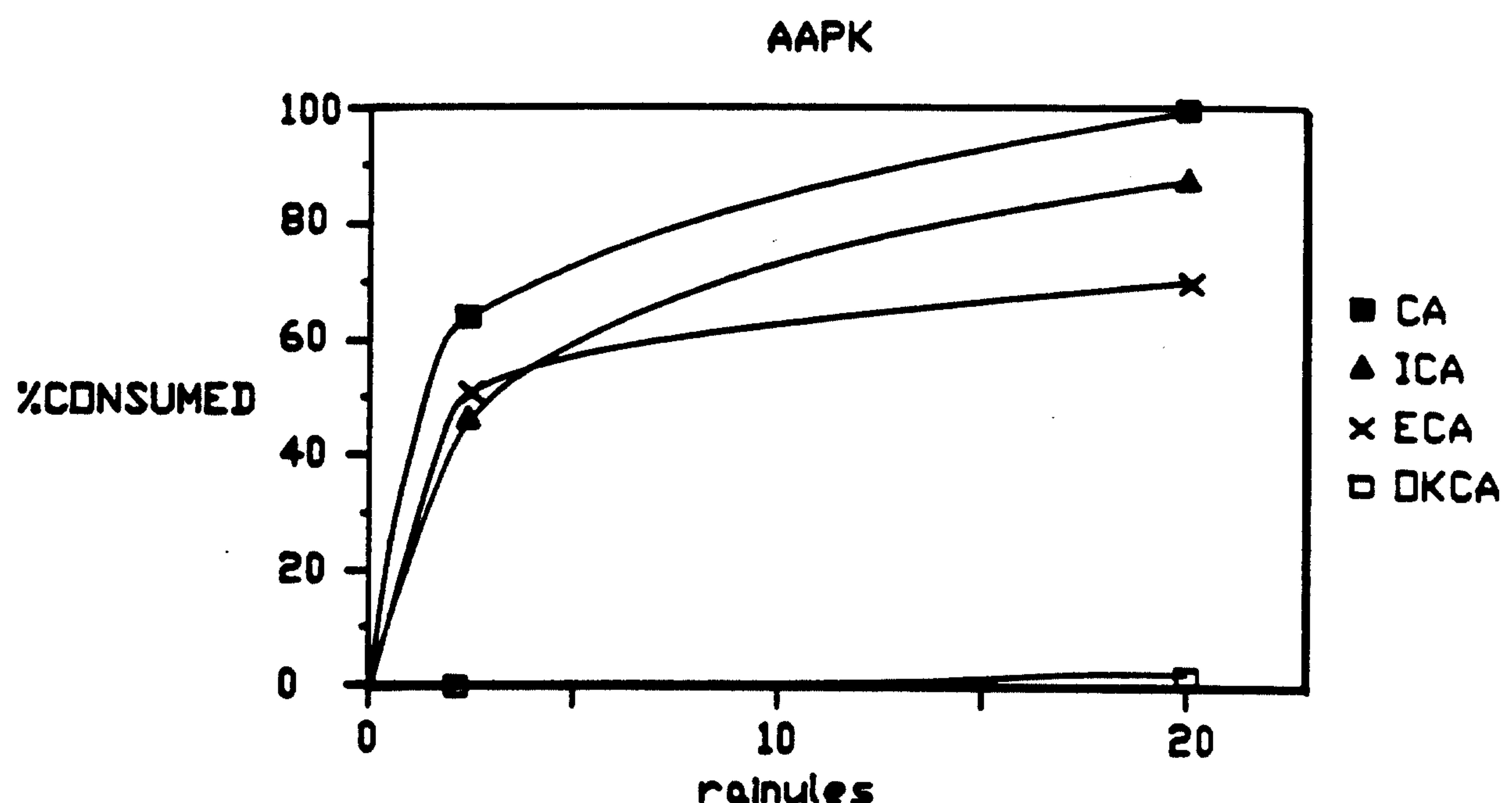


FIG.- 11A

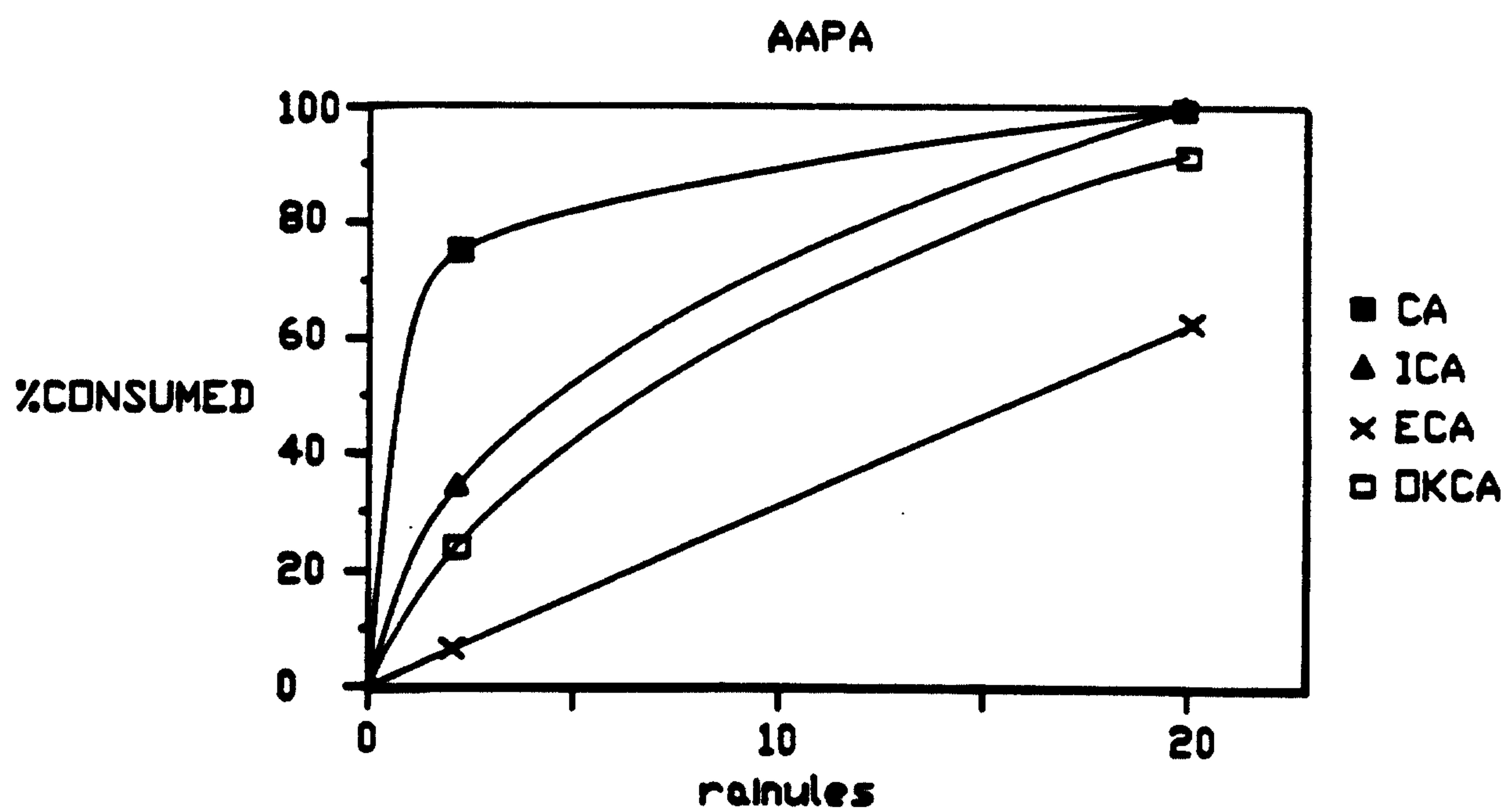


FIG.- 11B

12/15

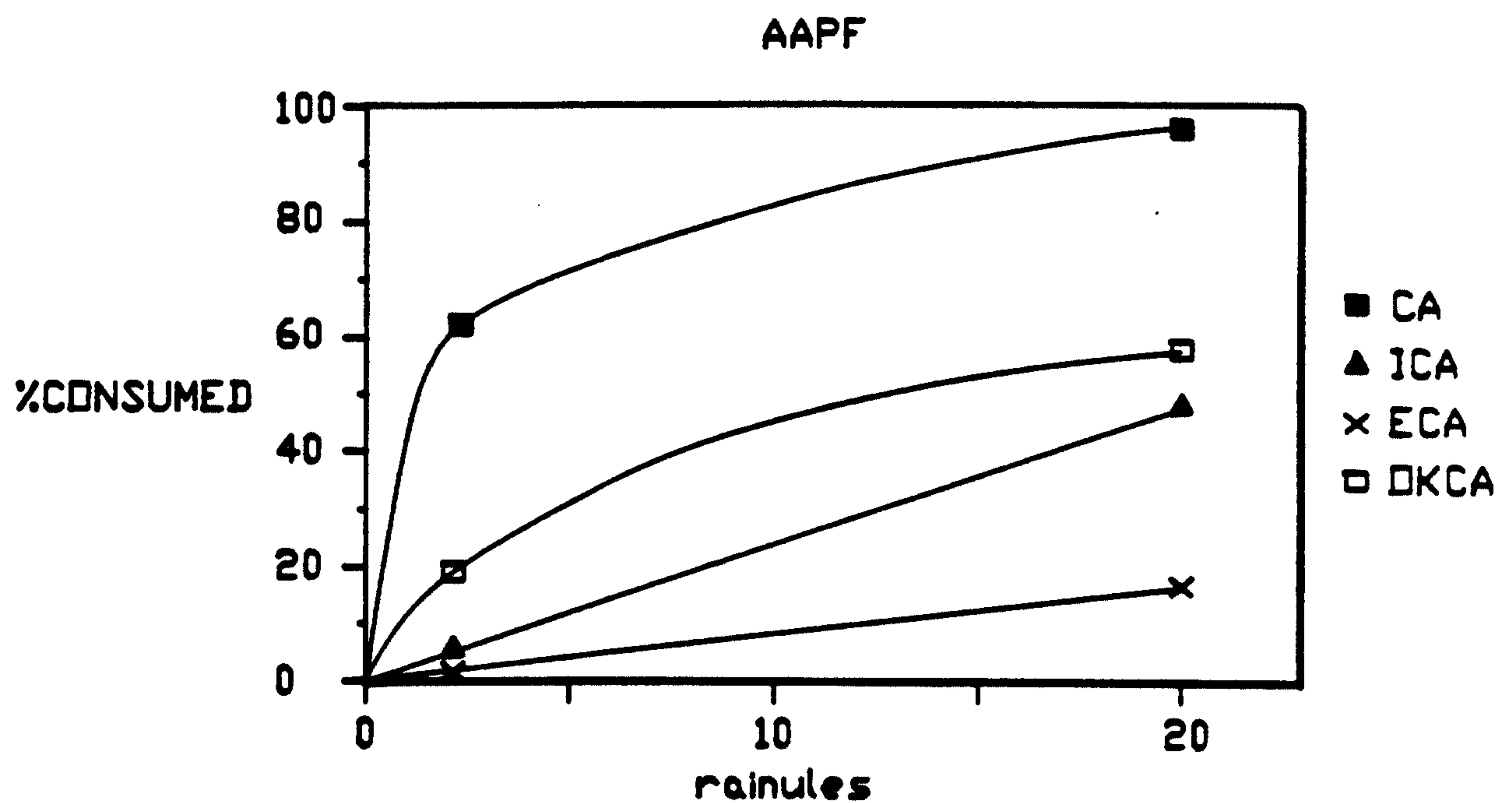


FIG.-11C

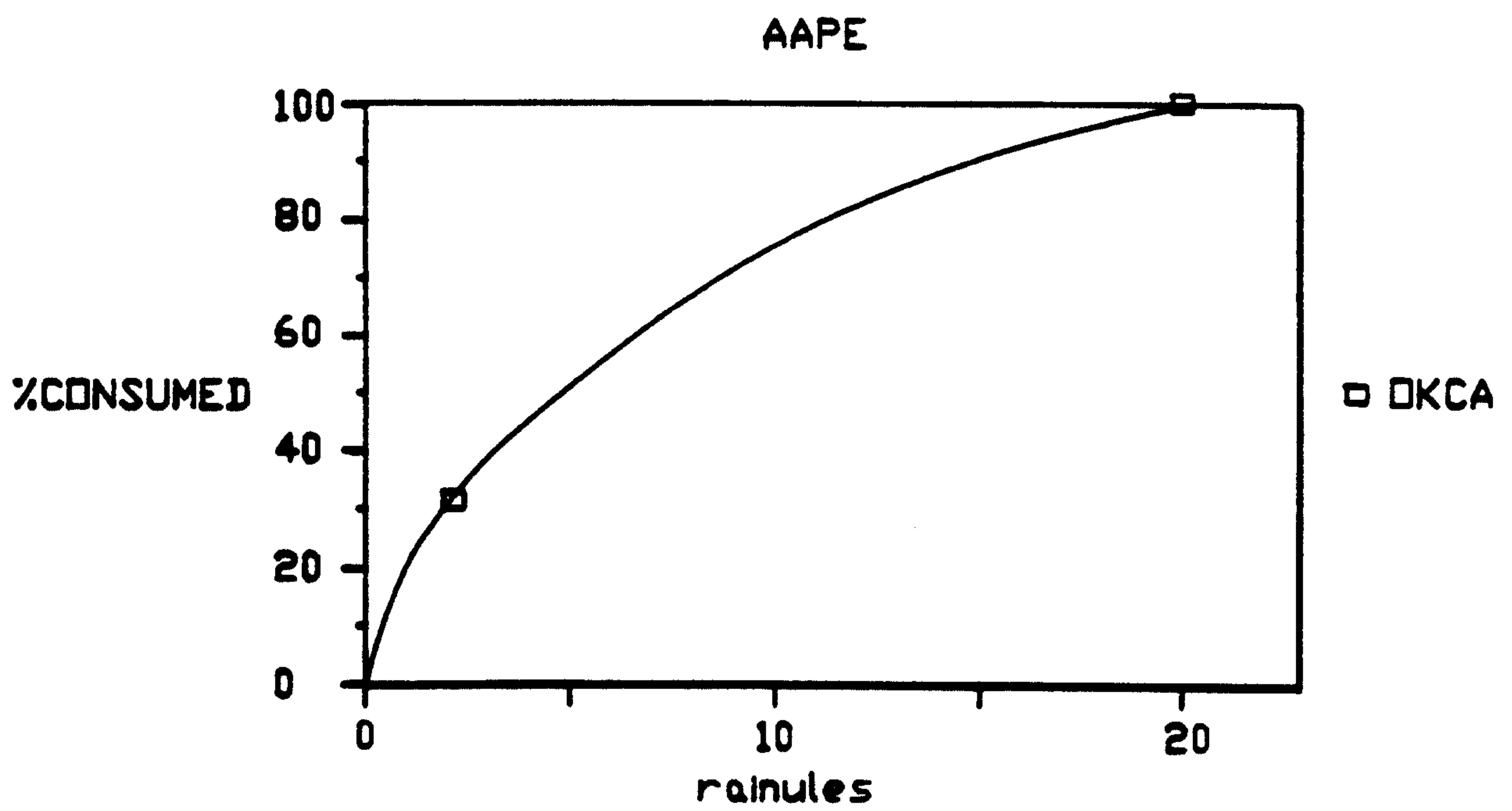


FIG.-11D

13/15

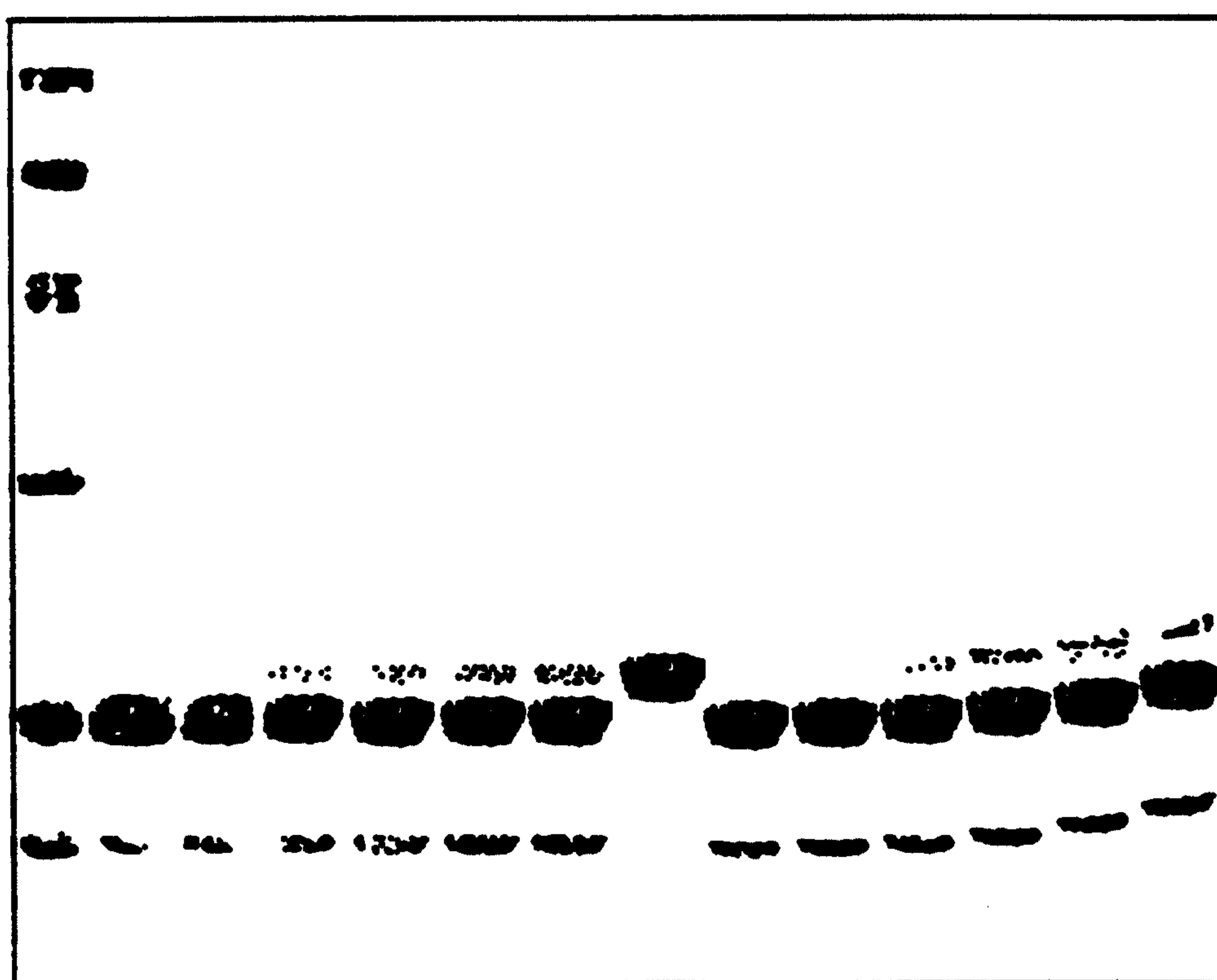
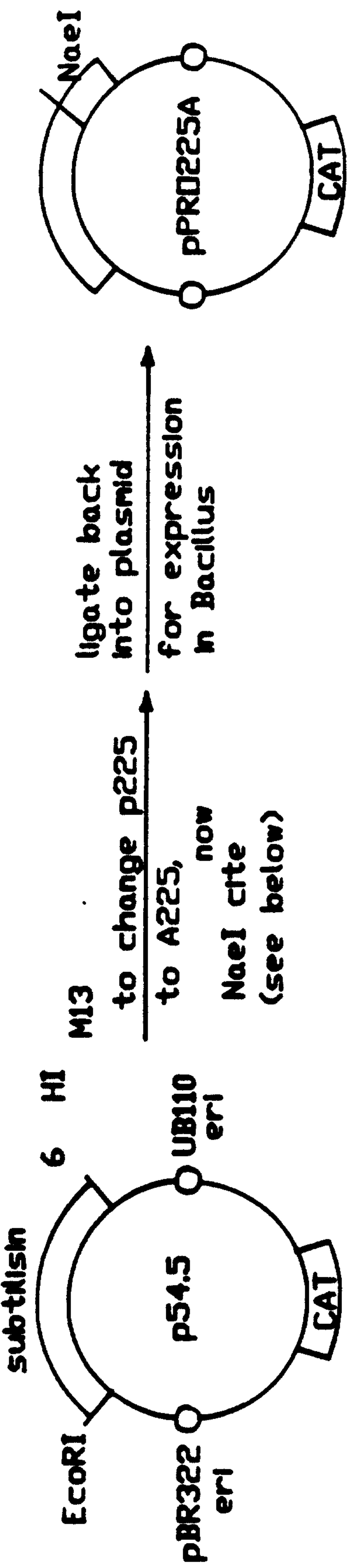


FIG.-12

14/15



W.T. acid sequence : 5'-TCA·ATG·GCA·TCT·CCG·CAC·GTT·GCC·GGA·GCG·GCT·GCT·

W.T. DNA sequence : Ser Met Ala Ser Pro His Val Ala Gly Ala Ala Ala
sequence 221 222 223 224 225 226 227 228 229 230 231 232

mutant A225 : 5'-TCA·ATG·GCA·TCT·GCC·CAC·GTT·GCC·GCG·GCT·GCT·
DNA sequence : Ser Met Ala Ser Pro His Val Ala Gly Ala Ala Ala
221 222 223 224 225 226 227 228 229 230 231 232

Asterisks denotes changes from wild type (W.T.) Barillus Angloliquifaciens subtilisin gene sequence. Underlined is a phenotypically silent Nael restriction site introduced to facilitate isolation of mutant plasmid from amongst a mixed W.T. and mutant population

FIG. - 13

15/15

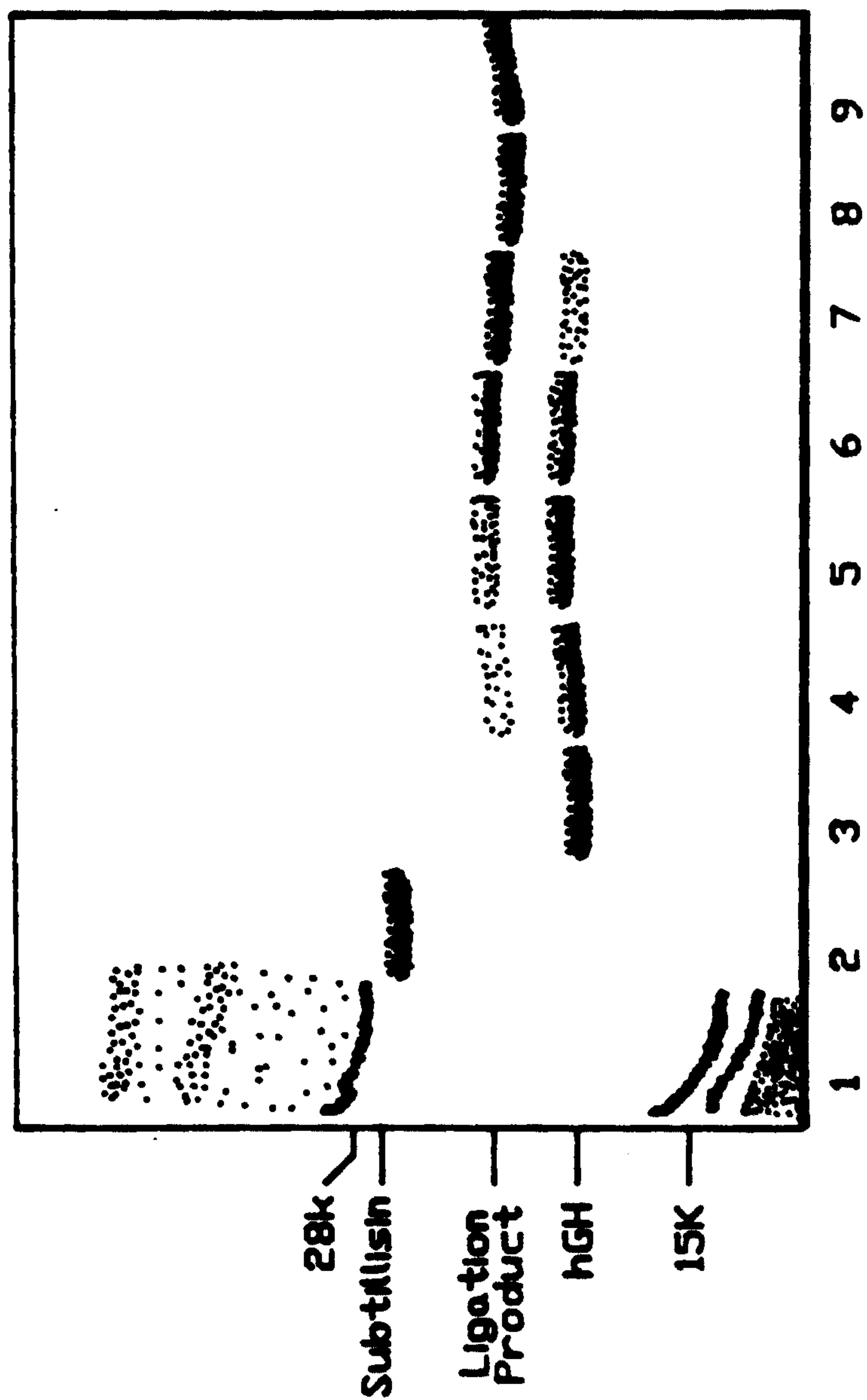


FIG. - 14