AMINO ACID SEQUENCES AND ACTIVE SITE FUNCTION IN YEAST HEXOKINASE B

A dissertation submitted to the University of London in candidature for the degree of Doctor of Philosophy

by

Andrew James William GRAY

August 1984

Department of Biochemistry
Imperial College
LONDON SW7 2AZ

ABSTRACT

The primary structure of yeast hexokinase B was investigated chemically to evaluate the utility and accuracy the primary structure determined by X-ray analysis, which has been termed the 'X-ray sequence', and to characterise the active site

Hexokinase purified by a scaled-up laboratory procedure was cleaved chemically with cyanogen bromide and hydroxylamine. The peptides derived were purified by chromatographic methods and were subjected to automated sequence determination. Approximately one half of the sequence was thus obtained. Confirmation of sequences was produced by sequencing by mass spectrometry peptide mixtures derived from a digestion with elastase.

Sequences were matched with the X-ray sequence by maximising the number of common residues and optimising the steric resemblance of the two sequences. The accuracy of the X-ray sequence varies along the polypeptide chain. It is about 30% correct in its allocations, less than the figure claimed.

Most of the residues identified in substrate binding have been identified. Although in some cases the proposed identity was incorrect, in all cases it was consistent with the postulated role.

The prediction that residue 189, which acts as a general base catalyst in the forward direction of reaction, should be aspartic acid, has been found to be correct.

Two of the four cysteine residues of hexokinase were located. One is believed to be that essential for activity. This essential thiol was misidentified by the X-ray sequence,

which identified instead an asparagine.

Because of the poor accuracy of the X-ray sequence it is unlikely that this approach to sequence determination is very useful. However, fitting non-overlapping peptides to tertiary structures by matching gross structural features could be a useful means of ordering a sequence.

CONTENTS

Page

LIST OF FIGURES

Page

LIST OF COMMON ABBREVIATIONS

ADP	Adenosine-5'-0-diphosphate
AMP	Adenosine-5'-0-phosphate
ATP	Adenosine-5'-0-triphosphate
ATPase	Adenosine triphosphatase
8-BrATP	8-bromoadenosine-5'-0-triphosphate
CrATP	Chromium (III) adenosine-5'-0-triphosphate
Del	α-amino pentanoic acid
Eps	α-amino hexanoic acid
G	Glucose
G6P	Glucose-6'-0-phosphate
Gam	α-amino butanoic acid
HPLC	High Performance Liquid Chromatography
IU	International units of enzyme activity
K_m	Michaelis constant
MgADP	Magnesium (II) adenosine-5'-0-diphosphate
MgATP	Magnesium (II) adenosine-5'-0-triphosphate
NMR	Nuclear Magnetic Resonance Spectroscopy
ORD	Optical Rotatory Dispersion
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
Tris	Tris (hydroxymethyl) aminomethane
V_{max}	Maximum specific velocity in the forward direction of reaction

γ (Gam), δ (Del) and ε(Eps) are employed in peptide sequences to indicate residues that were not identifiable, but where sidechains correspond roughly in size to two, three or four carbon atoms respectively.

ACKNOWLEDGEMENTS

I would like to thank the following people for their invaluable assistance without which none of this work would have been possible:

Professor E.A. Barnard, who appointed me and who organised the project for ready discussion, advice, direction, help, information and supervision; for reading and criticising this manuscript, and for having enough faith in the project to keep it going.

Dr. Rudolf Shipolini, who supervised the bulk of the work described in this thesis, for teaching me how to use a protein sequencer and an HPLC; for tolerance and intolerance in fashioning some sort of protein chemist, and for much else besides.

Dr. Anne Dell, for teaching me the rudiments of protein chemistry, for supervision of all the mass-spectrometric work described, and for running my samples on the mass spectrometer.

The staff of the pilot plant of the Department of Biochemistry at Imperial College, and particularly Mr. Jim Allan, for help and facilities in purifying hexokinase.

Mr. David Featherbe for rapid and efficient amino acid analysis; Mr. Ian Blench for expert assistance with protein sequencing.

Miss Deanne Eastwood, for the speedy, accurate and efficient typing of this thesis.

Imperial College of Science & Technology, for the award of a bursary during the tenure of which much of this work was carried out.

Finally, I would like to thank those many members, past and present, of the Department of Biochemistry of Imperial College, whose help, encouragement, advice and, above all,

friendship, made doing this worthwhile.

CHAPTER 1

INTRODUCTION

1.1 Properties of hexokinases from yeast

Hexokinase (E.C.2.7.1.1) is one of a class of enzymes that catalysesthe transfer of phosphate groups from nucleoside triphosphates to sugars:

$$\text{Hexose} + \text{ATP} \longrightarrow \text{Hexose 6-phosphate} + \text{ADP}$$

This reaction is of considerable metabolic importance because it is by this procedure that glucose is made available both to the glycolytic pathway and to polysaccharide synthesis.

Yeast (Saccharomyces cerevisiae) has three distinguishable hexose-phosphorylating activities: two hexokinases, active towards both glucose and fructose, and a glucokinase inactive towards fructose. The hexokinases are not required for growth on glucose (Maitra, 1970) but are essential for growth on fructose. They represent the principal hexose-phosphorylating capacity in the cell, being several times more active than the glucokinase.

Hexokinase activity purified from baker's yeast is resolved chromatographically into two distinct forms, A and B (Lazarus et al, 1966; Ramel et al, 1971; Barnard, 1975) which are identical with the P_I and P_{II} forms of Schulze and Colowick (1969).

Hexokinases A and B are non-interconvertible isoenzymes. They differ in their specific activities towards glucose, 300 and 800 units/mg respectively; in their isoelectric points, 5.3 and 5.1 respectively; and in the ratios of their activities towards fructose and glucose, 2.0 and 1.1 respectively (Lazarus et al, 1966). They are antigenically distinct (Womack et al, 1973) although their amino acid compositions (Schmidt and

Colowick, 1973a) indicate a considerable degree of similarity: 27 of the 54 tryptic peptides are common to both forms.

The advantage to yeast of possessing two such activities is unclear. It is of some interest that in exponentially-growing S.cerevisiae the majority of hexose-phosphorylating activity is hexokinase B (Gancedo et al, 1977). A similar but less pronounced trend is also observed for S.carlsbergensis (Kopperschlager and Hoffman, 1969).

Both the hexokinase isoenzymes have a molecular weight of 104 000 (± 2000) and contain two identical subunits of molecular weight 52 000 (Rustum et al, 1971). The enzymes exist as dimers under conditions of low ionic strength, pH and glucose concentration. Increasing any of these factors induces dissociation (Derechin et al, 1972). The effect of glucose alone is slight but is greatly increased by the presence of MgADP.

1.2 The Structure of the Hexokinases

Hexokinase B can be crystallised in the dimeric form from either ammonium sulphate or ammonium phosphate solutions, and in the monomeric form from potassium phosphate solution. These forms are termed BI, BII and BIII respectively. The monomeric form was prepared from a form of hexokinase that had been slightly proteolysed during purification; it is known that such proteolysis removes the ability to dimerise in solution (Schulze and Colowick, 1969) without substantially affecting catalytic properties.

The two dimeric forms differ in the relative orientation of their subunits and in their abilities to bind sugar and nucleotide substrates. Whereas in the BI form, crystallised from ammonium sulphate solution, the two subunits are arranged

Fig 1.1

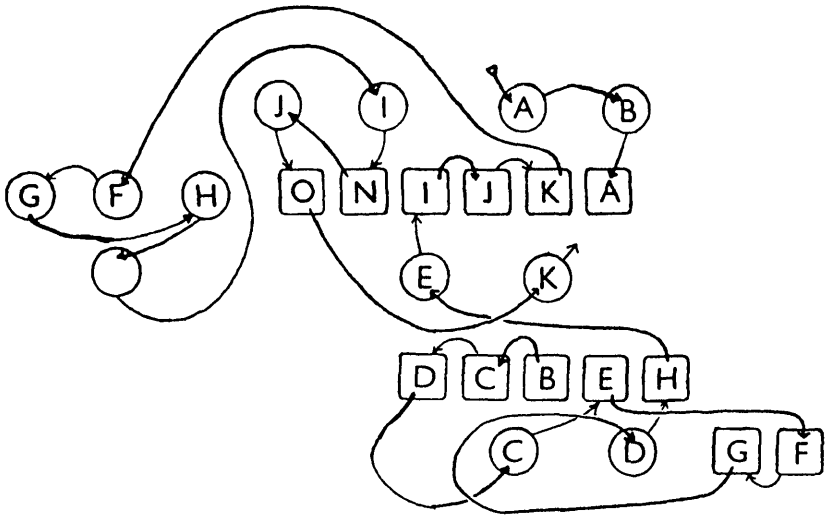
Amino acid compositions of Hexokinase isoenzymes A and B, and of the model of isoenzyme B derived from the X-ray sequence.

	Hexokinase A	Hexokinase B	X-ray sequence
Cys	4	4	5
Asx	56	53	32
Thr	30	29	14
Ser	23	25	48
Glx	52	54	22
Pro	25	29	17
Gly	42	40	35
Ala	32	33	59
Val	27	24	24
Met	11	11	8
Ile	30	36	32
Leu	49	35	27
Tyr	15	15	8
Phe	18	23	19
Lys	38	34	19
His	8	5	6
Arg	18	18	11
Trp	4	4	4
Gam			28
Del			33
Eps			17
	482	472	457

The N-terminal 11 residues of Hexokinase B are not present in the crystal form from which the X-ray sequence was derived, and are not included in the published composition of that sequence (Anderson et al, 1978b). The composition of the X-ray sequence above has been amended to take these residues into account.

symmetrically, related by a 180° rotation about and a 3Å translation along the dyad axis (Steitz, 1971; Steitz et al, 1973): in the BII form the subunits are arranged heterologously, and are related by a 156° rotation about, and a 13Å translation along, the dyad axis (Anderson et al, 1974; Anderson and Steitz, 1975; Steitz et al, 1976, 1977). The BI form will not bind substrates or products, whereas the BII form will. Such differences must arise from the mutual orientation of the subunits since the tertiary structure of the subunits is identical in all the crystalline forms studied.

The elucidation of the subunit tertiary structure has been extended from 7 (Anderson et al, 1974) to 2.1Å (Anderson et al, 1978b) and the path of the backbone of the polypeptide chain established. Each subunit has an elongated bilobal structure whose principal feature is a deep cleft that divides it into two domains, one having approximately twice the size of the other (ibid, passim). The lobes can be characterised by the arrangement of their secondary structure: the large lobe is of mixed all-α - α/β type and contains a six-stranded mixed β-sheet and eight α-helices; the small lobe is of α/β type and contains a five-strand β-sheet flanked on either side by two α-helices. The helix content of the molecule is 40-50%, which is consistent with the ORD measurements of Kenkare and Colowick (1965). The longest helix is 50Å in length and runs from one end of the large domain into the small one. The size of each subunit is 80 x 40 x 50Å.



1.2.1 The 'X-ray sequence' of hexokinase B

In an attempt to obtain amino acid sequence information from electron density data, the crystal structure determination was extended to 2.5Å resolution (Anderson et al, 1978 a,b). About 80% of the residues showed some side-chain electron density, which could in 30% of cases best be accounted for by a single amino acid. To sites possessing side-chain electron density not attributable to a specific residue, aliphatic side-chains of comparable length were allocated ('Gam', 'Del' and 'Eps'). This technique assigned 2909 of 3596 non-hydrogen atoms, and 455 out of 461 residues. This model is, however, biased towards small side-chains and is unable to extract sequence information from regions of low electron density.

The resolution was extended to 2.1Å by crystallographic refinement in an attempt to relieve the shortcomings of the model. Further non-hydrogen atoms could be added to the structure, the number of bond angles disallowed in protein conformation was reduced and the path of the polypeptide backbone altered slightly. Although this refined model accommodated two chemically-sequenced nonapeptides so well

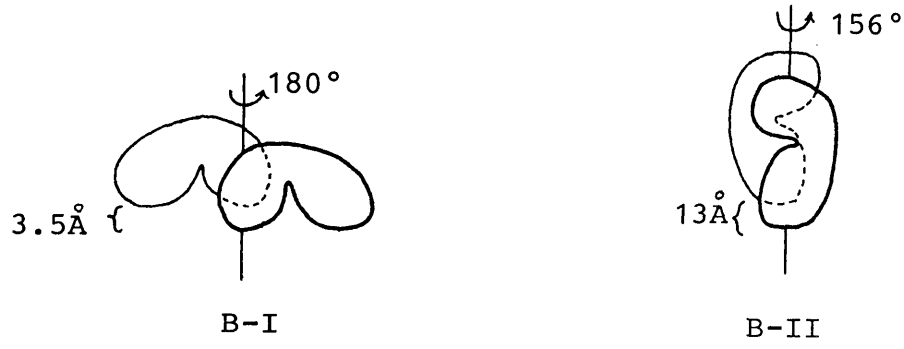
Fig 1.2: The 'X-ray sequence' of Yeast Hexokinase B (Anderson et al,1978b)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0		Ala	Ala	Ala	Gam	Asx	Del	Ser	Leu	Val	Glx	Val	His	Del	Eps	Val	Phe	Ile	Val	Pro
20	Pro	Del	Ile	Leu	Glx	Ala	Val	Val	Ser	Ile	Leu	Thr	Thr	Arg	Del	Asx	Asx	Gam	Asx	Ser
40	Ser	Ala	Ala	Ser	Ile	Pro	Met	Val	Pro	Gly	Trp	Val	Leu	Lys	Glx	Val	Ser	Gly	Ala	Glx
60	Ala	Gly	Ser	Phe	Leu	Ala	Ile	Val	Met	Gly	Gly	Gly	Asx	Leu	Glx	Val	Ile	Leu	Ile	Ser
80	Leu	Ala	Gly	Arg	Glx	Glx	Ser	Ser	Ile	Gam	Ala	Ser	Arg	Ser	Leu	Ala	Ala	Ala	Met	Ser
100	Thr	Thr	Ala	Ile	Pro	Ser	Asx	Leu	Trp	Gly	Asx	Eps	Ala	Del	Ser	Asx	Ala	Ala	Phe	Ser
120	Ser	Del	Glx	Phe	Ser	Ser	Del	Ala	Gly	Ser	Val	Pro	Leu	Gly	Phe	Thr	Phe	Gam	Glx	Ala
140	Gly	Ala	Lys	Glx	Eps	Val	Ile	Lys	Gly	Glx	Ile	Thr	Gam	Glx	Ala	Del	Ala	Phe	Ser	Leu
160	Ala	Gam	Leu	Del	Lys	Leu	Ile	Ser	Ala	Met	Del	Asx	Ala	Gam	Phe	Pro	Ala	Gly	Asx	Del
180	Eps	Eps	Ser	Val	Ala	Asx	Ile	Del	Asp	Ser	His	Gly	Ile	Leu	Gam	Gam	Val	Asx	Tyr	Thr
200	Asx	Ala	Del	Ile	Lys	Met	Gly	Ile	Ile	Phe	Gly	Ser	Gly	Val	Asx	Ala	Ala	Tyr	Trp	Cys
220	Asx	Ser	Thr	Eps	Ile	Gly	Asx	Ala	Ala	Asx	Gly	Gly	Del	Gam	Gly	Gly	Ala	Gly	Gam	Met
240	Eps	Ile	Cys	Cys	Asx	Glx	Ser	Ser	Phe	Arg	Lys	Ala	Phe	Pro	Ser	Leu	Pro	Glx	Ile	Gam
260	Tyr	Leu	Del	Thr	Leu	Asx	Gam	Eps	Ser	Pro	Gam	Ala	Del	Lys	Thr	Phe	Eps	Lys	Asx	Ser
280	Gam	Ala	Lys	Asx	Eps	Gly	Glx	Ser	Leu	Arg	Asx	Val	Leu	Met	Del	Phe	Lys	Del	Del	Gly
300	Glx	Del	His	Del	Gam	Eps	Ala	Eps	Ser	Phe	Gam	Ala	Ala	Asx	Val	Glx	Asx	Thr	Ser	Tyr
320	Pro	Ala	Lys	Ile	Glx	Lys	Leu	Pro	His	Phe	Asx	Leu	Arg	Eps	Gam	Eps	Asx	Leu	Phe	Gam
340	Gly	Asx	Glx	Gly	Ile	Ala	Eps	Lys	Thr	Del	Met	Lys	Del	Val	Val	Arg	Arg	Gam	Leu	Phe
360	Leu	Ile	Ala	Ala	Tyr	Ala	Phe	Arg	Leu	Val	Val	Cys	Gam	Ile	Del	Ala	Ile	Cys	Glx	Lys
380	Lys	Gly	Tyr	Ser	Ser	Gly	His	Ile	Ala	Ala	Eps	Gly	Ser	Del	Arg	Ser	Tyr	Ser	Arg	Phe
400	Ser	Del	Asx	Ser	Ala	Thr	Gam	Asx	Del	Asx	Ile	Tyr	Gly	Trp	Pro	Glx	Ser	Ala	Del	Del
420	Ser	Lys	Pro	Ile	Del	Ile	Thr	Pro	Ala	Ile	Asx	Gly	Del	Gly	Ala	Ala	Ser	Gam	Val	Ile
440	Gam	Ser	Ile	Ala	Ser	Ala	Eps	Del	Ser	Gam	Ala	Del	Gam	Ser	Ala	Gam	Gam	Ala		

that a claim was made that the model described 60% of the residues correctly, its amino acid composition shows considerable discrepancy from that determined by careful analysis (Schmidt and Colowick, 1973a and fig. 1.1) in over half the amino acid types. It is still weighted towards residues of short side-chains and it still possesses 17% of the total amino acids in the form of unidentified aliphatic side chains (fig. 1.2).

1.2.2 Substrate binding by yeast hexokinase

The majority of the X-ray crystallographic work has been carried out on the B-II asymmetric dimer.

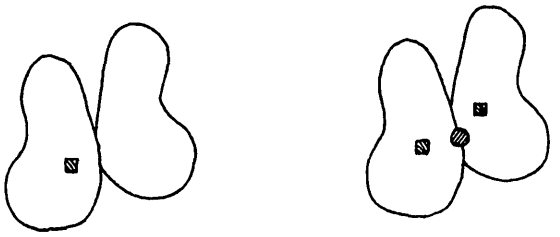


B-I B-II

Crystalline complexes of hexokinase with its substrates were prepared by soaking the crystals in 70% ammonium sulphate solution containing the ligands at 10-50mM concentration. Low resolution (7Å) studies of these crystals reveal that glucose, other sugar substrates and glucose analogue inhibitors such as o-toluoyl glucosamine and o- and p-iodobenzoylglucosamine bind in the deep cleft that separates the two subunits (Anderson and Steitz, 1975)(see diagram).

Although the dimeric crystalline enzyme has two such glucose-binding sites per dimer, in fact it binds only one molecule per dimer. However, in the presence of the nucleotide substrates two molecules of glucose and one of nucleotide are bound per dimer, the nucleotides binding to a site between the

two subunits on the dyad axis, as is shown on the diagram below:



10mM glucose

10mM glucose
+20mM MgADP

substrate binding to crystalline hexokinase

Nucleotides will not bind to the crystal in the absence of sugars. Thus substrate binding is highly synergistic, a phenomenon also seen in the kinetics of hexokinase.

The heterotropic interactions of substrate binding to the BII crystals may be due to the heterologous arrangement of the subunits and indeed be an artefact of crystallisation, unrepresentative of the properties of the enzyme during its catalytic cycle in solution. Early solution studies (Colowick et al, 1969; Colowick and Womack, 1969; Noat et al, 1969) concluded that two molecules of sugar were bound per molecule of dimeric enzyme in the absence of nucleotide, and that the dimer bound sugars less well than the monomer (Colowick et al, 1969), suggesting that glucose-binding sites became more accessible upon dissociation and providing a 'thermodynamic' model for the phenomenon of glucose-induced dissociation of dimeric hexokinase (Derechin et al, 1972).

Since the direct-binding methods of measurement showed that binding of glucose by hexokinase is too weak to measure any cooperativity directly, glucose binding has perforce been studied indirectly, by utilising the observation that glucose binding quenches the intrinsic fluorescence of hexokinase, to an extent proportional to the extent of binding (Zewe et al, 1964).

Hoggett and Kellett (1976a) showed that the binding of glucose to dimeric hexokinase in solution could be represented by a single binding constant, and that at saturating glucose the relative fluorescence of the monomeric and dimeric forms were quenched equally, which suggested that in solution the dimer bound two molecules of glucose equally well. This they proved conclusively by studying the response of the monomer-dimer equilibrium to changes in the glucose concentration.

Although they found the dissociation constant of the monomer for glucose to be 8-20 fold less than that of the dimer over the pH range 6.5-8.3, in accordance with the thermodynamic model, these findings are somewhat invalidated by the fact that their constants were measured under widely differing conditions of pH and ionic strength, factors known to influence the compactness and flexibility of proteins (Tanford, 1961). At pH 8.3 glucose binding to the intact monomer is increased four-fold by increasing the ionic strength from 0.03 to 0.13 (Feldman and Kramp, 1978). Mayes et al. (1982) have claimed that the binding of glucose to the monomeric S form of hexokinase B shows no ionic strength dependence at this pH, but it is possible that this result may be related to use of the proteolysed S form.

By studying the variation in the hexokinase-glucose dissociation constant, and the association constant of hexokinase dimerisation with ionic strength and glucose concentration, Feldman and Kramp (1978) showed that at pH 8.3 the two glucose dissociation constants for the dimer are approximately equal, but are only twice the dissociation constant for the monomer.

The sugar-binding site has been investigated by difference electron density mapping between hexokinase and its complexes with glucose and glucose derivatives.

The glucose moieties of all such derivatives bind in the same orientation and conformation as glucose itself (Anderson and Steitz, 1975; Fletterick et al, 1975; Steitz et al, 1977; Anderson et al, 1978c): in the C-1 chair equatorial conformation, with the 1-hydroxyl axial (α-anomer), although both the α and β anomers are substrates and hexokinase shows no anomerase activity (Wurster and Hess, 1974). Bound glucose molecules interact with protein sidechains through the C-1,-3,-4 and -6 positions, a pattern of interactions that accounts for the observed sugar specificity of hexokinase (Sols et al, 1958). The groups on the protein that coordinate to glucose have been provisionally identified from the X-ray sequence (Anderson et al, 1978 b,c) and several have been found in the chemically derived sequences described in this thesis (see discussion).

The nucleophilic 6'-hydroxyl of glucose interacts with an aspartic acid carboxyl group, which is protected by the substrate from water and may act as a general base catalyst, assisting nucleophilic attack on the γ-phosphate of ATP by deprotonation. Involvement of a carboxylic acid is strongly suggested by Viola and Cleland (1978) from studies of the variation with pH of the profile of V_{max} and V_{max}/K_m for hexokinase. The temperature dependence of the pK_a of these profiles suggested the involvement of a group of pK_a 6.15, acting as a general base and general acid catalyst in the forward and reverse directions respectively. Perturbation of the pK_a with organic solvent confirmed that this catalytic group is a neutral rather than a cationic acid (ibid).

The interstitial ATP binding site of the BII crystalline form is unlike to be the catalytic site, since the γ-phosphate of ATP bound in this site is 20Å from the 6'-hydroxyl of bound glucose, which is too far distal for phosphoryl transfer to

occur. It is unlikely to be an effector site, since the hysteretic properties observed for hexokinase at low pH (Kosow and Rose, 1971; Shill and Neet, 1971, 1975; Peters and Neet, 1977) were shown to be due simply to contamination of commercial preparations of ATP with aluminium ions (Womack and Colowick, 1979).

Another type of nucleotide binding site is observed in the crystalline monomer, hexokinase BIII, to which AMP and 8-Br AMP bind in the ratio one per monomer (Steitz et al, 1977; Shoham and Steitz, 1980). Binding of glucose and AMP to this crystalline form is mutually exclusive, so no crystalline ternary complex has been studied. The fact that AMP is an inhibitor that is competitive with ATP (Rudolph and Fromm, 1971) suggests that this is the true catalytic site for ATP.

The structure of this binding site has been investigated by means of difference electron density mapping between the hexokinase BIII-AMP complex and the apoenzyme. The triphosphate chain of ATP was model built into the site by overlapping the crystal structure of tetraammine cobalt^{III} tripolyphosphate (Merritt et al, 1978) with the α-phosphate of AMP and a putative site for the γ-phosphate of ATP located from the position of an enzyme bound sulphate ion.

The adenine group lies on the surface of the small lobe in a shallow depression at the entrance to the glucose-binding cleft. The plane of its ring interacts with a tyrosine residue at position 397. The ribose is bound anti- to the adenine. Although the preferred configuration of nucleotides is syn-, in the anti-configuration the 5'-hydroxyl points towards the glucose-binding site. The dihedral angle χ at the glycosidic bond of ATP is 180°. Enzymes that cause χ to

change from its free solution value show a high degree of base specificity (Mildvan, 1981). Upon binding to hexokinase this dihedral angle increases from 140 to 180° (T.A. Steitz, unpubl.), and indeed hexokinase shows a high base specificity for adenine (Hohnadel and Cooper, 1972).

The ribose pucker is C-2'-endo, which allows both of the ribose cis-hydroxyl groups to interact with the protein, in a manner consistent with the specificity of hexokinase (ibid.). The α-phosphate is seen as a disordered region of low electron density on difference maps, and does not appear to coordinate to the enzyme. The tripolyphosphate chain binds in an extended conformation pointing towards the glucose 6'-hydroxyl. Both the β- and γ-phosphates form hydrogen bonds to the protein.

Although the nucleotide substrate seems to bind to hexokinase in an ordered manner in a configuration optimised for phosphate transfer, crystals of a hexokinase-glucose complex are catalytically inactive. The γ-phosphate binding site is 6Å from the 6'-hydroxyl of glucose, a distance too great for direct phosphate transfer. If the model-built ternary complex represents a reaction intermediate that is viable or nearly so, then the rigidity of the crystalline lattice may be preventing the expression of a conformational change necessary to bring the two groups sufficiently close together for reaction to take place.

1.3 The conformation of the nucleotide substrate

Electron density mapping alone will not yield detailed information about the conformation of the nucleotide or its interaction with its coordinating metal ion.

Because the various coordination isomers of MgATP exist

in solution as a set of rapidly interconverting stereoisomers it is not possible to determine the particular conformation towards which hexokinase shows activity. This problem has been solved by the use of coordination inert complexes of nucleotides with trivalent metal ions (de Pamphilis and Cleland, 1973; Cornelius et al, 1977) which interconvert sufficiently slowly to permit their separate purification.

The coordination-inert compound chromium-adenosine-5'-0-triphosphate ($Cr^{III}ATP$) prepared as a dead-end inhibitor of hexokinase (de Pamphilis and Cleland, 1973; Dunaway-Mariano and Cleland, 1980a) is a slow substrate for hexokinase (Danenberg and Cleland, 1975) which will bind about 25% of the complex present in solution (ibid). It appears that of the isomeric forms present in solution only the β, γ -bidentate forms are inhibitors, and on the basis of this specificity of chromium nucleotides as slow substrates and inhibitors (Janson and Cleland, 1974) it was suggested that the substrate for hexokinase is β,γ-bidentate MgATP and that the product is β-monodentate MgADP (ibid; Dunaway-Mariano and Cleland, 1980b).

Cr^{III} complexes are paramagnetic and thus not amenable to study by either ^{1}H or ^{31}P NMR; however, the closely related Co^{III} complexes do not suffer from this restriction, and the active isomer of tetraammine $cobalt^{III}ATP$, $Co(NH_3)_4ATP$, which is a slow substrate for hexokinase (Cornelius and Cleland, 1978) was indeed demonstrated by ^{1}H and ^{31}P NMR (Cornelius et al, 1977) to be β, γ -bidentate.

A β, γ-bidentate metal nucleotide complex should exist as a mixture of two diastereomers and indeed hexokinase shows substrate activity towards only half of a solution of $Co(NH_3)_4$ ATP (Cornelius and Cleland, 1978). The inactive isomer was degraded without loss of chirality to tetraammine $cobalt^{III}$

tripolyphosphate, which was crystallised and the structure solved by X-ray crystallography (Merritt et al, 1978). This inactive isomer is of Δ,(right hand) ring sense and thus the active isomer in the hexokinase reaction is the Λ,(left hand), β,γ-bidentate form (Dunaway-Mariano and Cleland, 1980b). Both ring conformers of the correct screw sense are substrates (ibid).

Hexokinase is active towards both the Δ and Λ ring sense isomers of β,γ-bidentate $Cr^{III}ATP$ through catalysis of the epimerisation of the unfavourable isomer at the observed slow enzymatic rate (ibid). It is possible that it catalyses the epimerisation of ΔMgATP, thus allowing both ring sense isomers to act as substrates and eliminating unproductive binding of ΔMgATP.

The α-phosphate of ATP does not coordinate to the metal ion at all, as is shown by the lack of metal ion-dependent reversal of stereochemistry for ATPαS (Jaffe and Cohn, 1979).

1.4 The kinetics and mechanism of hexokinase

Hexokinase seems not to have any role in the cellular control of metabolism. It shows none of the criteria of Newsholme and Start (1973) of a control enzyme, operating neither at equilibrium nor at a crossover point in metabolism. It shows Michaelis-Menten type kinetics under the conditions typically employed for assay (approx. 1.0µg/ml) but these conditions may not be an accurate representation of those found in vivo.

Hexokinase, like all the other enzymes of glycolysis, is present at high concentrations in the yeast cytosol, of the order of 10^{-4}M (Hess et al, 1969); about equimolar with the concentrations of some of the glycolytic intermediates.

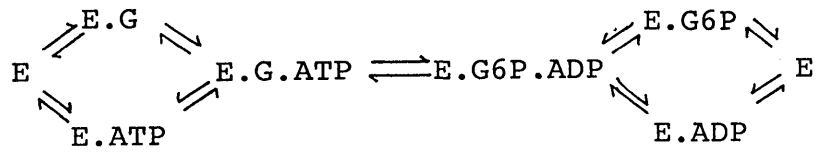
The turnover numbers measured *in vivo* for the glycolytic enzymes are similarly balanced, but lower than the maximum literature values. The apparent kinetic constant of hexokinase, of the order of 1mM, is well above the steady-state concentrations of most of the glycolytic intermediates. Together with the high dissociation constants for enzyme-substrate complexes, these data indicate a low degree of saturation of enzymes with substrates, between 1 and 40%. Thus at medium glycolytic flux the enzymes are at substrate levels well below those of half-maximal saturation, and the flux is very responsive to small changes in effector levels or metabolic demands.

Investigations into the kinetics of the hexokinases have made use of both isoenzymic forms almost indiscriminately. It seems that they share a common mechanism, both in terms of structural events and kinetic nature, but with differing kinetic constants. Consequently the following description of the mechanism has, wherever possible, eschewed placing numerical values to such constants, but attempts to describe the mechanism in general terms.

The majority of the data from kinetic methods such as initial rate studies (Fromm and Zewe, 1962), competitive inhibitors (Fromm and Zewe, op. cit., Zewe *et al*, 1964; Fromm, 1969; Danenberg and Cleland, 1975), product inhibition studies (ibid), isotope exchange reactions at equilibrium (Fromm *et al*, 1964) and the use of alternative substrates (Rudolph and Fromm, 1970) tend to support the hypothesis that reaction proceeds via ternary substrates and products complexes formed by random but not rapid equilibrium addition of substrates, whose interconversion is slow but not rate-limiting (Purich and Allison, 1980). Product release rather than interconversion of ternary

complexes seems to be the rate-limiting factor of the reaction (Fromm et al, 1964; Purich and Allison, 1980). The mechanism of the reaction does not involve a kinetically significant ping-pong type mechanism via a phosphoryl-enzyme intermediate (Solomon and Rose, 1971; Purich and Fromm, 1972).

Although either substrate may be the first to bind, the qualitative contributions of the two pathways may differ depending on the conditions of measurement. Kosow and Rose (1970, 1971) have suggested that the lower pathway becomes important at low pH and high concentrations of ATP.



The actual extent of ordering is unclear: Kosow and Rose (1970) and Britton and Clarke (1972) claim that the reaction is substantially but neither totally nor compulsorily ordered; Viola et al (1982) claim from a reappraisal of much of the existing data that the mechanism is partly ordered. Kinetic evidence from the use of analogues of ATP (Hohnadel and Cooper, 1972; Danenberg and Danenberg, 1977) suggest that glucose is the obligatory first substrate, as does evidence from direct binding studies (Noat et al, 1969) and rapid reaction kinetics (Wilkinson and Rose, 1979) which suggest that hexokinase cannot form a functional binary nucleotide complex in the absence of glucose.

It is clear that hexokinase can bind nucleotides in the absence of sugars, since it has a weak ATPase activity with a K_m for nucleotides approximately 50x that seen for the phospho-

transferase reaction (Trayser and Colowick, 1961; Kaji and Colowick, 1965). DelaFuente and Sols (1970) propose that ternary substrates complexes formed by prior addition of nucleotides are kinetically abortive. However, the failure to observe other than weak direct nucleotide binding to hexokinase may be due to the observed synergism in binding rather than to formation of an abortive complex, since such synergism is most marked between combinations of sugars and nucleotides that are good substrates.

Isotope-trapping experiments have shown that functional prior binding of glucose can occur, since all the bound glucose of a hexokinase-glucose complex is trapped as glucose 6-phosphate by ATP (Rose et al, 1974). Glucose dissociation is slow for the binary complex, and partly rate-limiting in the reverse direction of reaction, but slower still from the ternary substrates complex, much slower than the rate-determining step of product formation. This substrate synergism, a nucleotide-induced tightening of sugar binding, may be due to a conformational change that accompanies nucleotide binding.

The rate of formation of the functional hexokinase-glucose complex is of the order of $10^6 M^{-1}s^{-1}$. If formation of this complex were diffusion-controlled, the rate should be of the order of $10^9 M^{-1}s^{-1}$. Thus a protein conformational change or some other event must be rate-limiting.

In the back-reaction, similar experiments (Cleland, 1977) indicate that glucose 6-phosphate binds very tightly to the ternary products complex but less so to the binary complex. Thus binding of the products pair shows the same nucleotide-induced synergism as the substrates pair and may involve the same conformational change. These experiments also show

that the products ternary complex releases MgADP much more rapidly than it does glucose 6-phosphate, indicating that product release is largely ordered.

Inhibition (Kosow and Rose, 1970) and flux-ratio (Britton and Clarke, 1972) data for the reverse reaction also suggest that release of MgADP precedes release of glucose 6-phosphate, and that even at saturating levels of MgADP, only 36% of the flux goes by prior release of glucose 6-phosphate. The hexokinase-MgADP complex decomposes much more rapidly than the hexokinase-glucose 6-phosphate complex, since whereas a net incorporation of labelled material from ADP into ATP occurs during the forward reaction, none occurs from labelled glucose 6-phosphate.

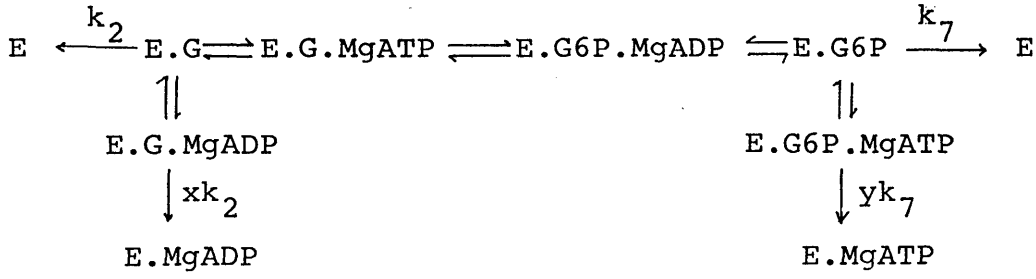
The extreme conditions (of low pH and high ATP/ADP ratio) required *in vitro* for ATP binding to precede glucose binding are unlikely to be approached *in vivo* even under extreme aerobic conditions. It is most probable that *in vivo*, hexokinase shows an ordered mechanism in which glucose binding is the obligatory initial event.

The *in vitro* mechanism, although fundamentally, is not simply ordered. MgADP and MgATP inhibit non-competitively at low levels but competitively at high, when they must thus be competing for the same site.

Thus at high levels of MgADP, this nucleotide must bind both to the hexokinase-glucose and hexokinase-glucose 6-phosphate complexes; similarly, when MgATP acts as product inhibitor vs MgADP, it binds to both binary complexes. The abortive ternary complexes, hexokinase-glucose-MgADP and hexokinase-glucose 6-phosphate-MgATP release the bound sugar or sugar-phosphate at rates comparable to their

rates of release from product binary complexes. There is thus only slight synergism in the formation of these unproductive complexes, whereas more might have been expected from molecules that fill the active site virtually completely.

The accumulation of the enzyme molecules into abortive complexes at high nucleotide levels was postulated by Fromm et al (1964) to explain the fact that at high nucleotide level the glucose $\leftrightharpoons$ glucose 6-phosphate exchange reaction was not inhibited. Exchange of free and bound sugars would occur via the abortive complexes E.G.ADP and E.G6P.ATP rather than via E.G or E.G6P, which were trapped as the ternary complexes E.G.MgATP and E.G6P.MgADP. The decreased bimolecular rate constant for G6P binding accounts for the observation of a higher K_m than an ordered mechanism employing rate constants for formation and breakdown of functional complexes alone would suggest.



This mechanism is proposed by Viola et al (1982) to represent only one of the possible courses of a steady state random mechanism, since alone it cannot account for the well-characterised binding of chromium nucleotides to hexokinase in the absence of sugar substrates (Danenberg and Cleland, 1975), nor for the observed incorporation of labelled material from glucose 6-phosphate into glucose at saturating levels of

MgATP (Fromm et al, 1964).

In the steady state at saturating concentrations of substrate, approximately half of the enzyme forms have transferred the phosphate group (Wilkinson and Rose, 1979). The rest of the enzyme must be present as the substrate ternary complex, E.G.MgATP, since the substrates are saturating. The equilibrium constant for phosphate transfer between the enzyme-bound ternary complexes is thus approximately unity. This represents an equilibrium highly displaced towards glucose and ATP compared to the equilibrium constant of $\sim 10^3$ for the overall reaction in solution at pH 7.0 (Robbins and Boyer, 1957). The substrates pair must be bound 10^3x more tightly than the products pair. The enzyme has evolved the phosphate transfer step in the interconversion of ternary complexes to ideality. A similar phenomenon is observed for adenylate, arginine, creatine, 3-phosphoglycerate and pyruvate kinases (Nageswara Rao et al, 1976;1978; 1979; Nageswara Rao and Cohn, 1981). It may well be a phenomenon general to kinases and common among phosphotransferases.

1.4.1 Nature of the sugar-induced conformational change

Kinetic evidence that a conformational change in the protein does occur in formation of the binary hexokinase-glucose complex comes from the observation of considerable substrate synergism in the intrinsic nucleoside triphosphatase activity of hexokinase. Non-phosphorylatable analogues of glucose cause the K_m for nucleotides to fall to the value seen for the phosphotransferase reaction, and bring about an increase in the catalytic efficiency of the phosphatase activity (de la Fuente et al, 1970). The sugar "substrate" must be inducing a conformational change that increases the lability of the γ-phosphate as well as the affinity of the enzyme for the

nucleotide. Similarly the presence of sugars causes hexokinases to bind CrATP as much as 40 times more tightly over a period of several minutes (Danenberg and Cleland, 1975). The better a sugar as a substrate for the phosphotransferase reaction, the better it is able to induce this progressive tightening of nucleotide binding. Since CrATP has equal affinity for hexokinase in the absence as in the presence of sugars, the tigher binding seems to be due to a conformational change preceding the catalytic events.

That a change in the conformation of hexokinase is induced by sugar binding and is not dependent upon any other factor is shown by a variety of evidence; thus the temperature-jump relaxation kinetics of glucose binding show a single relaxation time, the kinetic analysis of which is consistent with a slow isomerisation or conformational change that occurs subsequent to glucose binding (Hoggett and Kellett, 1976b); likewise, glucose binding to hexokinase causes changes in the ultraviolet chromophores of the enzyme (Cohn, 1963; Peters and Neet, 1978); crystals of the hexokinase-glucose complex produced by diffusing low concentrations of glucose into the crystalline enzyme disintegrate upon exposure to high concentrations of glucose as crystalline lattice forces prevent full expression of a sugar-induced conformational change (Anderson et al, 1978c); finally, inelastic neutron scattering analysis shows that a stiffening of the structure occurs upon binding glucose (Jacrot et al, 1981).

To investigate the nature of the glucose-induced conformational change the structures of the native enzyme and the binary complex with glucose have been compared. Isoenzyme A of hexokinase has been crystallised in the presence

of glucose (Womack et al, 1973) under conditions that should permit expression of any glucose-induced conformational change, and its structure solved to 3.5Å resolution (Bennett and Steitz, 1980a). Hexokinase A is chemically very similar to hexokinase B (Womack et al, 1973) and electron density mapping shows the tertiary structure of the isoenzyme A-glucose complex and the isoenzyme B apoenzyme to be very similar except in the relative orientations of their small lobes (Bennett and Steitz, 1980b). Superimposition of the structures of the large lobes shows that upon sugar binding, the polypeptide backbone of the small lobe moves by up to 8Å towards the large lobe, closing the interdomain cleft in which sugars bind. This is equivalent to a rotation of 12° around and a translation of 0.9Å along a screw axis of a domain consisting of one third of the entire polypeptide chain (Bennett and Steitz, 1978).

The surface area of each bound glucose molecule is reduced by 94% when the two lobes close around it. Of the remaining 6% over half is due to the 6'-hydroxyl, which is left the only group available for reaction (McDonald et al, 1979).

The surface area accessible to solvent of the enzyme is reduced upon binding glucose, and again upon closure of the interdomain cleft. From the reduction in surface area upon glucose binding, the hydrophobic free energy change has been estimated from an empirical relationship. Its magnitude is comparable with that calculated from solution equilibria. Despite being favoured energetically the closed structure does not form in the absence of glucose, presumably because the presence of a water-filled cavity within the protein prevents favourable interactions via hydrogen bonds or van der Waals contacts (ibid).

To investigate whether the conformational change observed is due to differences between the two isoenzymic forms, solutions of hexokinase B in the presence and absence of glucose were studied by small angle X-ray scattering (Pickover et al, 1979). The radius of gyration of the enzyme molecule fell by 0.95Å upon binding glucose. The radius of gyration calculated from the atomic coordinates of the crystalline enzyme forms in the presence and absence of glucose shows a net decrease of 0.90Å to occur upon binding glucose. These two values are directly comparable in magnitude and direction, and so the conformational change is due solely to glucose binding and not to any differences between the two isoenzymic forms. It is unlikely that any other event in solution contributes to the observed change.

The conformational change is an essential part of the catalytic process, since molecules such as substituted glucoses which can bind to the enzyme with the glucose moiety in the correct orientation for catalysis but obstruct the movement of the two lobes are not substrates.

The function of the substrate-induced closure of the interdomain cleft is to exclude water from the active site formed by that closure. If the binding and catalytic sites of hexokinase were contiguous then the enzyme would function as an ATPase rather than as a phosphotransferase. The solution evolved by hexo- and phosphoglycerate kinases (and possibly by other kinases as well) is to divide the binding and active sites functionally and to merge the two, with the concomitant exclusion of water, by closure of the interdomain cleft. Thus water is prevented from interfering with phosphoryl transfer, and the active site region is increased in dielectric constant, which aids the attack of the substrate nucleophile on the

γ-phosphorus atom. The aspartic acid sidechain carboxylate that acts as a general base catalyst in hexokinase and adenylate-, fructo- and phosphoglycerate kinases (and which may be a feature common to phosphokinases that are unidirectional) is increased in effectiveness by removal from aqueous solvent. The molecular architecture of the enzyme seems to have evolved to carry out the reaction in this way.

The essential determinant for the sugar substrate-induced conformational change seems to be the presence of the nucleophilic 6'-hydroxymethyl group of the hexose substrate. Hexokinase B crystallised in the presence of xylose, the pentose analogue of glucose, is in the open form (Shoham and Steitz, 1982) which is contrary to expectations, since the effect of xylose in causing the K_m of the ATPase activity of hexokinase to fall is well documented (de la Fuente et al, 1970) and had been considered to be due to the same conformational change demonstrated to accompany glucose binding in the phosphotransferase reaction.

1.4.2 Crystalline structures represent kinetically viable intermediates

Despite evidence that a binary hexokinase-glucose complex is a catalytic intermediate, neither the binary complex formed by diffusion of glucose into crystals of hexokinase in the BIII form (Anderson et al, 1974) nor that formed by crystallisation of hexokinase A in the presence of glucose (Womack et al, 1973; Bennett and Steitz, 1980a,b) are catalytically active (Bennett and Steitz, 1978; Anderson et al, 1978, a,b; Wilkinson and Rose, 1980,1981). Glucose-induced closure of the interdomain cleft, although essential for catalysis, is not alone sufficient to induce a catalytically

active crystalline conformation.

If ADP is present in the crystallisation medium as well as glucose, crystals of hexokinase A containing tightly bound glucose and weakly bound ADP are formed (Wilkinson and Rose, 1980). Unlike those used for crystallographic work and described above, these crystals are catalytically active and, if suspended in a solution containing ATP, proceed to form products via their bound glucose. They exchange bound with free glucose at a rate that is less than that shown by crystals of hexokinase A crystallised in the presence of glucose alone, implying that a nucleotide-dependent conformation is stabilised by lattice forces. It is unfortunate that these crystals are not of a form suitable for X-ray crystallography (ibid).

The nature and catalytic status of glucose-containing crystalline hexokinase complexes has been investigated by performing isotope-trapping experiments on their freshly-dissolved solutions (Wilkinson and Rose, 1981). If they represent functionally different states from those of solution complexes, then they should exhibit different behaviour in solution.

The nonfunctional hexokinase-glucose crystal possesses a highly ordered glucose-binding site in which glucose is optimally oriented for phosphate transfer from ATP. The correctness of this binding is demonstrated by the fact that when freshly dissolved in a solution containing ATP, this complex traps 80% of its bound glucose as glucose 6-phosphate. Since this exceeds the equilibrium value of 50% observed in solution kinetics (Wilkinson and Rose, 1979), the complex must represent a form both capable of reacting with ATP and

functionally well advanced along the reaction pathway.

The functional crystal, studied in the same way, shows a requirement for ATP to trap glucose in product forms that is much lower than that of either the solution or the non-functional crystalline binary complexes, and so may be considered to represent a reaction intermediate that is further advanced. Clearly, this complex does not dissolve to a form functionally preceding ATP binding, but to one in which the binding site for ATP has been stabilised by lattice forces in addition to already possessing a similarly stabilised glucose-induced conformation. The nonfunctional crystal dissolves to the form preceding ATP binding, in which the ATP binding site is not fully optimised for phosphate transfer.

Because inhibitors competitive with ATP stabilise glucose binding and slow its rate of release (Wilkinson, K.D. and Rose, I.A., unpublished), binding of nucleotides may stabilise the glucose-binding site as well as the nucleotide site through an adenine nucleotide-dependent conformational change. MgATP has indeed been observed to induce an ultra-violet difference spectrum for hexokinase when glucose is on the enzyme (Roustan et al, 1974).

1.4.3 The stereochemical consequences of phosphate transfer

Kinetic and structural data provide no evidence about the mechanism of phosphate transfer, although they suggest strongly that it is direct. Lowe (1981) has described four possible mechanisms, of which one, involving a phosphoryl-enzyme intermediate, seems in the case of hexokinase to be ruled out on kinetic grounds (Solomon and Rose, 1971). These mechanisms can be distinguished by their stereochemical consequences at the transferred phosphate group. In-line can

be distinguished from axial transfer since the latter requires retention and the former inversion of the stereochemical configuration of the phosphorus atom.

The stereochemical consequences of phosphate transfer can be determined by making this group chiral. Phosphate and thiophosphate monoesters have been synthesised by incorporation of the stable heavy isotopes of oxygen, ^{17}O and ^{18}O (Abbott et al, 1978; Cullis and Lowe, 1978), or of sulphur and ^{18}O (Orr et al, 1978), and their absolute configuration determined. These groups can be transferred to ADP to produce ATP with a chiral γ-phosphate group of known configuration. Orr et al (1978) demonstrated by means of gas chromatography and mass spectrometry that the reactions catalysed by hexokinase, glycerol kinase and pyruvate kinase have identical stereochemical consequences. Glycerol kinase was then shown to catalyse the inversion of [γ(R)^{16}O,^{17}O,^{18}O]ATP (Blättler and Knowles, 1979), so this must also be true of hexokinase. Glycerol kinase was then independently demonstrated to cause the inversion of both phosphate and phosphorothiolate substrates (Pliura et al, 1980). Using ^{31}P NMR of D-glucose 6-[^{16}O,^{17}O,^{18}O] phosphate (Jarvest et al, 1980), Lowe and Potter (1981) confirmed unequivocally that hexokinase-catalysed phosphate transfer involves inversion of configuration at phosphorus. In both cases the inversion has stereoselectivity in excess of 90%.

Associative (S_N2) and dissociative (S_N1) mechanisms of phosphate transfer have identical stereochemical consequences and cannot thus be distinguished. However, if reaction involves a metaphosphate intermediate, then unless the β-phosphat of ATP is rotationally restricted, there will be a transfer of

the β,γ -bridging oxygen atom into a β-nonbridging position (Midelfort and Rose, 1976; Rose, 1979). In the case of hexokinase no such transfer occurs either in the absence of glucose or in the presence of lyxose. In the presence of glucose and ATP half the quantity of ATP that proceeds to products undergoes oxygen interchange (Rose, 1980). Thus no more scrambling occurs than can be accounted for by simple partition between enzyme-bound ternary complexes, and so no mechanism of phosphate transfer other than by direct in-line transfer to glucose can take place.

1.5 Summary

The complete mechanism of hexokinase, as deduced from data of all origins, involves conformational changes upon addition of both substrates. Upon binding of each substrate the enzyme closes up around that site and tightens the binding - in the case of glucose by closure of the interdomain cleft and in the case of ATP by an as yet uncharacterised conformation change that tightens sugar binding as well.

Activation of the sugar 6'-hydroxyl could probably occur immediately subsequent to glucose binding since in the hexokinase-glucose binary complex the catalytic aspartate group is protected from solvent water. Conformational change following nucleotide binding forces the chelate ring of the metal-nucleotide complex into an active conformation without breaking the metal-phosphate coordination bonds. This conformational change brings the reacting groups sufficiently close together.

Activated, nucleophilic glucose displaces the γ-phosphate of ATP with inversion of configuration in an S_N2 mechanism

without formation of a metaphosphate intermediate. The transition state is stabilised by delocalisation of negative charge onto the magnesium ion and other groups hydrogen-bonded to the γ-phosphate oxygen. These interactions must occur only during catalysis of the phosphotransferase reaction and are thus brought about by conformational changes, otherwise hexokinase would be a better ATPase.

The immediate product is cis $Mg(H_2O)_4(ADP)(GDP)$ as demonstrated by the isolation of the corresponding product from $Cr^{III}ATP$. After transphosphorylation the magnesium ion migrates from β,γ -bidentate to β-monodentate coordination. Coordination to the γ-phosphate, now transferred to glucose, is lost. This change in the pattern of coordination acts as the signal for product release. The inability of Cr(ADP)(G6P) to undergo this shift accounts for its slow rate of release as a substrate. MgADP is released first as the β-monodentate complex, followed by glucose 6-phosphate.

1.6 The requirement for primary structure determination

The kinetic and structural phenomena observed to occur during the catalytic cycle of hexokinase are inexplicable without determination of the molecular nature of the conformational changes that occur upon substrate binding. It is essential to determine the amino acid sequence or arrangement not merely at but also around each substrate-binding site, since upon substrate binding signals for some unspecified alteration of the binding site are transmitted across the molecule, presumably by making or breaking of interactions within the protein molecule. The residues that coordinate to the substrate molecules may in the apoenzyme be involved in other, unspecified, interactions. Upon coordination these

interactions are broken and this presumably causes a slight shift in the relative positions of parts of the polypeptide backbone that results in tighter binding at the other substrate-binding site. Only by extensive sequence determination can the interactions other than substrate-binding be identified and their modes of action proposed.

In the absence of the primary structure of a protein molecule it is of considerable interest and importance to crystallographers to be able to maximise the information obtainable from an electron density map, particularly with regard to explanation of kinetic data for that protein in molecular terms. Obviously to be able to satisfy the amino acid sequence from an electron density diagram with complete accuracy requires that the resolution of that diagram should not exceed the average interatomic spacing in the molecule. Unfortunately, such resolution is rarely even approached. Nevertheless, sufficient gross structural features may be observed to enable the prediction of the identity of particular amino acids, and even to propose a sequence. Comparison of the actual primary structure of a protein with that deduced from high-resolution X-ray crystallography will reveal whether or not this exercise is justified by its accuracy. If so, it may be that without having to resort to sequencing either polypeptide or polynucleotide, it is possible to identify the nature of regions of structural interest or catalytic importance from X-ray crystallography alone. If, however, the accuracy of the crystallographic data is low, then production of an amino acid sequence in this way may not after all be a worthwhile exercise. Although the X-ray data would be informative on gross structural arrangement, it would not alone be sufficient data

for explanations of mechanisms, which would require the classical methods of the protein chemist.

We have recently heard (T.A. Steitz, personal communication) that, following a report that the gene for hexokinase had been cloned, the nucleotide sequence of both the A and B isoenzymic forms had been completed. Despite the fact that considerable delay would elapse before publication of such sequences, it was decided to terminate the work at the stage described in this thesis and not to proceed to subdigest for sequence confirmation.

CHAPTER 2

PURIFICATION OF HEXOKINASE

2.1 Materials and Methods

The peptide sequencing facilities available to the author were typically capable of elucidating 25-35 amino acid residues from 7-15 nmoles of peptide, the actual extent of sequence determination depending both on peptide composition as well as quantity.

An inevitable consequence of this was that it was necessary to purify hexokinase in relatively large quantities, of the order of 2 micromoles, that were not available commerciall except at prohibitive expense. Since modern methods of peptide purification produce high yields of soluble material, it might not be thought necessary to produce such large quantities. However, experience has shown that peptides derived from carboxymethylated hexokinase tend to precipitate readily in salt-free buffers, and that considerable losses may thus be expected to occur during purification.

The procedure employed to purify such large quantities was to scale-up method A of Barnard (1975). This bench-size method obtained typical yields of 200mg of hexokinase from 2.0kg of yeast. Moreover, it introduced dual steps of treatment with serine protease inhibitors which satisfactorily eliminated the extensive cleavage of hexokinase by the multiplicity of such proteases present in the cell supernatant. The treatment thus both greatly reduced the chromatographic complexity of the hexokinase activities produced and provided the reliably homogeneous product necessary for sequencing.

The procedure was scaled-up by a factor of 10. An increase in scale of this magnitude required the provision

of pilot plant facilities for cell breakage and centrifugation during the initial steps. Fortunately, such facilities are available at Imperial College. Following both absorption and ammonium sulphate precipitation, the preparation reduced to one of laboratory scale which was satisfactorily carried out in the cold room.

The procedure is described for a batch of either 20 or 24kg of yeast. With the exception of the volume of suspension buffer, smaller quantities of yeast employed a proportionally scaled-down procedure.

All the operations described were carried out at 4°C. Chemicals were of the highest grade commercially available in bulk. All solutions were made up with deionised distilled water.

The yeast (United Yeast Co. Ltd., Morden, Surrey) is suspended in 60l of 20mM Tris-HCl pH7.5 to which is added phenylmethanesulphonylfluoride (PMSF) to 10^{-4}M. The presence of this protease inhibitor is necessary and sufficient to inactivate serine proteases in the suspension. The cells are ruptured by passing the suspension twice through a chilled Manton-Gaulin homogeniser at 8 000 psi pressure, which gives 80-90% cell breakage. The cellular debris is removed by centrifugation on a Sharples disc-bowl centrifuge. The supernatant is taken as the crude extract. The cellular debris has no hexokinase activity and is discarded.

The amount of hexokinase activity in the crude extract varies both seasonally and with the batches of yeast used, but a typical value for 20kg of the yeast commercially available in London would be 2.5×10^{6}IU, based on the buffered-indicator

assay of Darrow and Colowick (1962).

The crude extract is cautiously acidified to pH 5.0 with 3M acetic acid. The precipitate formed contains mainly nucleic acids and is removed by centrifugation on the Sharples. The pH should not be allowed to fall below this figure since hexokinase is irreversibly inactivated by a pH of 4.6 or less. The extract is brought back to pH 7.5 with a solution of 2M Tris base. The suspension is then diluted with cold distilled water until the conductivity falls below 3mS.

To the diluted suspension is added 12 kilograms (damp weight) of precycled DEAE-cellulose (Whatman, DE23) equilibrated with Tris-HCl buffer pH 7.5 of a conductivity not exceeding 3mS. The slurry of resin and suspension is left cooled and stirred for 30 minutes to allow complete binding of hexokinase to the resin; or until the supernatant of a filtered aliquot shows no hexokinase activity.

The resin is removed from the supernatant by centrifugation in a muslin bag on a Spinco dryer. The damp resin is then transferred to 30 litres of 100mM Tris-HCl pH 7.5 and after a brief wash is baled out and centrifuged as rapidly as possible, since delay causes more activity to be lost than can be justified by the removal of other proteins. The resin is finally transferred to 35l of 500mM Tris-HCl pH 7.0. Most of the adsorbed enzyme is eluted by a 45 minute wash. It is impractical to transfer the resin with any hexokinase still adsorbed to a more concentrated wash since the hexokinase eluted thereby is heavily contaminated with other proteins.

The active supernatant is slowly acidified to pH 5.0 by dropwise addition of 3M acetic acid. Early attempts to purify hexokinases reported that three isoenzymic forms were present in yeast (Ramel et al, 1971), termed A, B and C in

their order of elution from DEAE cellulose by a gradient of decreasing pH. Isoenzymes A and B differ in primary structure; isoenzyme C can be converted into B by a combination of high salt concentration and low pH (ibid).

To the acidified suspension were added 17kg of ammonium sulphate, sufficient to bring it to 0.80 saturation. It was found necessary to make the addition over a period of 1-2 hours to obtain good precipitation. The suspension was held cooled and stirred for a further 2 hours before two passes of centrifugation on the Sharples centrifuge. The solid precipitate contains all the hexokinase. The pellet is stable at 4°C.

The solids are redissolved in the minimum volume of buffer, either 10mM Tris-HCl pH 7.0 or 10mM sodium succinate pH 5.8. The solution is adjusted to pH 7.0 with 2M Tris. Sufficient of a 1M solution of di-isopropylfluorophosphate in dry ethane-1,2-diol is added to bring its concentration to 10^{-4}M. The solution is left stirred and cooled for 30 minutes before centrifugation for 1 hour at 8 000 g on a Sorval RC-5 centrifuge. The clear supernatant contains all the hexokinase activity. It is desalted by gel filtration on a column (15 x 90cm) of Sephadex G-25 equilibrated with 10mM sodium succinate pH 5.8. The flow rate is approximately 20 litres/hour. All the protein separates cleanly from the salt and is eluted from the column as a visibly brown band.

This solution is acidified to pH 5.0 with 3N acetic acid, added dropwise. This causes considerable precipitation. The suspension is centrifuged for 1 hour at 8 000g on the Sorval RC-5 centrifuge. The clear supernatant contains all the hexokinase activity. The supernatant is adjusted in pH back to 5.8 and is diluted until its conductivity does not

exceed 1.1mS. The solution is subjected to chromatography on DEAE-cellulose (Whatman, DE-52) equilibrated with 10mM sodium succinate, pH 5.8, by employing a gradient from pH 5.8 to pH 4.6. The two hexokinase isoenzymes are largely resolved by this gradient into peaks eluting at about pH 5.5 (A) and 5.3(B). The pools thus made of each isoenzyme are then separately subjected to rechromatography on a smaller scale under the same conditions. The resolution between the hexokinase isoenzymes is improved. It is found that the apparent isoelectric pH of each isoenzyme falls, to 5.3 (A) and 5.1 (B). These figures approximate better to the values found from analytical isoelectric focussing by Schmidt and Colowick (1973b) and Hoggett and Kellett (1976a). The apparent isoelectric pH found by chromatographic elution tends to vary with the conditions employed and the amount of protein present. The hexokinase is substantially pure at this stage. It is concentrated upon a column of DEAE-cellulose (Whatman, DE52) (approximately 1ml of resin per 10mg of protein) equilibrated with 20mM Tris-HCl pH 8.0. The enzyme is removed from any volume of solution at this pH. It is eluted with 1.0M sodium chloride in 10mM sodium succinate, pH 5.0. The enzyme is stable in this solution for several months if 1mM ethylenediaminetetraacetic acid is added as preservative, and will not precipitate provided that the protein concentration does not exceed 10 mg/ml.

The concentrated solution is finally subjected to gel filtration on a column of Bio-gel P60 (2.5 x 85cm) equilibrated with 1.0M sodium chloride in 10mM sodium succinate, pH 6.0. The protein is now pure by the criteria of producing a single band upon SDS-PAGE and having a specific activity of 650-800 IU/mg. The net yield of the purification is about 20% (hexo-

kinases A and B) for a purification factor of the order of 10^3.

Data for a typical preparation are shown in fig. 2.1. Since SDS-PAGE alone is not an unequivocal demonstration of the purity of a protein, and because existing sequence information was restricted to the N-terminal 11 residues, the protein product was characterised and its purity proved by sequencing.

A 10-fold increase in the scale of the preparation was unfortunately not able to produce an equivalent increase in the magnitude of the yield. The unavailability of any means of removing enzyme-bound DE-23 rapidly from either the suspension buffer or, particularly, from the 100mM washing buffer meant that considerable quantities of hexokinase were eluted therefrom and were thereby lost. Although losses in the suspension buffer could be eliminated or reduced by provision of sufficient DE-23, losses in the washing buffer were never actually eliminated but merely reduced to acceptable levels (< 20%). The strength of the eluting buffer was increased from the 450mM of Barnard (1975) to 500mM, and the fractional saturation of that buffer with ammonium sulphate from 0.75 to 0.80 likewise, in successful attempts to produce slight increases in yield. Further development of the method would have required considerable revision of the conditions for the batch absorption procedure, a task which, although necessary, was not fully justifiable on grounds of the time available, since the procedure used and developed produced adequate, if unexceptional, quantities of pure hexokinase.

2.2 Characterisation by sequencing of the protein product

Hexokinase B (8 mg) was gel filtered on a column

Fig 2.1 Purification of Hexokinase B from 20kg of Yeast

	Volume, ml	Activity, units	Protein, mg	Sp.Act., U/mg	Purific^{n} factor	Yield, %
Suspension buffer, 20mM Tris-HCl pH 7.0	47×10^3	2.35×10^6	18×10^6	0.13	1.0	100
Elution buffer, 500mM Tris-HCl pH 7.5	40×10^3	1.00×10^6	2.8×10^6	0.36	2.76	42
Desalted Ammonium Sulphate precipitate	2×10^3	1.00×10^6	2.2×10^5	3.7	28.4	42
Chromatography on DE-52:						
Hexokinase A	1620	450×10^3	2430	185	1423	19
Hexokinase B	1920	164×10^3	1058	155	1192	7
Total Hexokinase		614×10^3				26
Chromatography on DEAE-Sephacel:						
Hexokinase B	420	178×10^3	336	530	4070	7.6

(90 x 1cm) of Bio-Gel P-60 (Bio-Rad Laboratories Inc., Richmond, Calif., USA) equilibrated with 10% (v/v) formic acid. The protein ran as a single homogeneous peak which was pooled and lyophilised.

The lyophilised protein was reduced and carboxymethylated (see 5.1 for experimental details), dialysed with 5% (v/v) formic acid and lyophilised. The carboxymethylprotein was taken up in 500µl of 10% (v/v) aqueous triethylamine for application to the spinning cup sequencer.

43 cycles of the Edman degradation were performed. The sequence was determined by a combination of PTH-analysis by HPLC and back-conversion of PTH-amino acids to the free amino acid by acid hydrolysis, followed by amino acid analysis (5.1).

The sequence was determined for the first 29 residues:

Val-His-Leu-Gly-Pro-Lys-Lys-Pro-Gln-Ala-
Arg-Lys-Asp-Gly-Ser-Met-Ala-Unk-Val-Pro-
Lys-Glu-Leu-Met-Gln-Thr-Ile-Unk-Asp-

A considerable drop in the repetitive yield occurred at step 13. It is probable that position 13 is asparagine rather than aspartic acid, and that internal cyclisation between positions 13 and 14 is blocking the sequence (Bornstein, 1970), which would account for the drop in repetitive yield.

The first 11 residues of the sequence are identical with the sequence of a peptide removed from intact hexokinase by gentle tryptic digestion (Schmidt and Colowick, 1973b) and known to be essential for dimerisation to take place (ibid).

CHAPTER 3

STRATEGY FOR PROTEIN SEQUENCING

The complexity of mixtures of peptides derived from cleavage of hexokinase with proteolytic enzymes such as Trypsin or Chymotrypsin means that even with the availability of HPLC-based methods of peptide purification the labour involved would be too great, and the average peptide length and yield too low, to justify the adoption of such methods as primary digests.

Of available reagents of limited proteolytic specificity, cyanogen bromide seemed the most suitable. Hexokinase B possesses 11 methionine residues per monomeric subunit. The peptide chain length of this monomeric subunit is 472 residues, so digestion of hexokinase with cyanogen bromide should produce 12 peptides of average length 40 residues. This presents an ideal basis for sequencing, since the labour involved in purification will not be unrealistic and the sequence of each peptide thus produced should be mostly determinable, since the Beckman sequencer can produce sequences of about 30 residues from a peptide present in sufficient quantity. The peptides would also be highly suitable for subdigestion with good quality proteolytic enzymes to obtain the complete sequence.

As a means of confirming sequence data obtained from one primary digest it is advisable to perform a second, preferably on material obtained from a different bulk purification. This digest should remove any ambiguities in the data from the primary digest and provide overlaps between the primary peptides. One means of obtaining sequence data

rapidly and nonspecifically is electron-impact mass spectrometry.

The maximum amount of sequence information can be obtained if the protein substrate is digested fairly randomly, such as by the use of a non-specific enzyme. The serine protease elastase has such a limited specificity, cleaving principally at lysine and the aromatic and hydrophobic amino acids. The labour involved in purifying peptides obtained by such methods to the standard required for Edman degradation would be exhausting, but by employing the mixture analysis technique of interpreting Mass Spectra it is possible to examine simple mixtures of peptides and thus to avoid many of the time-consuming and loss-creating steps incumbent upon exhaustive peptide purification.

CHAPTER 4

AUTOMATED SEQUENCE DETERMINATION

4.1 The modified Beckman sequencer

Sequence determination was performed on a Beckman 890B sequencer which had been adapted with some of the recommendations of Hunkapiller and Hood (1978) and Wittman-Liebold (1973). The original vacuum system was modified: rough and restricted vacuum systems were taken from the same pump by means of valves connected in parallel: the fine vacuum pump was replaced by a pump from Edwards High Vacuum Co., the valve operating the transition from rough to fine vacuum was replaced by a more rapidly responsive system from Leybold-Heraeus.

The vacuum outlets from both the fine and restricted vacuum systems were connected by flexible metal hosing to a cold-trap immersed in a 60 l dewar of liquid nitrogen, which was refilled daily. Interposition of this dewar greatly improves the long-term vacuum performance, which otherwise falls off rapidly within the timescale of a single sequence run. The traps were emptied every 50-60 cycles.

The vacuum cell and housing were cleaned out with the same frequency to remove extensive deposits of impurities (principally diphenylthiourea) that might otherwise contaminate the sample.

For sequencing runs of over 10 residues a modified Beckman 0.1M Quadrol program was employed. This incorporated the recommendation of Hunkapiller and Hood (1978) of adding the coupling buffer in two stages, thus preventing drift of the pH during the coupling stage. The program incorporated double coupling, double cleavage and double extraction steps. The

first chlorobutane extract was directed to the fraction collector, the second was discarded.

For analytical runs of 5-10 residues a 'short' 0.1M Quadrol program was employed. This incorporated a long coupling stage (34 minutes) during which the coupling buffer was added in two steps, followed by single cleavage and extraction steps.

Typically, with the unpurified commercial reagents available it was possible to determine 30-35 residues from 5-20 nmoles of peptide.

4.2 Abbreviations of sequencing terms

ATZ	Anilinothiazolinone-amino acid
BuCl	n-butyl chloride
DMAA	N,N dimethylallylamine
HFBA	Heptafluorobutyric acid
HPLC	High Performance Liquid Chromatography
PITC	Phenylisothiocyanate
Polybrene	1,5 dimethyl 1,5 diazaundecamethylenepolymethobromide
PTC	Phenylthiocarbamyl-amino acid
PTH	Phenylthiohydantoin-amino acid
Quadrol	tetrakis (2-hydroxypropyl) N,N,N',N', diaminoethane
SPITC	4-sulphophenylisothiocyanate
TFA	Trifluoroacetic acid

4.3 Materials and Methods

Materials for automated sequencing

PITC was sequencer grade from Pierce Chemical Co., Rockford, Ill., USA. It was employed as a 5% solution in

n-heptane (Rathburn Chemical Co., Walkerburn, Scotland).

SPITC and DMAA were sequencer grade from Pierce.

HFBA, Benzene, Ethyl Acetate and n-Butyl Chloride were all sequencer grade from Rathburn. The ethyl acetate was 0.1% in acetic acid.

Quadrol was 1.0M in n-propanol/water:3/4 (v/v) from Pierce. It was diluted to 0.1M with n-propanol/water:3/4 (v/v) before use.

Water was distilled and subsequently deionised by passage through a Milli -Q^{TM} water purification system (Millipore Associates).

Polybrene and glycylglycine were from Sigma Chemical Co., Poole, Dorset, U.K.

Before sequencing peptides derived from cleavage of hexokinase by hydroxylamine, n-propanol, HFBA and heptane were all repurified according to the procedures of Hunkapiller and Hood (1978). Ethyl acetate was redistilled.

For peptides derived from cleavage of hexokinase by cyanogen bromide reagents were used without further purification.

Polybrene (Tarr, 1978) was routinely used to prevent peptide washout. 3 to 4mg of polybrene with 100 nmoles of glycylglycine were loaded into the cup and six cycles of Edman degradation performed, to ensure that any impurities in the polybrene capable of blocking free peptide α-amino groups were removed. The actual sample was not loaded until analysis of the sixth precycling step had proved to be free of any impurity.

Because commercially available sequencer-grade chemicals are not of a reproducibly high standard, this analytical run of the sixth precycling step was essential to investigate the

background level of impurities that would be present during the actual sequencer run. Faulty chemicals producing an unacceptably high background were discarded and the analytical precycling recommenced.

Although necessarily time-consuming, these precautions were essential to ensure reproducibly high yields at each cycle of sequencing of "clean" peptides.

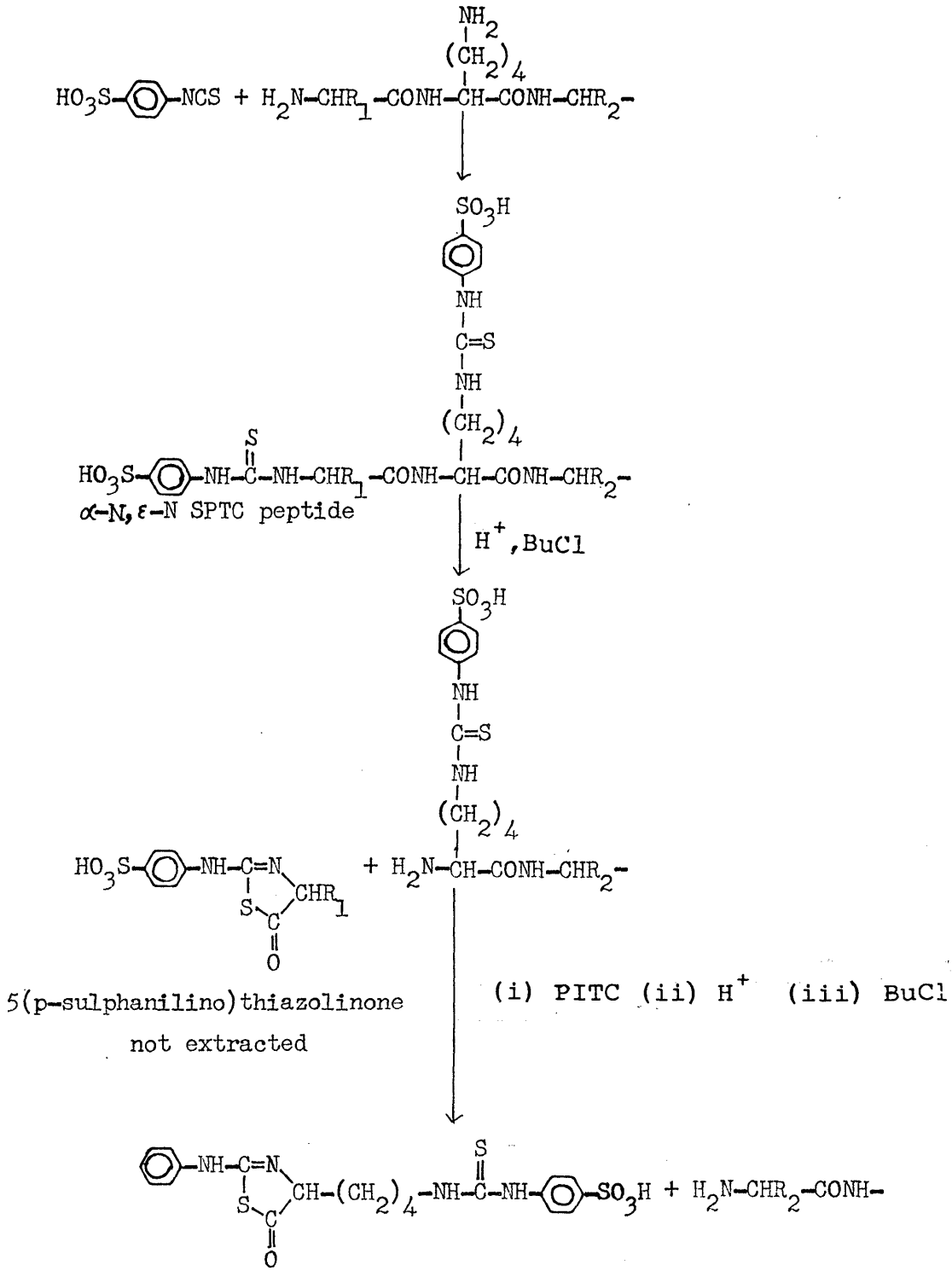
4.3.1 Coupling of peptides with Braunitzer's reagent

Braunitzer's reagent (4-sulphophenylisothiocyanate) (SPITC) (Braunitzer, 1970) is a form of PITC modified in order to increase the hydrophilicity of any peptide to which it is coupled. In the case of particularly hydrophobic peptides such treatment reduces the extent of washout from the sequencer cup by the non-aqueous solvents benzene and ethyl acetate employed after the coupling stages of the sequencer cycle. The reagent has the same specificity as PITC and blocks free N-termini and the ε-amino group of lysine residues (fig. 4.1). Lysine residues labelled thus are not extracted from the sequencer cup by n-butyl chloride and appear as gaps in the sequence. Their identity can be surmised from the absence of the degradation products of serine and threonine, which also appear as gaps, and from the presence of a slight trace at the normal position of lysine, possibly due to an incomplete labelling.

The procedure for derivatisation was adapted from that of Blakley (1977). Coupling of N-terminal and lysine ε-amino groups was performed under the controlled conditions possible in the cup of the protein sequencer. The peptide was introduced to the cup and was dried under vacuum in the usual way. The peptide film was then taken up in 300 µl of 0.4M DMAA/TFA

Fig 4.1

Coupling of peptides with Braunitzer's reagent

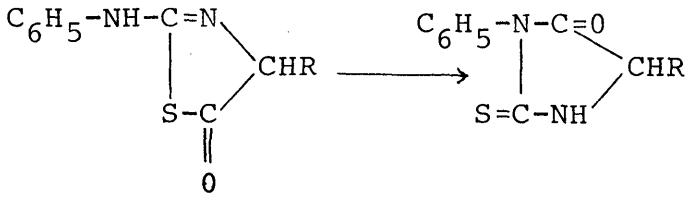


buffer, pH 9.5, to which were then added 20µl of a fresh 2% aqueous solution of SPITC. The cup was flushed with argon for 1 minute. Incubation was carried out at 50°C for 1 hour. The solvents were removed by drying under vacuum. Excess reagent and any byproducts of reaction were removed by employing an extensive wash with n-butyl chloride. The normal sequencer program was then commenced at step 1, so that the peptide was exposed to PITC in case the coupling with SPITC had been incomplete.

4.3.2 Conversion of ATZ- to PTH-amino acid

The n-chlorobutane solution of the extracted ATZ-amino acid was first dried down under a stream of Argon. The residue was dissolved in 50µl of 20% (v/v) trifluoroacetic acid and the aqueous extract incubated under Argon for 10 minutes at 80°C in a Grant JB1 water bath. The aqueous solution was dried under vacuum. It was then taken up in 50µl of methanol (HPLC grade from Fison's) and a suitable aliquot analysed by HPLC.

If either serine or threonine were known in advance to be the residue in question, a much shorter incubation at 80°C of only 2 minutes was employed, to reduce destruction of the PTH-amino acid by dehydration. This treatment was found to be more successful for threonine than for serine. Deamidation of asparagine and glutamine was reduced in the same manner by use of a conversion time of 5 minutes.



ATZ-amino acid H^+(aq) PTH-amino acid

4.3.3 Identification of amino acid phenylthiohydantoins

The PTH derivatives of all 20 naturally occurring amino acids could be resolved on a column (4.5mm x 25cm) of Du Pont Zorbax ODS resin. The column was held in a water jacket and was maintained at 50°C by means of a circulating water bath. The phenylthiohydantoins were eluted by a gradient of increasing ethanol concentration.

Chromatography was performed on either a Waters Associates or a Varian Liquid Chromatograph. The Waters system consisted of two model 6000 pumps, a model U6K injector and a model 440 dual channel (254 and 280 nm) absorbance detector. The column was equilibrated with 0.030M sodium acetate pH 5.3/20% (v/v) ethanol and was developed with a concave gradient (No. 7 on the Waters Gradient Programmer) to 42% ethanol over 6 minutes, followed by isocratic elution.

The Varian system consisted of a model 5000 liquid chromatograph fitted with a Waters Intelligent Sample Processor with a 254 nm absorbance detector and a peripheral aerograph absorbance detector set to 280 nm. This system was initially equilibrated with 0.033M sodium acetate pH 5.3/17% (v/v) ethanol and was developed with a composite concave gradient to 40% ethanol over 8 minutes, followed by isocratic elution.

CHAPTER 5

CLEAVAGE OF HEXOKINASE WITH CYANOGEN BROMIDE

5.1 Materials and Methods

Reduction/ Carboxymethylation: Guanidine hydrochloride was ultrapure grade from Schwarz-Mann. Tris (Trizma base) and dithiothreitol were from Sigma Chemical Co. Iodoacetic acid (2x recrystallised) was the generous gift of Dr. M. Rangarajan Iodo (2-^{14}C) acetic acid was from Amersham.

Cleavage: Cyanogen bromide was from Sigma. It was purified by resublimation before use.

Gel Permeation: n-propanol was HPLC grade from Fison's. Acetic acid was Analar grade from BDH.

HPLC: TFA was sequencer grade from Rathburn's. Acetonitrile was HPLC grade from Fison's. All water for chromatography was either doubly distilled and further purified by passage through "Porapak Q" (Phase separations Ltd) or was singly distilled and purified by passage through a Millipore Associates Milli-Q^{TM} purification system.

Methods

Gel permeation and reverse phase high performance liquid chromatographies were done on a Waters Associates liquid chromatograph (described under 4.3.3, q.v.). Gel permeation used a dual Waters I-125 protein analysis column in 5% (v/v) acetic acid/10% (v/v) n-propanol; reverse phase HPLC was done

on a Waters Associates μ-Bondapak phenyl column with a gradient of acetonitrile in 0.1% TFA.

500 nmoles of pure hexokinase B (Sp.Act > 600 U/mg) were dialysed against 10% formic acid until the solution was sufficiently salt-free for freeze-drying.

Reduction and Carboxymethylation

The lyophilised protein was taken up in 6M guanidine-Tris HCl pH 8.5 to a concentration of 5mg/ml. Dithiothreitol was added to 2 fold molar excess over protein thiol groups and the solution incubated under nitrogen for 2 hours. Iodo-acetic acid dissolved in the same buffer was added to 5-fold molar excess over total thiol groups and the solution incubated in the dark under nitrogen for 30 minutes. Reaction was terminated by thorough dialysis with 10% formic acid, after which the protein was freeze dried.

Cleavage

Cyanogen bromide cleavage was carried out in 70% (v/v) formic acid at a protein concentration of 5mg/ml, with the reagent in 100-fold molar excess over protein methionine residues. Reaction was carried out at room temperature for 24 hours. Reaction was terminated by removal of the reagent by rotary evaporation. The solution was repeatedly reduced in volume and each time was diluted with progressively more dilute formic acid until acid concentration had reached 20%. The final protein concentration was about 10mg/ml.

Amino acid analysis

Dried samples were taken up in 50μl of 6N hydrochloric acid (Aristar) containing 2 nmole of norleucine/50μl of acid as internal standard. The sample tubes were evacuated and sealed. The samples were hydrolysed by incubation *in vacuo* for 18 hours at 105°C. After breaking open the sample tubes

and drying their contents in vacuo, samples were analysed on a Beckman model 750B automated amino acid analyser.

5.2 Pattern of cleavage (preliminary digest)

The primary digest of hexokinase, from which it was expected that the majority of the sequence data would be obtained, was made by digesting the protein with cyanogen bromide (Gross and Witkop, 1961). This cleavage was expected to produce 12 peptide fragments of average length 40 residues. The X-ray sequence predicted a range of sizes from 20-110 residues but it is known that this sequence underestimates the number of methionine residues present.

As a preliminary investigation of the practicality of cyanogen bromide digestion, 2mg of pure hexokinase B were digested with cyanogen bromide in 100:1 excess over methionine (methods). The progress of this cleavage was monitored by taking aliquots of 20µg for analysis by SDS-PAGE (Laemmli, 1970). This demonstrated that reaction was essentially complete after 10 hours.

The remainder of the digest was analysed by performance of one round of automated Edman degradation to give the nature and relative quantities of the N-termini produced by the digestion (fig. 5.1). The total number of peptides suggested by this preliminary investigation is 13, one more than expected from the number of methionine residues per monomeric molecule found by analysis.

5.3 Purification of peptides by chromatographic methods

Following transfer of the bulk digest into 20% formic acid by rotary evaporation, its contents were separated on

Fig 5.1

Residue found	Quantity, nmole	Number of peptides
D	0.20	2
E	0·18	2
G	0·18	2
A	0·18	2
Y	0·09	1
R	0·09	1
V	0·08	1
I	0·08	1
L	0·10	1

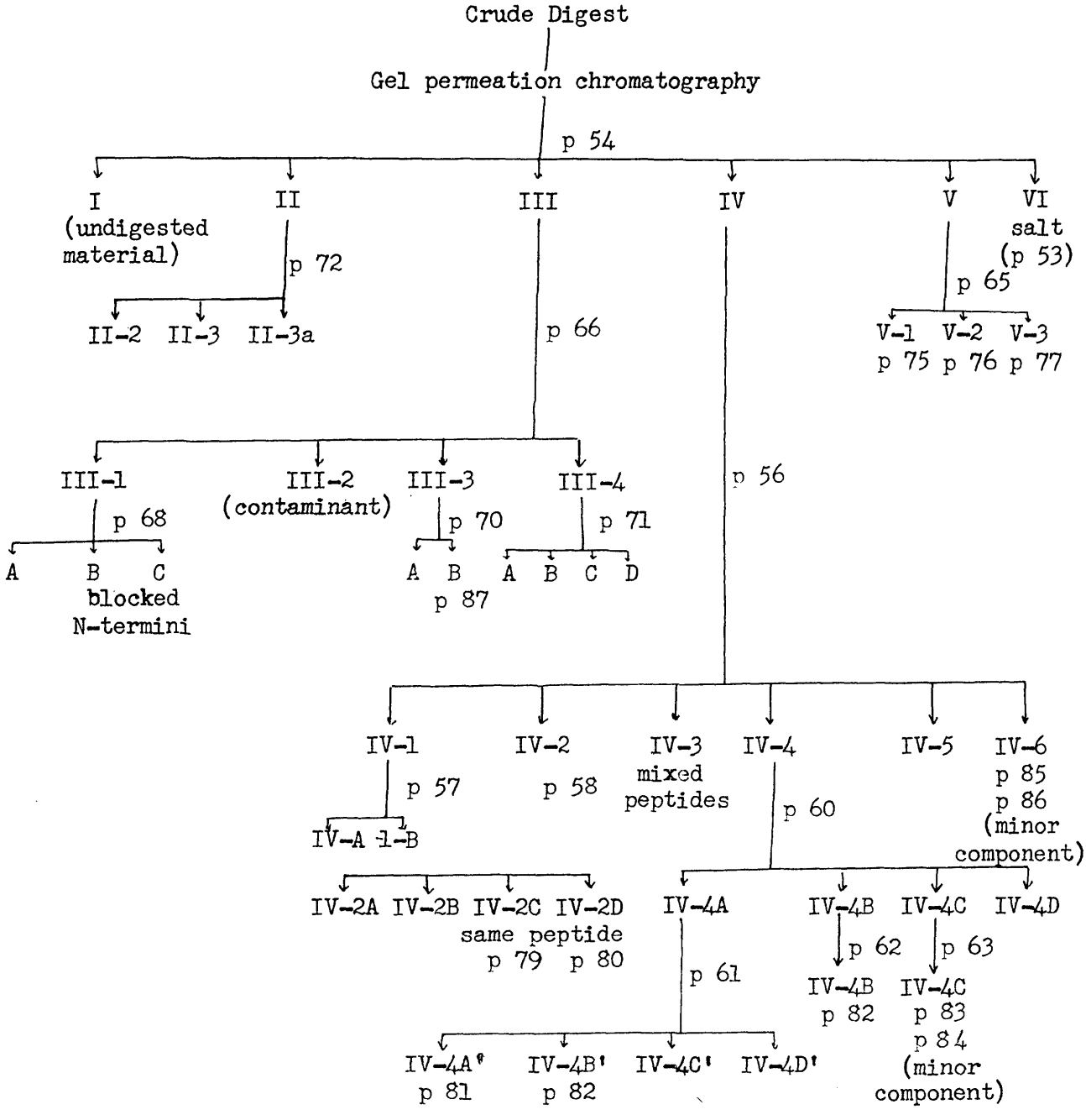
N-termini produced by cleavage of hexokinase with cyanogen bromide

a basis of size by gel permeation chromatography. Five major fractions were produced (fig. 5.2). The resolution of the system was lost if the volume loaded onto the column exceeded 250µl. The profile of the column was fully reproducible. The peptides were separated into pools by cutting at the points of local absorption minimum as indicated. Each of the six pools thus obtained was subjected to rechromatography on this column until it ran as a homogeneous peak.

Pool I is expected to contain material either totally or substantially undigested and was not subjected to further purification. Pool VI proved upon reduction of its volume to be heavily contaminated with salt. Despite its high absorption, it proved upon analysis to contain little material. It is expected to consist of very short peptides, if any, and was not subjected to further purification. Thus only poolsII-V were subjected to HPLC. Because pool IV was expected to be the most productive it was the subject of most effort.

A flow-chart summarising the process of peptide purification and stating where the detailed findings are available is shown overleaf.

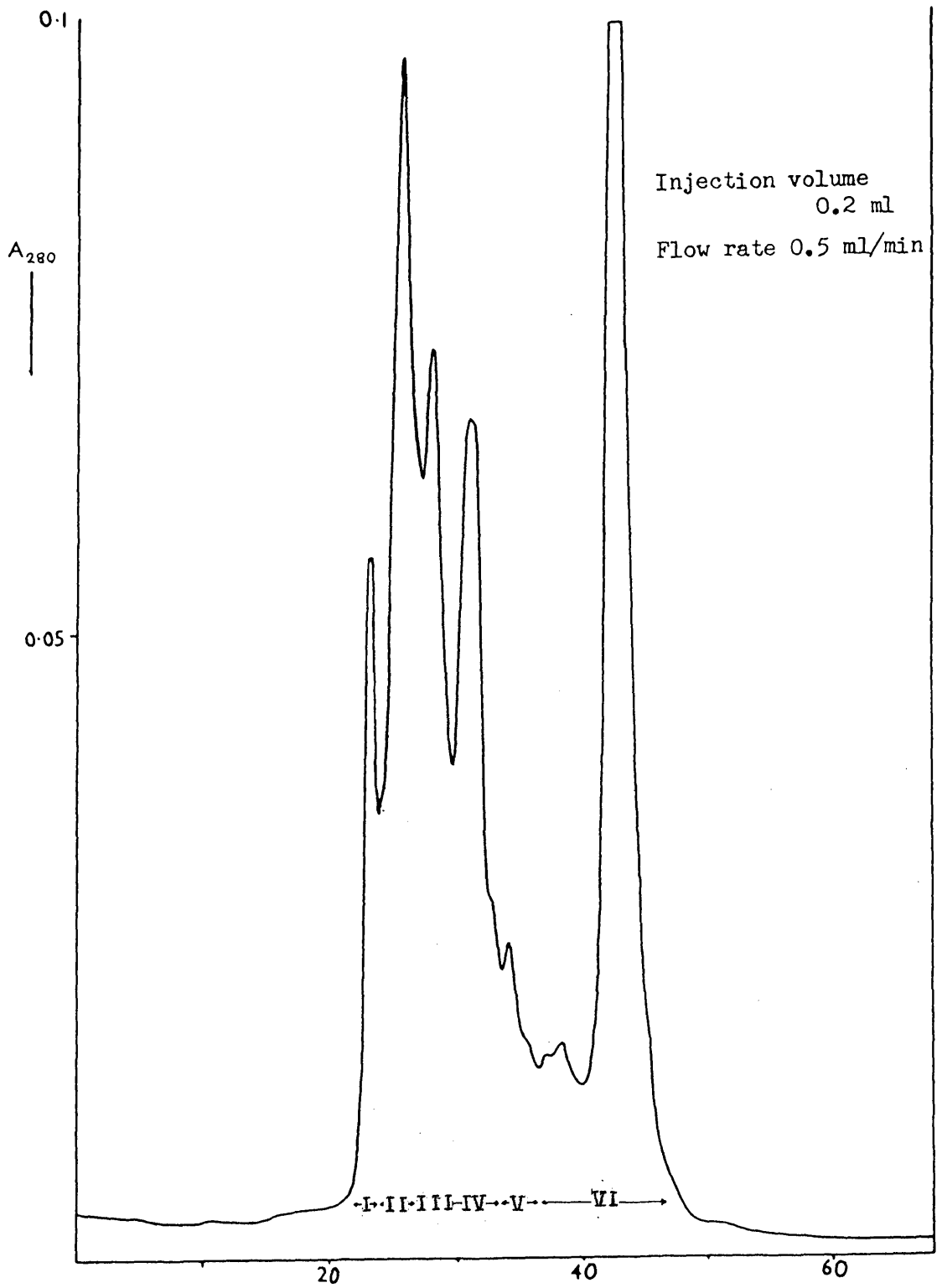
Flow-scheme for purification of peptides from a digestion of Hexokinase with Cyanogen Bromide



All peptide separations except the initial one were made by reverse phase HPLC. The page numbers by the side of each vertical arrow indicate the page numbers where the details of the chromatography performed and the peptide resolution achieved are to be found.

The page numbers beneath each peptide indicate the page where any sequence information obtained is to be found.

Fig 5.2: GEL PERMEATION CHROMATOGRAPHY OF CNBr DIGEST OF HEXOKINASE



5.3.1 Chromatography of Pool IV

The contents of this pool were resolved further by reverse-phase HPLC as described in methods (5.1). The column was developed with a linear gradient from 20 to 60% acetonitrile at 0.4% per minute. The peptides were extensively resolved as indicated in figure 5.3.

The contents of pools IV-1, -2 and -4 were all subjected to a further stage of purification by HPLC. The absorbance profile suggested that pool IV-6 contained a single component. An aliquot (10%) was taken for sequencing and confirmed that indeed this pool contained only one major peptide.

Although the absorbance profile of the resolution of pool IV did not suggest that pools IV-3 and IV-5 contained single peptides, aliquots of 10% were nevertheless taken for trial sequencing to characterise these pools further. Pool IV-5 proved to contain one major peptide in a quantity suitable for extended sequencing. Pool IV-3 contained several peptides and was not further employed.

The peptides of pools IV-1 and IV-2 were subjected to rechromatography, employing a linear gradient of acetonitrile from 24% at 0.25% min^{-1} with a flow rate reduced to 0.6 ml/min.

The resolution of pool IV-1 was marginally improved (fig. 5.4) but the two components suggested by the absorbance profile were unresolved. The column eluate was cut at the position of local absorption minimum.

Pool IV-2 was resolved into five components (fig. 5.5). The similarity of the amino acid composition of each suggested that they all contained the same peptide(s). Trial sequencing of pools IV-2C and IV-2D showed that for these two at least this was indeed the case.

The peptides of pool IV-4 were somehow resolved into four components by an isocratic concentration of acetonitrile

Fig 5.3

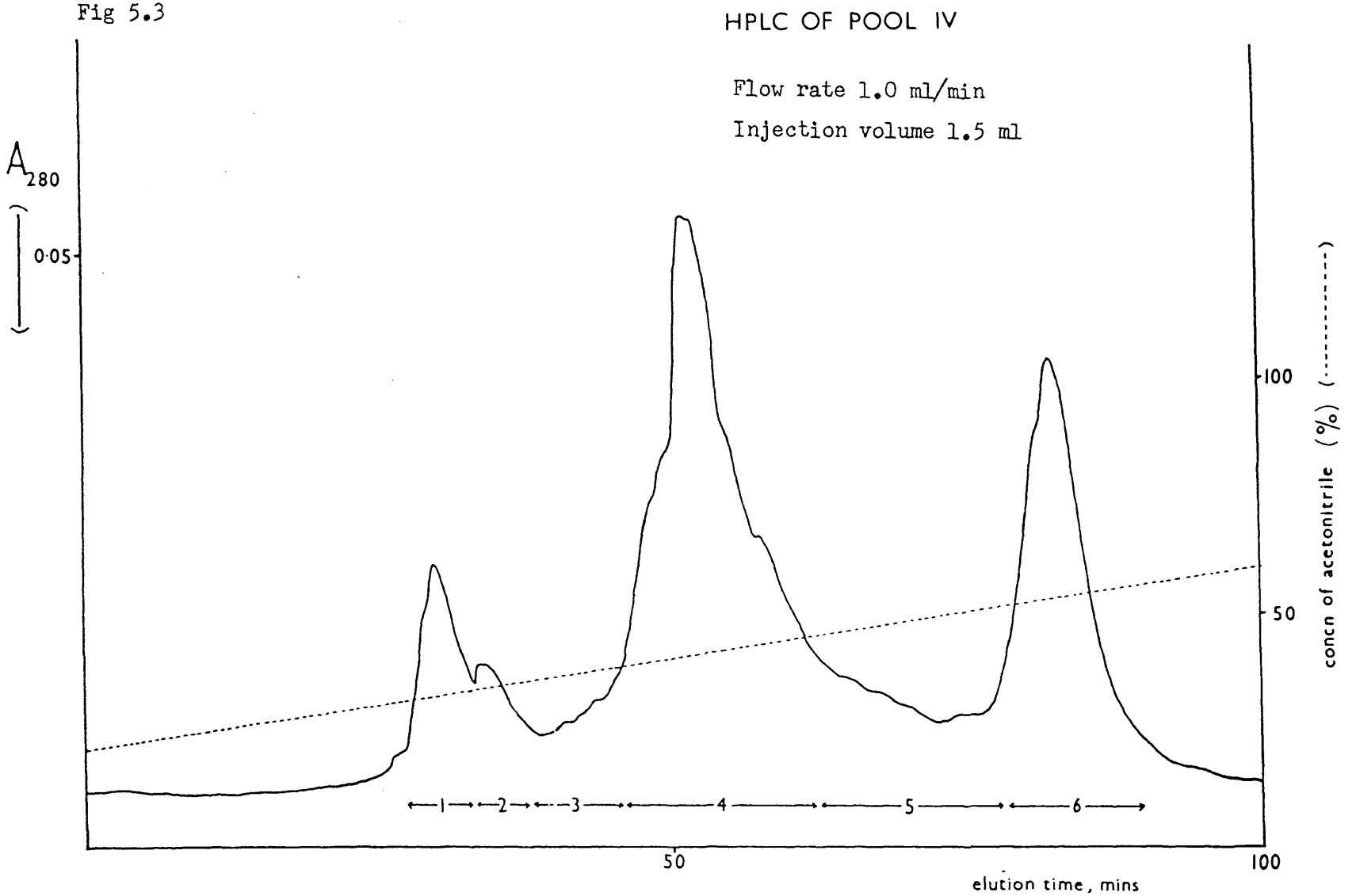

HPLC OF POOL IV
Flow rate 1.0 ml/min
Injection volume 1.5 ml
A280
0·05
100
50
concn of acetonitrile (%) (----------)
1
2
3
4
5
6
50
100
elution time, mins

Fig 5.4: HPLC OF FRACTION IV-I

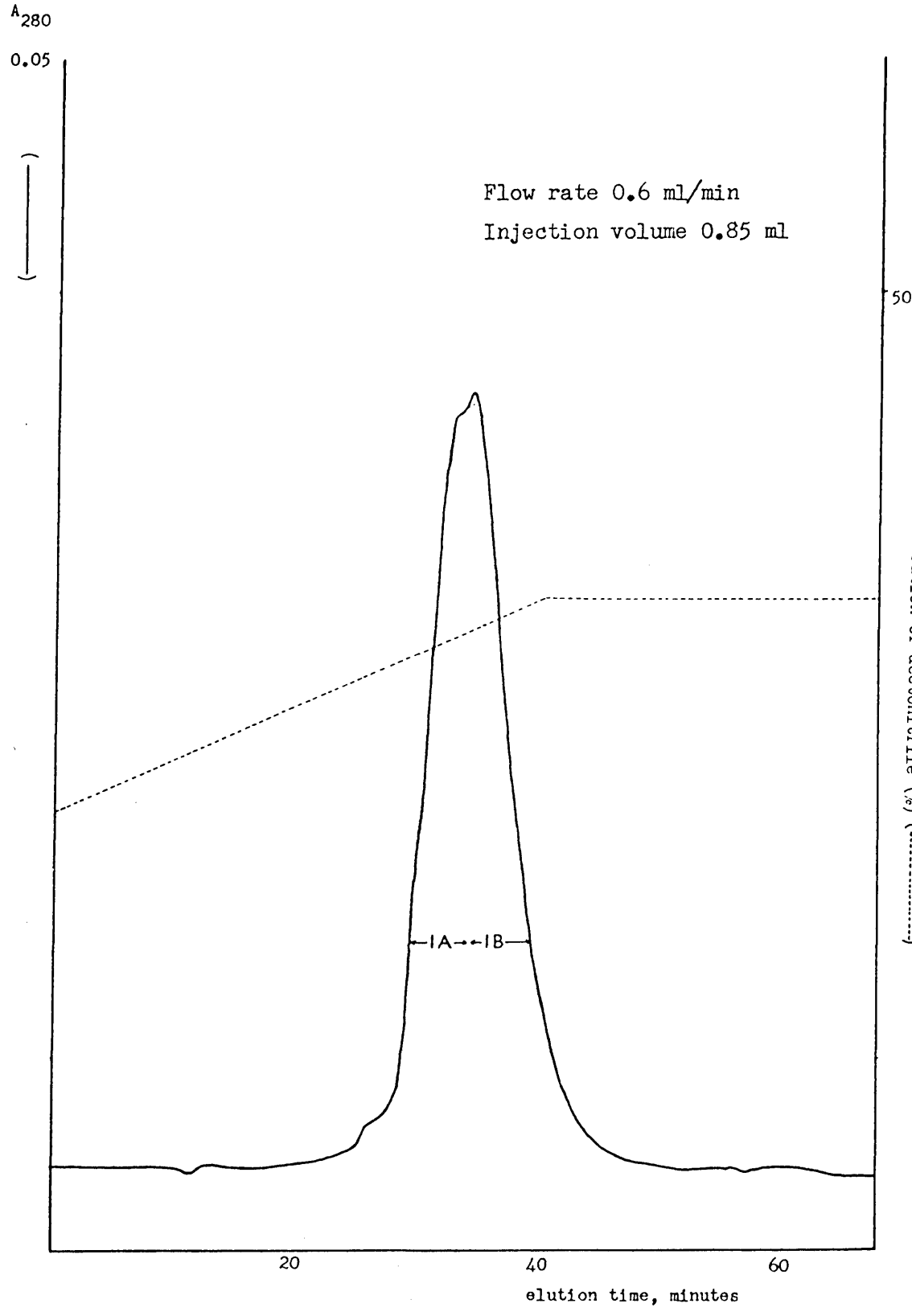
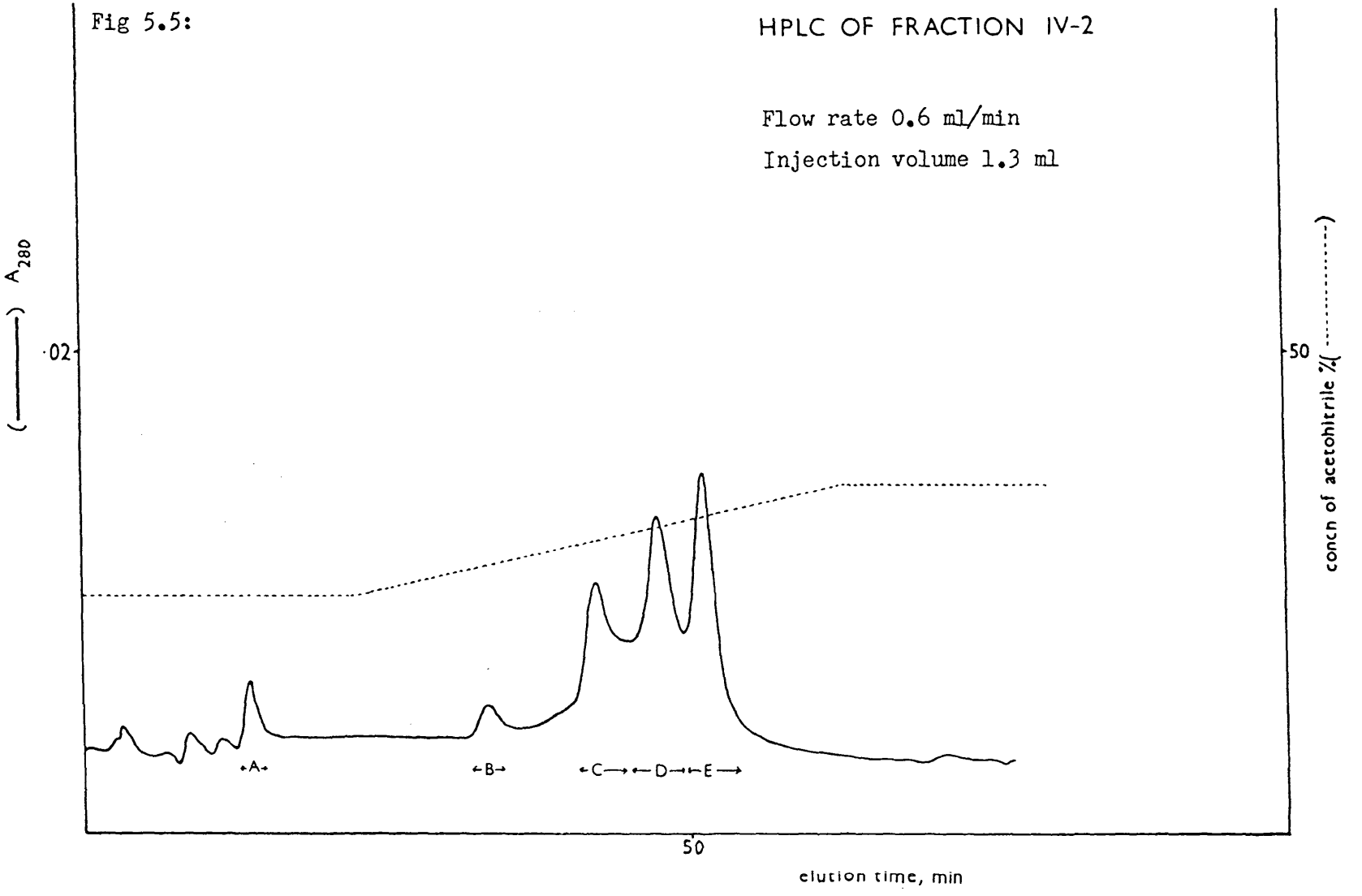


Fig 5.5:


HPLC OF FRACTION IV-2
Flow rate 0.6 ml/min
Injection volume 1.3 ml
(——) A_{280}
·02
50
concn of acetohitrile %(----------)
A
B
C
D
E
50
elution time, min

of 32% (fig. 5.6). The behaviour of these peptides had suggested that a concentration of about 40% would have been required for elution, so this isocratic separation may in fact have been an artefact of the system.

The first component to elute, IV-4A, was subjected to rechromatography (fig. 5.7) at 0.12% min^{-1}. It was thereby resolved into four components, IV-4A', -4B', -4C', and -4D'. The second and third components, IV-4B and -4C, were not, however, significantly further resolved upon chromatography (figs. 5.8-5.9).

Aliquots containing 10% of the material of each pool were analysed by producing sequences of 3-6 residues with the "short" 0.1M Quadrol program, as described in section 4.1. The results of these studies indicated the nature and quantity of each of the peptides present in each pool.

Only pools IV-1A, IV-2C, IV-2D, IV-4A', IV-4B', IV-4C, IV-5 and IV-6 contained peptides in sufficient quantity and purity to be deemed suitable for extensive sequence analysis. Only in these pools was the major peptide in sufficient excess over the other components to enable the allocation of each residue to its peptide to be unequivocal.

Fig 5.6 Isocratic separation of fraction IV-4

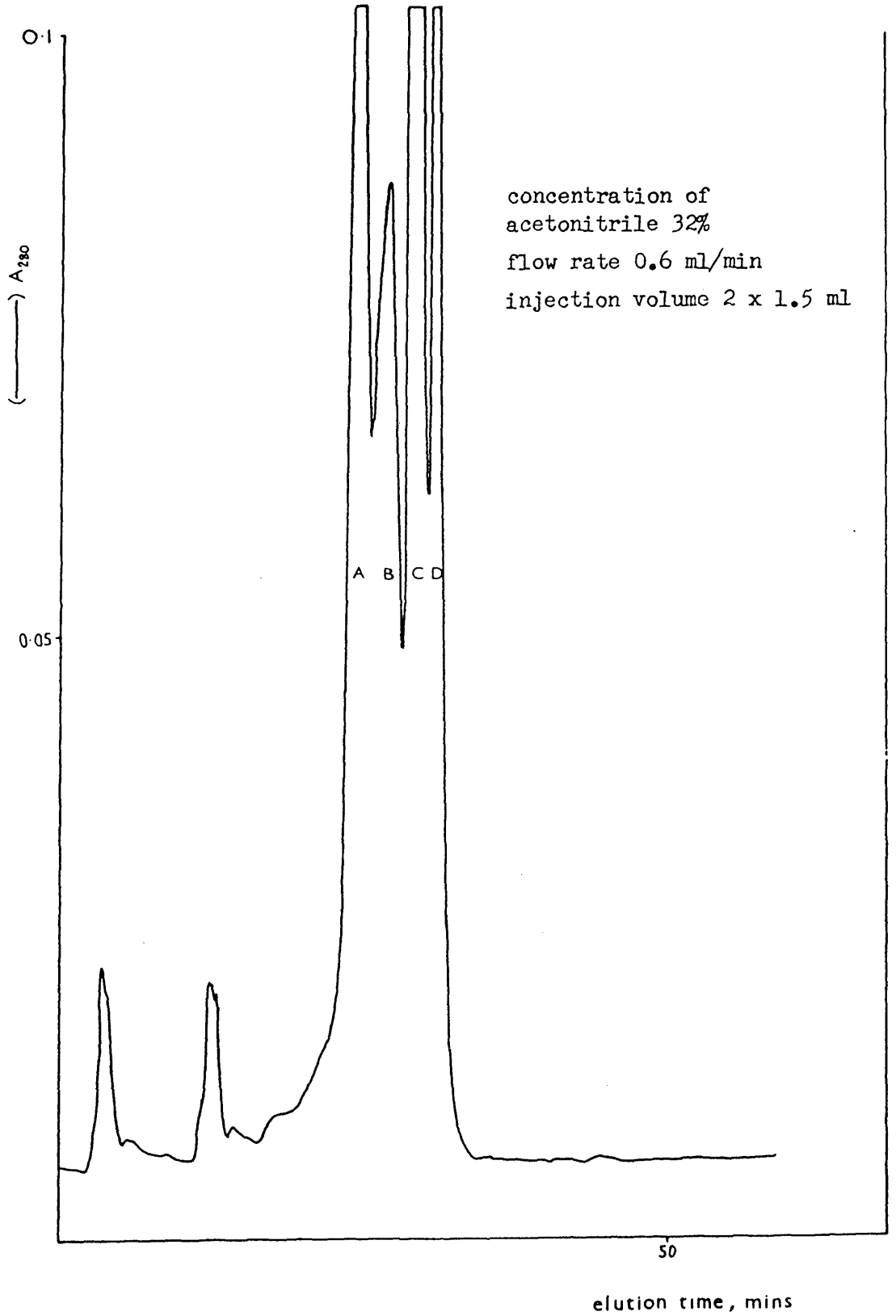


Fig 5.7

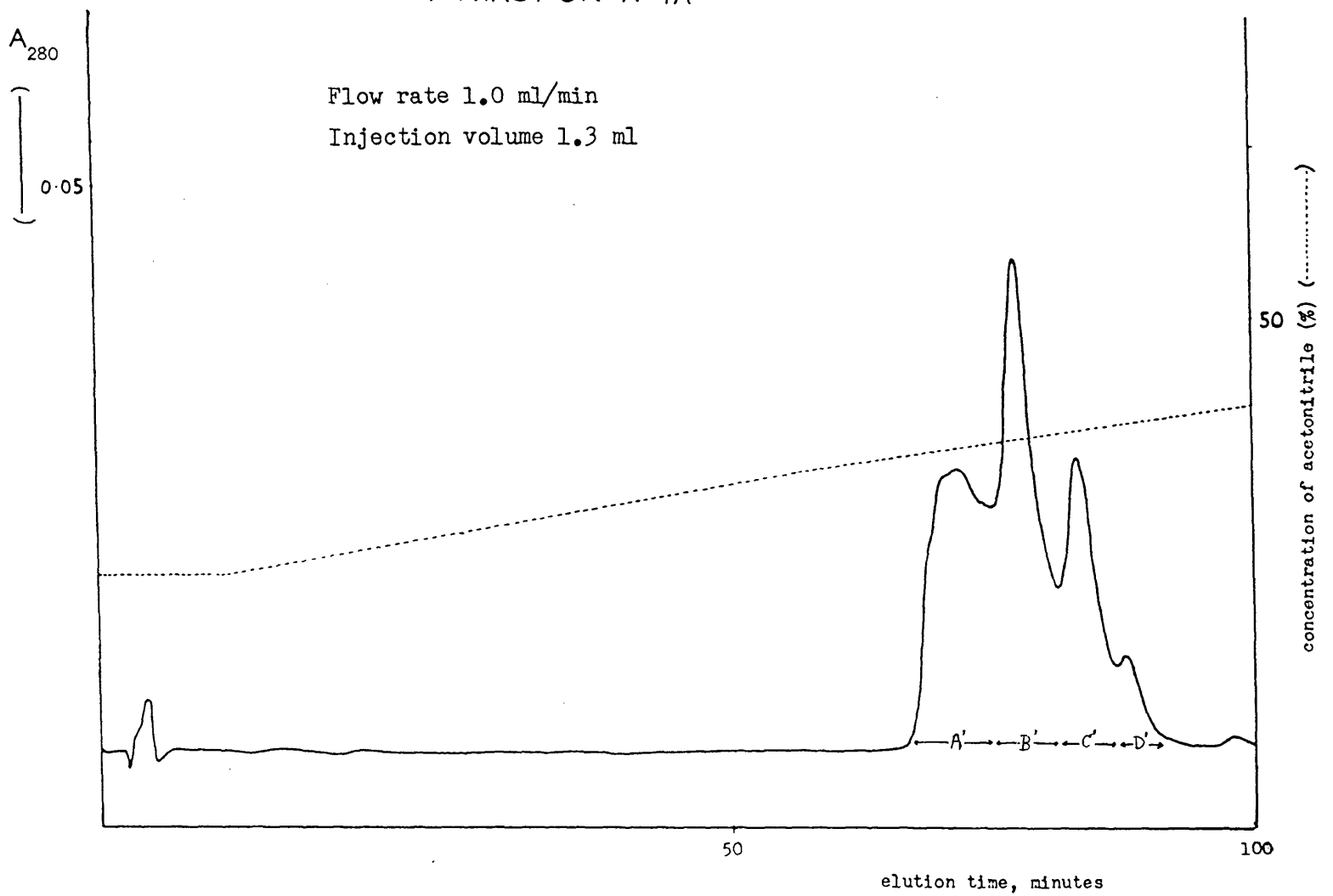

HPLC OF FRACTION IV-4A
Flow rate 1.0 ml/min
Injection volume 1.3 ml
A280
0·05
50
concentration of acetonitrile (%) (.....)
A'
B'
C'
D'
50
100
elution time, minutes

Fig 5.8

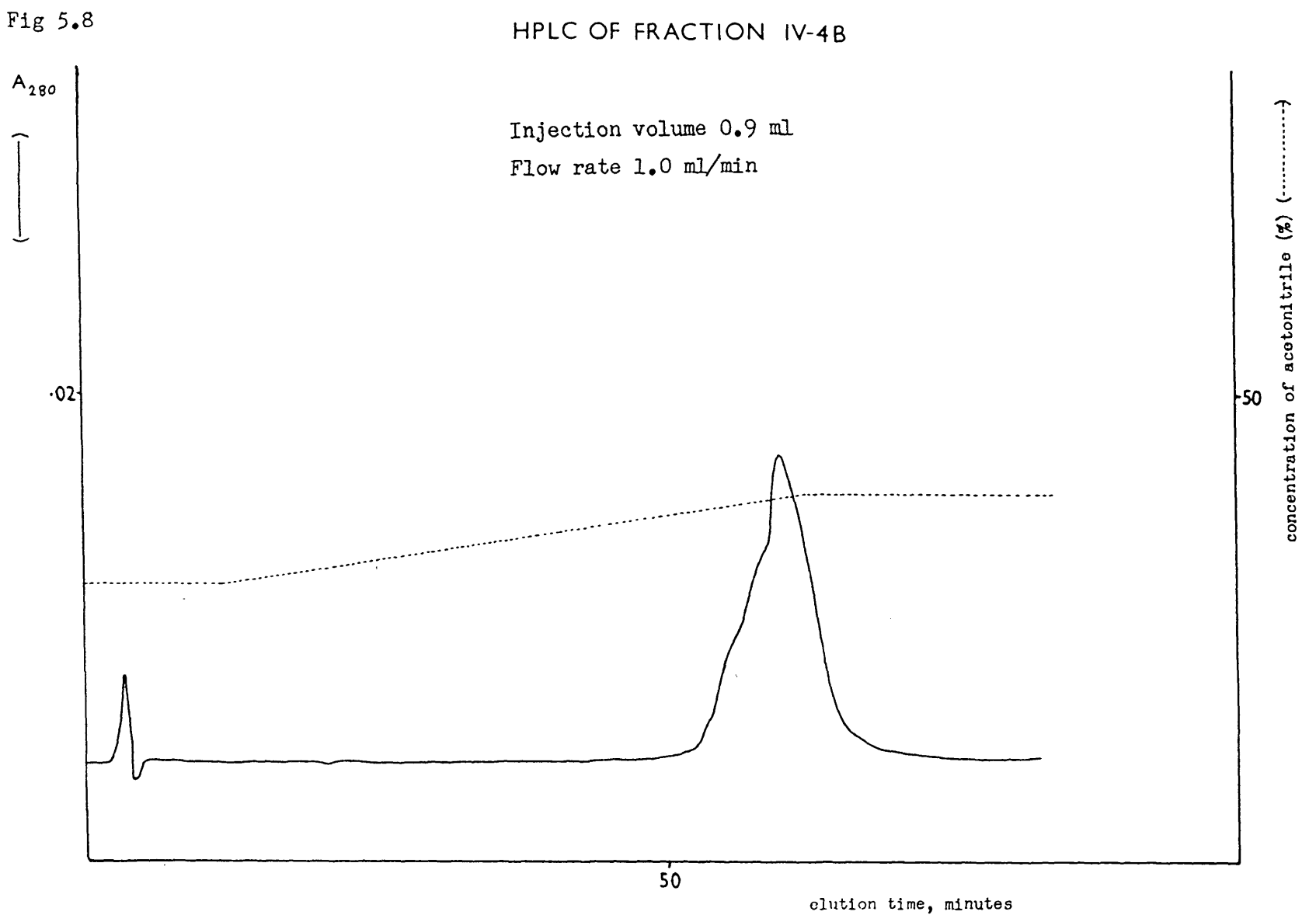
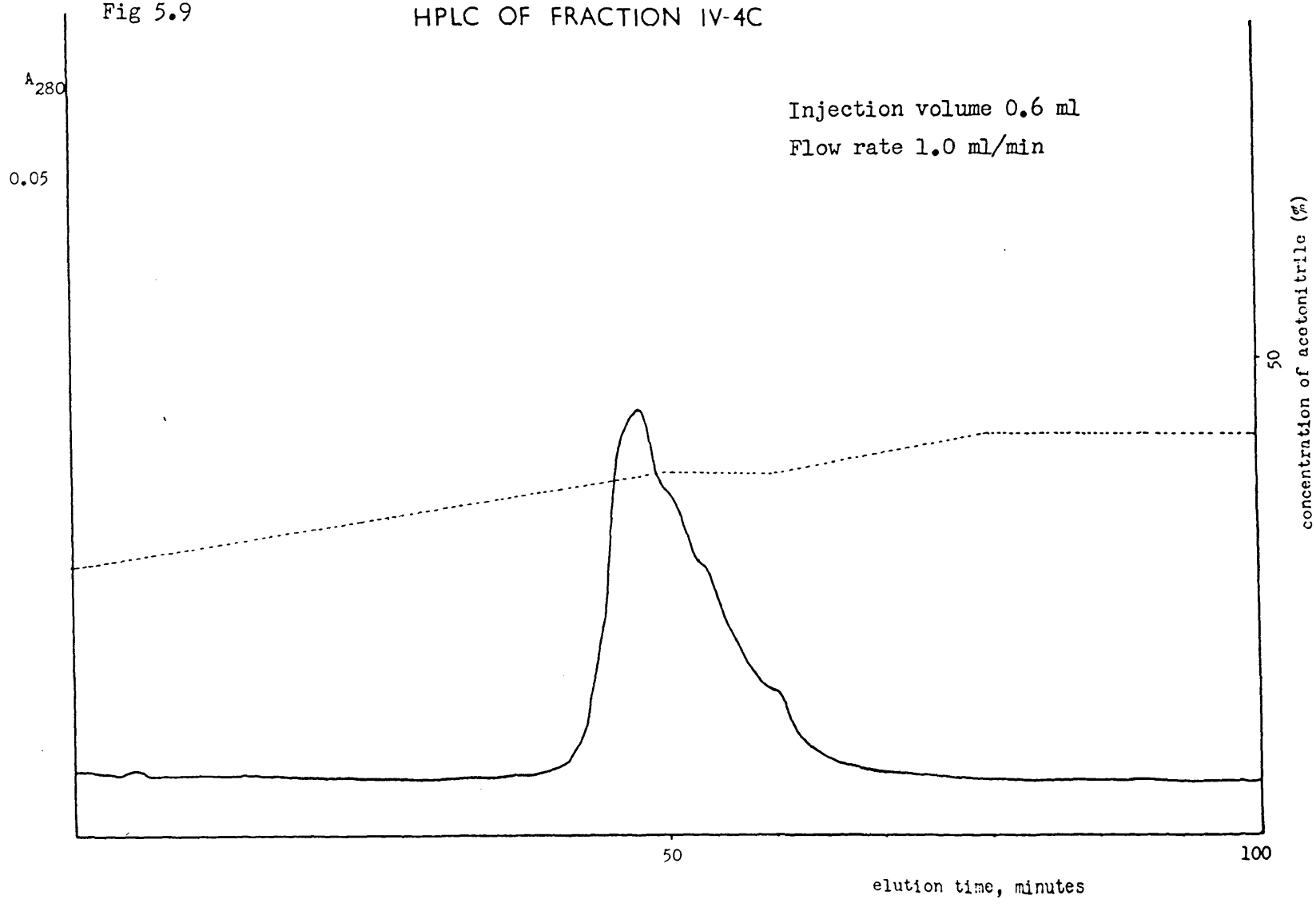

HPLC OF FRACTION IV-4B
Injection volume 0.9 ml
Flow rate 1.0 ml/min
A_{280}
·02
·50
50
elution time, minutes
concentration of acetonitrile (%) (.....)

Fig 5.9


HPLC OF FRACTION IV-4C
A_{280}
0.05
Injection volume 0.6 ml
Flow rate 1.0 ml/min
50
concentration of acetonitrile (%)
50
100
elution time, minutes

5.3.2 Chromatography of Pool V

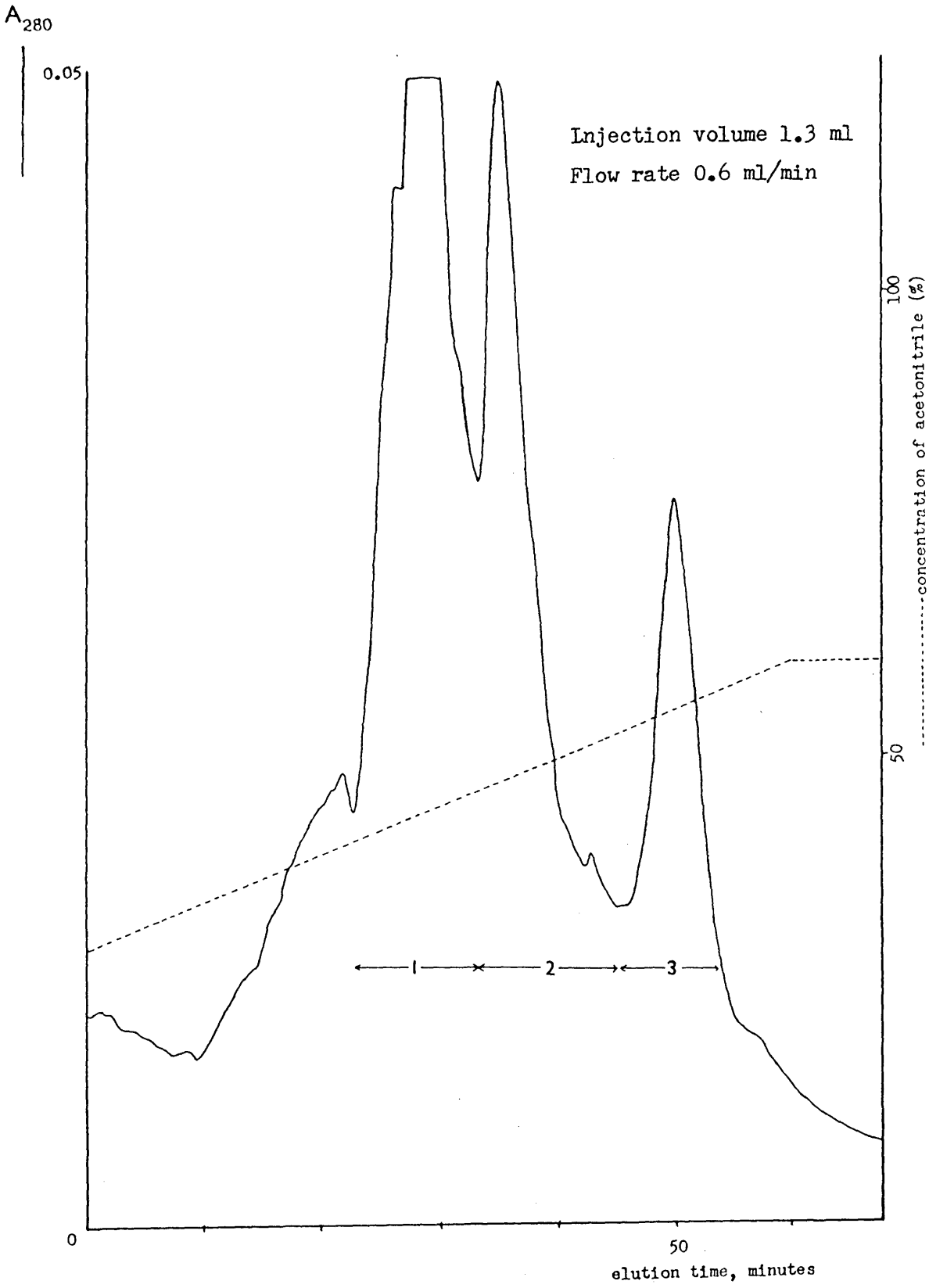
Because this pool represents a very small proportion of the total digested material, it was possible to purify its components by subjecting the entire pool to HPLC in a single run.

A linear gradient from 0 to 30% acetonitrile at 0.4% acetonitrile per minute produced only an unresolved smear, but further development of the column to 60% acetonitrile with the somewhat steeper gradient of 0.5% acetonitrile per minute produced three well-resolved components, which were termed V-1, V-2 and V-3 from their order of elution (fig. 5.10). The shape of the peak of V-1 suggested that it contained several components, but that they were so closely associated that rechromatography with a gentler gradient would be unlikely to achieve any well-resolved separation. A small aliquot, approximately 1/12, of V-1 was taken for preliminary sequencing to investigate the composition and quantity of this pool. The symmetry of the peaks of V-2 and V-3 suggested that each represented but a single species. Consequently the whole of each was taken for sequencing without further preparation. Each was indeed found to be substantially pure and was sequenced extensively.

The short trial run of peptide V-1 revealed that it contained large quantities of a single peptide with the N-terminal sequence:

Gly-Val-Ile-Phe-Gly

Fig 5.10: HPLC OF POOL V



5.3.3 Chromatography of Pool III

This pool was resolved by reverse-phase HPLC, as described in Methods. A linear gradient of acetonitrile from 30 to 60% at 0.5% per minute was employed. The peptides were thus resolved into three major components (fig. 5.11). A shoulder to the first peak was collected separately and was designated III-1.

Each of the four pools, III 1-4, was subjected to a further stage of purification by HPLC with a gentler gradient and reduced flow-rate to improve resolution (figs. 5-12-15). Peptide pools were made on the basis of the absorbance trace as indicated on the diagram.

Aliquots containing 10% of the material of each pool were analysed by sequencing with the short Quadrol program. The results of these studies indicated that only peptides III-2 and III-3B contained peptides suitable for extensive sequence determination. Peptides III-1B and III-1C possessed blocked N-termini and could not be analysed. The sequence of III-2 corresponded to no known sequence from hexokinase and may have been due to a contaminant of the enzyme as originally purified.

Because the average size of a peptide derived from cleavage of hexokinase by cyanogen bromide is expected to be ~40 residues, most sequence information will be derived from the peptides found in pools IV or V, as the products of complete cleavage by cyanogen bromide at each methionine residue. The rich variety of sequences obtained from these pools compared to the paucity of new sequences obtained from pool III indicates that this is indeed the case, and that the majority of the peptides purified do indeed derive from total internal cleavage. A sequence whose N-terminus places it simultaneously in pools III and IV (e.g. IV-6) may derive in the former from failure to cleave at the C-terminus of the latter.

Fig 5.11:

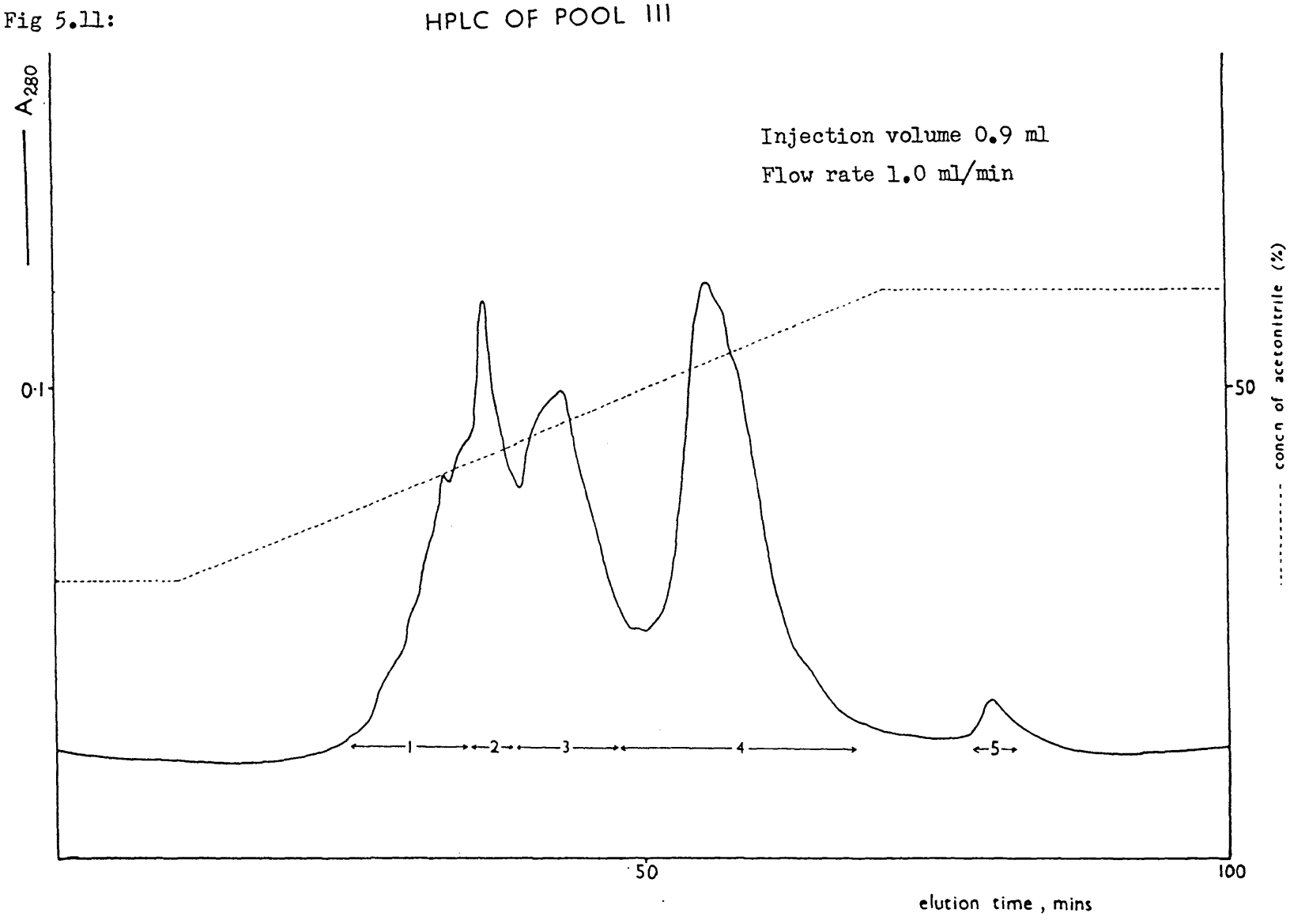

HPLC OF POOL III
Injection volume 0.9 ml
Flow rate 1.0 ml/min
A280
0·1
50
concn of acetonitrile (%)
1
2
3
4
5
50
100
elution time , mins

Fig 5.12

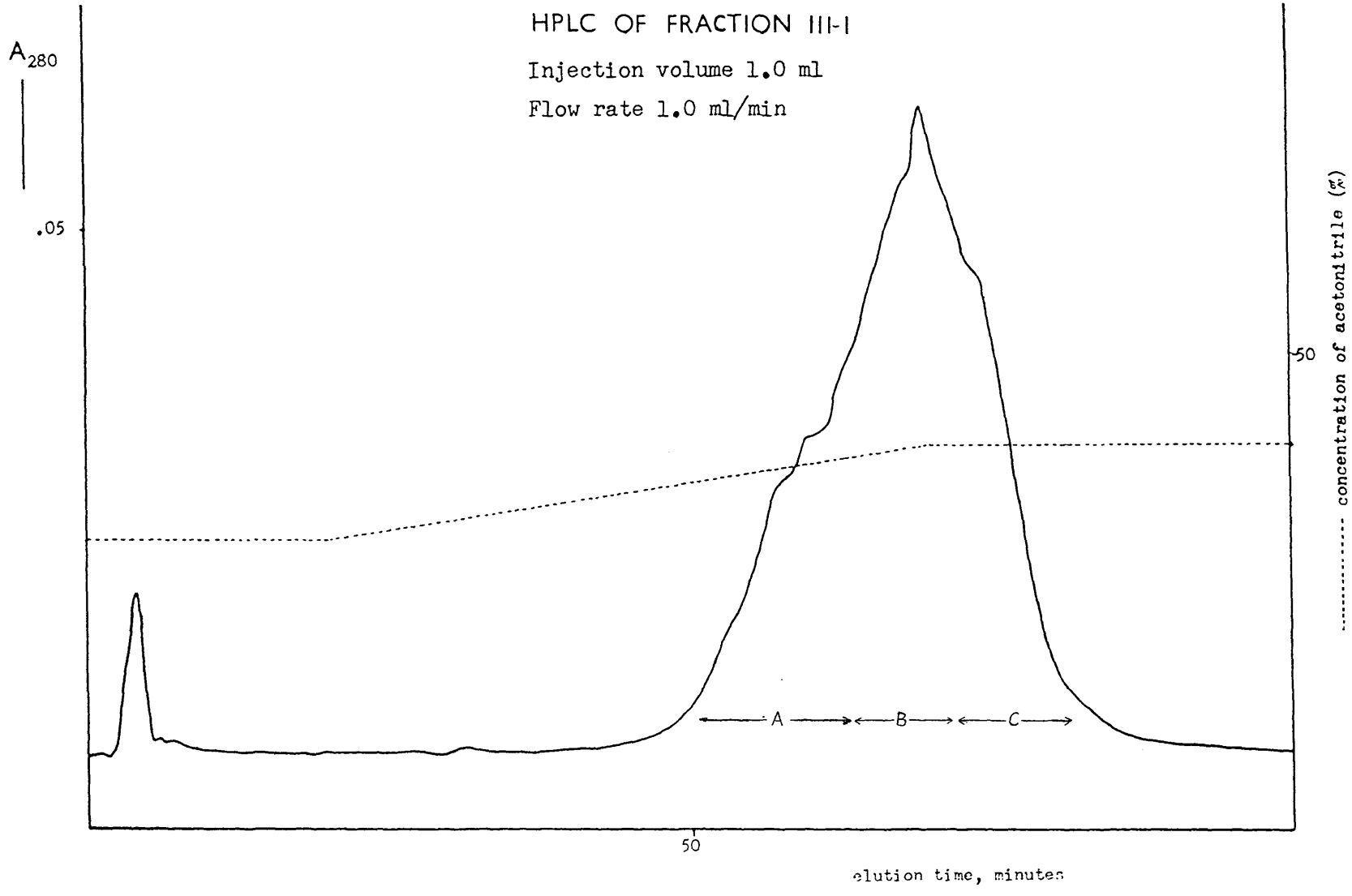

HPLC OF FRACTION III-1
Injection volume 1.0 ml
Flow rate 1.0 ml/min
A_{280}
.05
50
A
B
C
50
elution time, minutes
concentration of acetonitrile (%)

Fig 5.13

HPLC OF FRACTION III-2

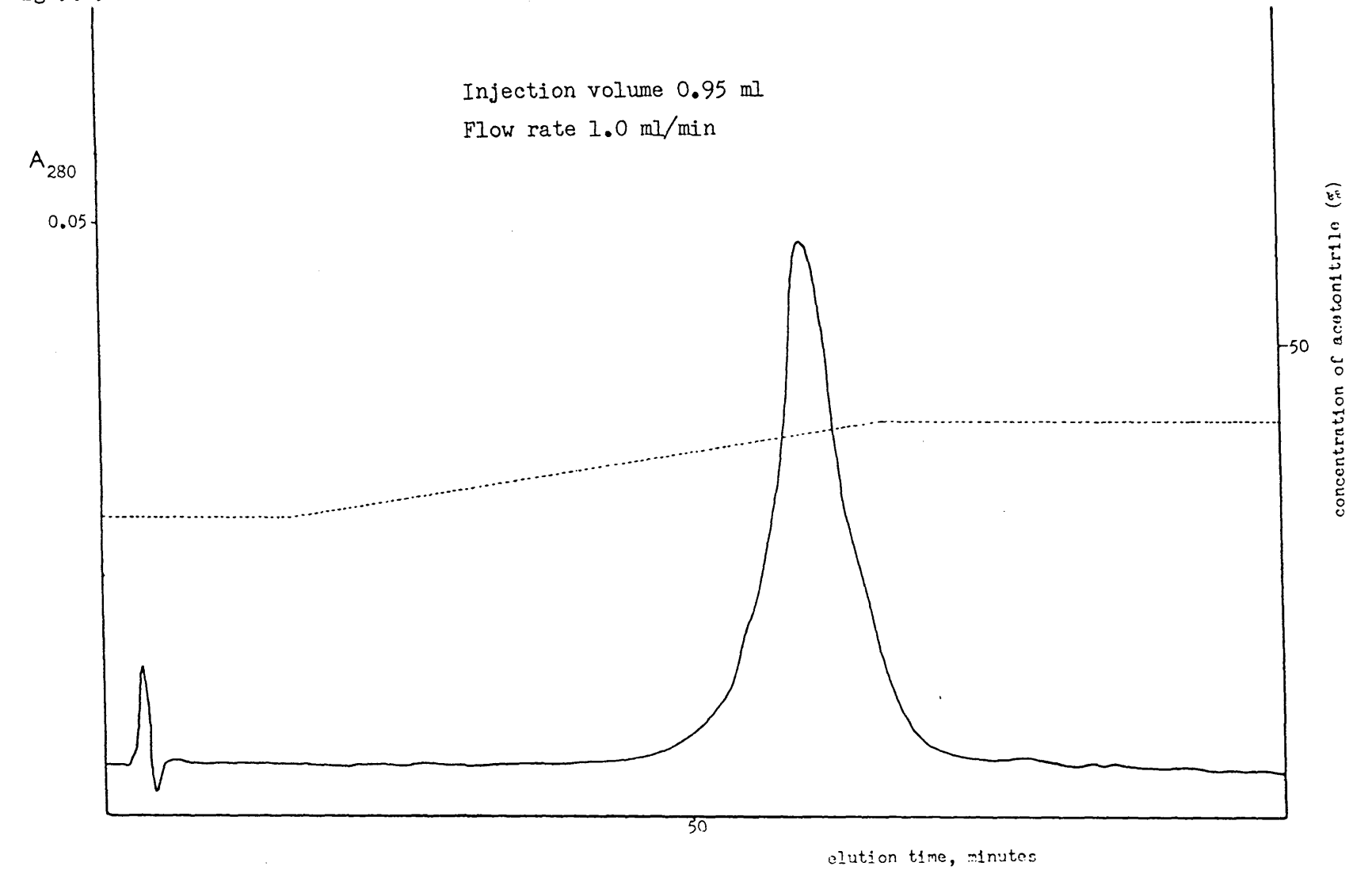


Fig 5.14

HPLC OF FRACTION III-3

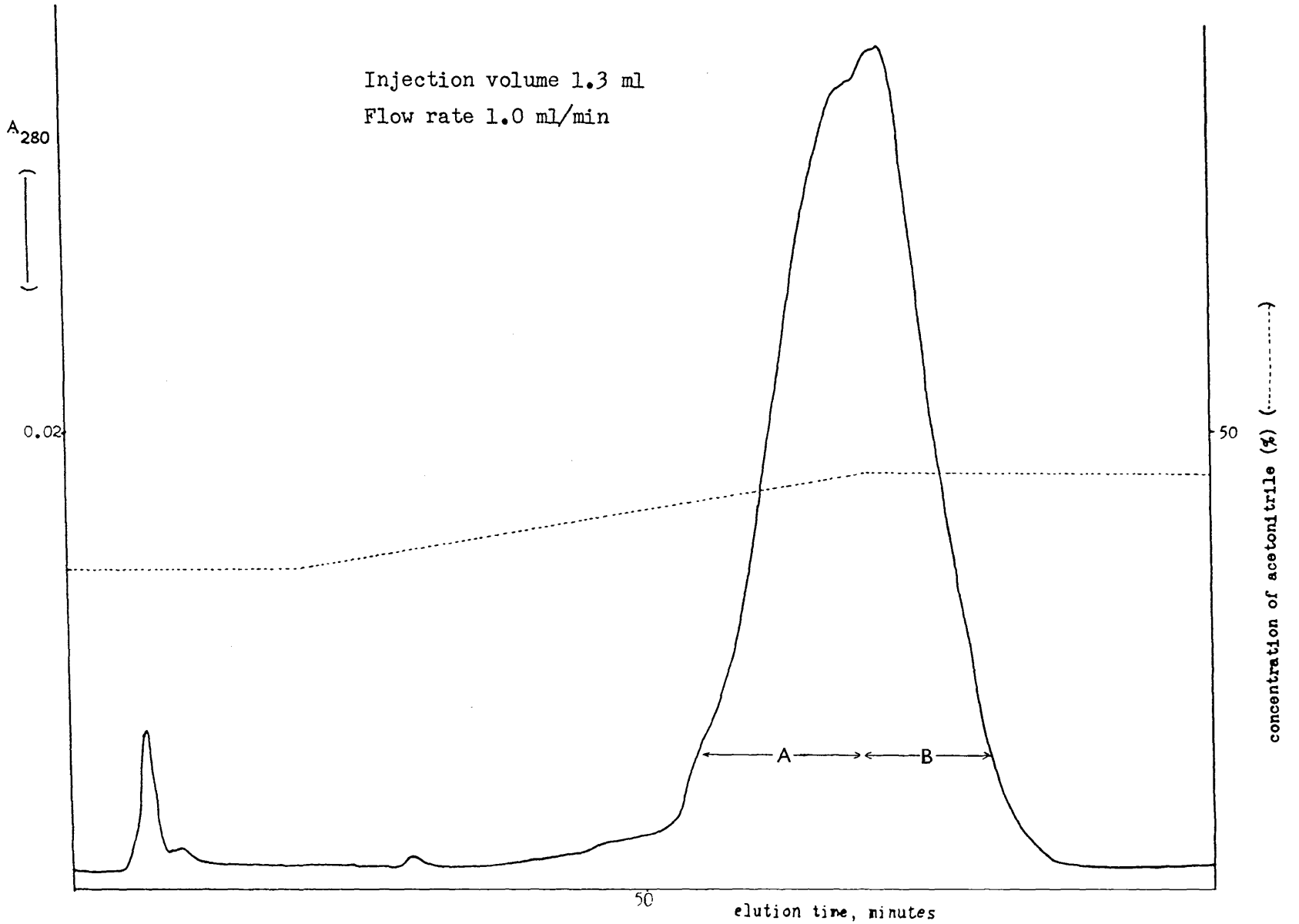
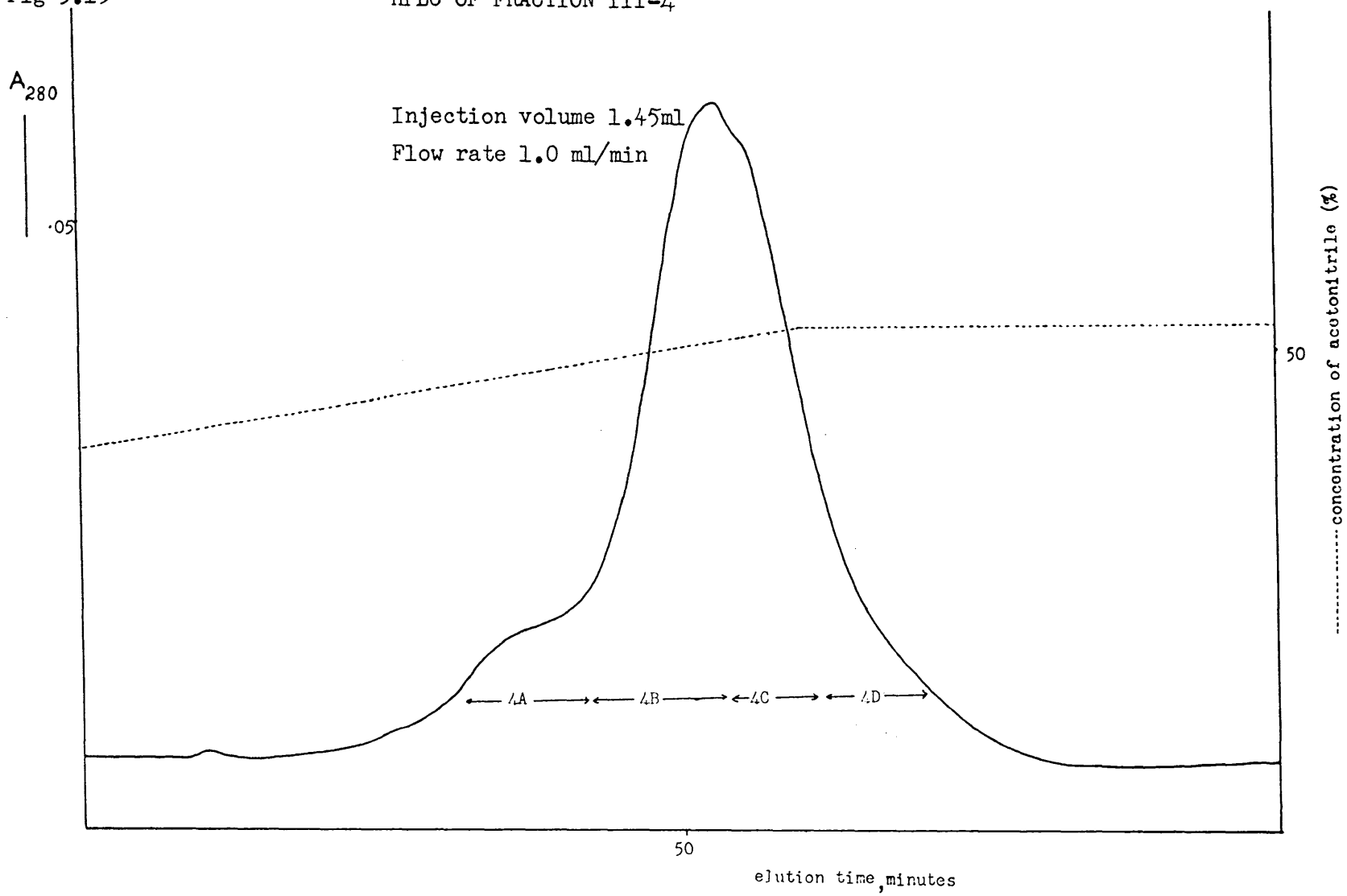


Fig 5.15

HPLC OF FRACTION III-4



5.3.4 Chromatography of Pool I1

This pool is expected to contain material of molecular wt ~20,000, corresponding to peptides 150-180 residues in length. It was resolved with a linear gradient of acetonitrile from 30% at 0.6%/min. Three major components were produced, of which the first two were further resolved by HPLC into single peaks. (fig. 5.16)

These two pools were analysed by short sequence runs. Both contained a single peptide, with the N-terminal sequence, Asp-Thr-Ser-Tyr-Pro-Ala, found in III-3B. They thus derive from cleavage at positions 316-7 and contain the C-terminal region of the enzyme.

Fig 5.16 HPLC OF POOL II

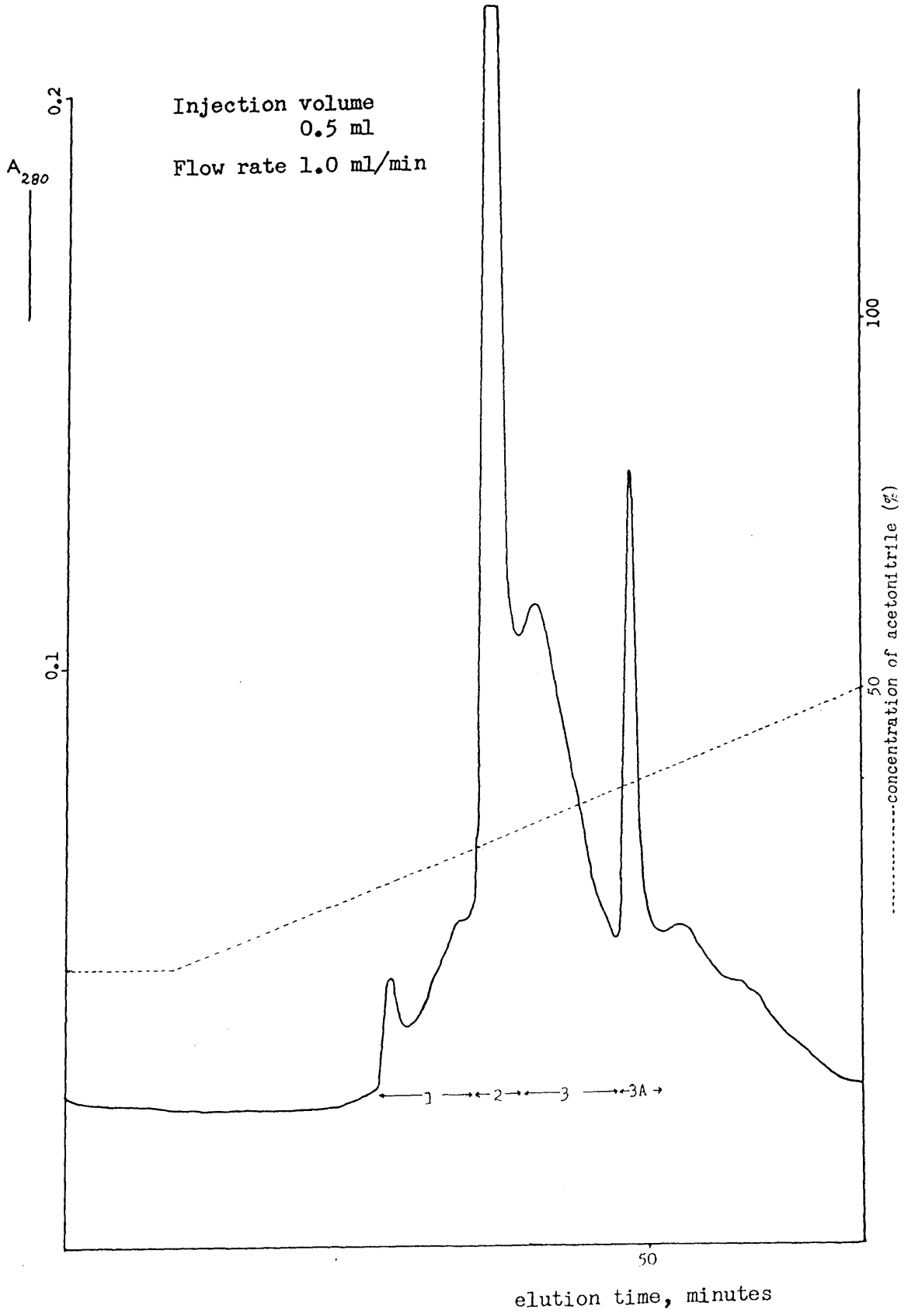


Fig 5.17

Amino acid nomenclature employed

Cysteine	Cys	C
Aspartic acid	Asp	D
Asparagine	Asn	N
	Asx	
Threonine	Thr	T
Serine	Ser	S
Glutamic acid	Glu	E
Glutamine	Gln	Q
	Glx	
Proline	Pro	P
Glycine	Gly	G
Alanine	Ala	A
Valine	Val	V
Methionine	Met	M
Isoleucine	Ile	I
Leucine	Leu	L
Tyrosine	Tyr	Y
Phenylalanine	Phe	F
Histidine	His	H
Lysine	Lys	K
Arginine	Arg	R
Tryptophan	Trp	W

With the exception of residue 189, the X-ray sequence fails to distinguish between free acids and amides. Consequently these residues are identified as Asx or Glx respectively in the X-ray sequence (q.v.).Diagrams employing the one-letter code similarly use the symbols D or E to indicate either form. In diagrams showing the correspondence between the X-ray and the chemically-determined sequences, where the X-ray sequence identifies such a residue correctly, the symbol appropriate to the acid or amide is used in that sequence, to indicate clearly the identity achieved.

γ, δ and ε are aliphatic sidechains of 2,3 and 4 carbon atoms respectively.

5.4 Peptide sequences obtained

PEPTIDE V-1

Estimated composition:

C_1 D_5 T_2 S_3 E_4 P_3 G_5 A_3 V_3 M_1 I_3 L_4 Y_2 F_2 K_2 H_1 R_1

(45 amino acid residues)

Quantity available:

6.5 nmoles

Sequence obtained:

Gly-Val-Ile-Phe-Gly-Thr-Gly-Val-Asn-Gly-
Ala-Tyr-Tyr-Asp-Val-Cys-Ser-Asp-Ile-Glu-
Lys-Leu-Gln (23 residues)

Repetitive yield:

98% (G(5) - G(7))

Comments:

The amount of material available for N-terminal formation fell sharply after step 8 as internal cyclisation at Asn(9)-Gly(10) blocked the peptide N-terminus.

This peptide could be allocated to positions 207-229 of the X-ray sequence. It contains one cysteineresidue, at position 222 of that sequence, which corresponds approximately to the cysteine residue proposed for position 220.

PEPTIDE V-2

Estimated composition:

$D_4\ T_2\ S_4\ E_4\ P_1\ G_5\ A_3\ V_2\ I_3\ L_7\ Y_2\ F_1\ K_1\ H_1\ R_1$

(41 amino acid residues)

Quantity available:

9 nmoles

Sequence obtained:

Gly-Ser-Gly-Tyr-Tyr-Leu-Gly-Unk-Ile-Leu-

Arg-Leu-Ala-Leu-Val-Unk-Unk-Asp-Ile-Glu

Asp-Leu-Gln-Gly (24 residues)

Repetitive Yield:

95% (G(3) - G(7))

Comments:

The sequence deteriorated rapidly beyond step 15 and must be considered unproven beyond this point. The short length of sequence available makes alignment with the X-ray sequence difficult (see discussion).

PEPTIDE V-3

Estimated Composition:

$D_8\ T_3\ S_3\ E_{10}\ P_5\ G_5\ A_3\ V_2\ I_6\ L_5\ Y_1\ F_4\ K_4\ H_1\ R_1$
(61 amino acid residues)

Quantity present:

3 nmoles

Sequence obtained:

Arg-Thr-Thr-Gln-Asn-Pro-Asp-Glu-Leu-Leu-
Unk-Phe-Ile-Ala-Asp- (15 residues)

Repetitive yield:

87% (Q(4) - N(5))

Comments:

This is the same peptide as IV-6, for which a better sequence was obtained. The peptide was lost rapidly from the cup after step 15, for reasons that remain unclear, since IV-6 was not extracted in this way.

PEPTIDE IV-1A

Estimated composition:

$D_6\ T_3\ S_2\ E_8\ P_5\ G_4\ A_2\ V_2\ M_1\ I_4\ L_2\ Y_1\ F_3\ K_3\ H_1\ R_1$
(48 amino acid residues)

Quantity available:

12 nmoles

Sequence obtained:

Ala-Ile-Asn-$\overset{CM}{Cys}$-Glu-Tyr-Gly-Ser-Phe-Asp-

Asn-Glu-Ser-Val-Val-$\overset{Thr}{Leu}$-Pro-Pro-Thr-Unk-

Tyr-Asp-Ile-Thr-Asn-Unk-Asp-Glu-Ser-Gly-

Asn-Pro-Gly-Gln-Pro-Phe (36 residues)

Repetitive yield:

95.4% (I(2) - N(3))

Comments:

This peptide was treated with Braunitzer's reagent to make it more hydrophilic. The effect of the reagent was to retain the peptide in the cup until sequence determination was terminated by overlap rather than by washout.

The peptide could be aligned with positions 241-277 of the X-ray sequence. It contains one cysteine residue, believed to be an essential component of the active site of hexokinase.

PEPTIDE IV-2C

Estimated composition:

$D_{11}\ T_2\ S_4\ E_{10}\ P_4\ G_6\ A_2\ V_3\ M_1\ I_3\ L_3\ Y_2\ F_7\ K_{10}\ H_1\ R_1$

(70 amino acid residues)

Quantity available:

12 nmoles

Sequence obtained:

Tyr-Lys-Gln-Gly-Phe-Ile-Phe-Lys-Asn-Gln-

Asp-Leu-Ser-Trp-Phe-Asp-Lys-Asn-Val-

(19 residues)

Repetitive yield:

96% (F(5)-F(7))

Comments:

Despite treatment with Braunitzer's reagent this peptide was lost rapidly from the cup after step 16. It is very hydrophilic up to this point and presumably hydrophobic after, resulting in rapid and ready loss.

Step 14 was identified as Tryptophan because none of the residues known to produce weak PTH-amino acid signals (serine, threonine and ε-SPTC lysine) were present.

This peptide could be aligned with positions 294-315 of the X-ray sequence, although the N-terminal region contains two positions in the X-ray sequence which, when the two sequences are compared, have no equivalent in the chemically-derived sequence:

294

X-ray sequence : M δ F K δ δ G E

IV-2C : (M) - Y K - Q G F

Peptide IV-2

The results obtained from peptide IV-2 confirmed the sequence obtained from peptide IV-2C.

Estimated composition

D_{11} T_2 S_4 E_{12} P_6 G_2 A_3 V_3 M_1 I_3 L_3 Y_2 F_8 K_{10} H_1 R_1
(74 amino acid residues)

Quantity available

8 nmoles

Sequence obtained

Tyr-Lys-Gln-Gly-Phe-Ile-Phe-Lys-Asn-Gln-
Asp-Leu-Ser-Trp-Phe-Asp-Lys-Asn-Val-
(19 residues)

Repetitive yield

94% (F(5)-F(7))

PEPTIDES IV-4A'/IV-4B/IV-4B'

IV-4A'

Estimated composition:

$D_9\ T_5\ S_4\ E_7\ P_4\ G_8\ A_5\ V_5\quad I_3\ L_7\ Y_1\ F_3\ K_5\ H_1\ R_1$
(68 amino acid residues)

Quantity present:

6 nmoles

Sequence obtained:

Asp-Phe-Pro-$\frac{\text{Leu}}{\text{Thr}}$-$\frac{\text{Gly}}{\text{Phe}}$-Glu-Glu-$\frac{\text{Val}}{\text{Ser}}$-Ala-Ser-

Ala-Val-Asn-Ile-Unk-Asp-Gly

(17 residues)

Repetitive yield:

85% (estimated)

Comments:

This peptide was treated with Braunitzer's reagent and was lost due to a technical malfunction after step 18. The extent of heterogeneity is considerable and may be due to contamination, possibly (although this is unlikely) with the corresponding sequence from hexokinase A. The yield could not be determined with any accuracy since the heights of the peaks of the PTH-analysis showed no consistent trend. It seems that either extraction or cleavage was poor and this caused the yield at each step to vary.

The peptide could be aligned with residues 174-190 of the X-ray sequence. This sequence proposes a methionine for residue 170, yet the peptide clearly commences with a residue equivalent to γ-174 of the X-ray sequence. Consequently the three residues 171-3 of the X-ray sequence have no chemically-derived equivalent.

	170									
X-ray sequence:	M	δ	Dx	A	γ	F	P	A	G	D
IV-4A':	(M)	-	-	-	D	F	P	L/T	G/F	E

In an attempt to eliminate or reduce the ambiguous sequence assignments, Pool IV-4B, which contained the same peptide in good yield, was sequenced. This peptide did not cleave well between residues 2 and 3, and so although some of the ambiguous assignments in the early part of the sequence were removed, others were introduced in the later part.

Sequence obtained:

Asp-Phe-Pro-Thr-Gly-Lys-Glu-Ser-Unk-Ser-

Phe/Ala-Ile-Ala-Ile-Asp-Unk-Gly-Gly

Peptide pool IV-4B' contains both this sequence and the sequence present in pool IV-4C, in which it is pure. Both sequences could thus be read off simultaneously. Due again to incomplete cleavage between residues 2 and 3, the sequence of interest was not determined unambiguously, but a number of the unidentified or incompletely identified residues of IV-4A' and IV-4B were characterised.

Sequence obtained:

Asp-Phe-Pro-Thr-Gly-Unk-Glu-Unk-Unk-Ser-

Phe-Leu-Unk-Ile-Asp-Asp-Gly-Gly-Thr

It is proposed that the sequence of this region of the hexokinase molecule is

Asp-Phe-Pro-Thr-Gly-Glu-Glu-Ser-Unk-Ser-

Phe-Leu-Unk-Ile-Asp-Asp-Gly-Gly-Thr-

This sequence is formed by combining the most accurately and clearly determined regions of the sequences of IV-4A', -4B and -4B

PEPTIDE IV-4C

Estimated composition:

$D_6\ T_4\ S_2\ E_7\ P_3\ G_4\ A_3\ V_4\ I_4\ L_5\ Y_1\ F_1\ K_2\ H_1\ R_1$

(48 amino acid residues)

Quantity available:

20 nmoles

Sequence obtained:

Leu-Gln-Lys-Gln-Ile-Thr-Lys-Arg-Asn-Ile-

Pro-Ile-Glu-Val-Val-Ala-Leu-Ile-Asn-Unk-

Thr-Thr-Unk-Gly-Leu-Val-Unk-Ala-Tyr

(29 residues)

Repetitive yield:

91% (I(10)-I(12))

Comments:

Preliminary sequence studies had shown that this peptide was not lost significantly from the sequencer cup within the first 10 residues. Together with the comparatively large quantity available and the low lysine content this influenced the decision not to use Braunitzer's reagent.

It is proposed to align with residues 140-170 of the X-ray sequence but the extent of homology is poor (discussion and fig. 8.4). It may also overlap with the nonapeptide sequence found by Kuromizu et al (1979) (and fig. 8.5).

PEPTIDE IV-4C

This pool contained minor components of N-terminal glycine and isoleucine. Inspection revealed the former to be identical to peptide V-1 (q.v.). The following sequence was obtained from the latter. There are estimated to be only 2 nmole of this peptide.

Sequence obtained:

Ile-Asp-Gly-Tyr-Val-Glu-Lys-Glu-Ile-Asn-

Asp-Leu-Glu-Asp-Val-Asn-Phe-Gly

(18 residues)

No reliable estimate was possible of the repetitive yield.

Comments:

Despite the small quantity of peptide available, the sequence is well defined as far as position 12, and from positions 16-18.

It corresponds remarkably well to the X-ray sequence, following methionine-47. It probably derives from cleavage at this point and at methionine-99. The truncated product expected to be derived from cleavage at methionine-69 would fall into pool VI.

PEPTIDE IV-6

Estimated composition:

$D_4\ T_3\ S_2\ E_8\ P_5\ G_4\ A_2\ V_2\ I_4\ L_3\ Y_1F_4\ K_3\ H_1\ R_1$

(47 amino acid residues)

Quantity available:

7 nmoles

Sequence obtained:

Arg-Thr-Thr-Gln-Asn-Pro-Asp-Glu-Leu-Trp-
Glu-Phe-Ile-Ala-Asp-Ser-Leu-Phe-Ala-Phe-
Ile-Thr-Glu-Gln-Phe-Pro-Pro-Gly-Ile

(29 residues)

Repetitive yield:

92% (Q(4)-N(5))

Comments:

This peptide was not treated with Braunitzer's reagent in an attempt to resolve the N-terminal signal, where serine was present. It is known that the sequence methionine-serine is somewhat resistant to cleavage by cyanogen bromide, due to formation of an internal iminolactone (Schroeder et al, 1969).

Tryptophan at position 10 was identified as a gap in the sequence, with no signal due to serine.

This peptide could be aligned readily with positions 100-129 of the X-ray sequence.

PEPTIDE IV-6 (Minor Component)

This pool contained approximately 2·5 nmoles of a peptide with N-terminal alanine.

Sequence obtained:

Ala-Asp-Val-Pro-Lys-Glu-Lys-Thr-Gln-Gln-
Ile-Val

Repetitive yield:

87%

Comments:

Beyond position 12 the sequence was lost in the background due to the major component.

This sequence overlaps with an internal cyanogen bromide cleavage-derived peptide found within a peptide derived from cleavage by hydroxylamine (see section 6.4) and with the sequence found from intact hexokinase (see section 2.2). It contains the residue proposed by Anderson et al (1978b) as the N-terminus of the X-ray sequence.

																	10
X-ray sequence:											A	A	A	γ	D	δ	S L V E V
Hydroxylamine peptide:	N	G	S	M	A	D	V	P	K	E	L	M	Q	L	I	E	N
IV-6 (minor component):					A	D	V	P	K	E	K	T	Q	Q	I	V	

No methionine was observed at step 8 of this peptide (which corresponds to position 7 of the X-ray sequence), whereas the amino acid sequences derived from the whole protein (p. 39, q.v.) and peptide 1 from cleavage by hydroxylamine (p. 99, q.v.) definitely showed it to be there.

PEPTIDE III-3B

Estimated composition:

$D_9\ T_4\ S_3\ E_9\ P_4\ G_7\ A_6\ V_4\ M_1\ I_6\ L_6\ Y_3\ F_3\ K_4\ H_2\ R_2$

(73 amino acid residues)

Quantity available:

3 nmoles

Sequence obtained:

Asp-Thr-Ser-Tyr-Pro-Ala-Arg-Ile-Glu-Unk-
Asp-Pro-Phe-Phe-Asn-Leu-Leu-Asp-Ala.

(19 residues)

Repetitive yield:

96% (Y(4)-A(6))

Comments:

This peptide was not treated with Braunitzer's reagent.

The peptide was readily allocated to position 317-355 of the X-ray sequence. There is a remarkably high proportion of correctly-identified residues in the X-ray sequence. It is perhaps then curious that methionine-316 is not identified correctly.

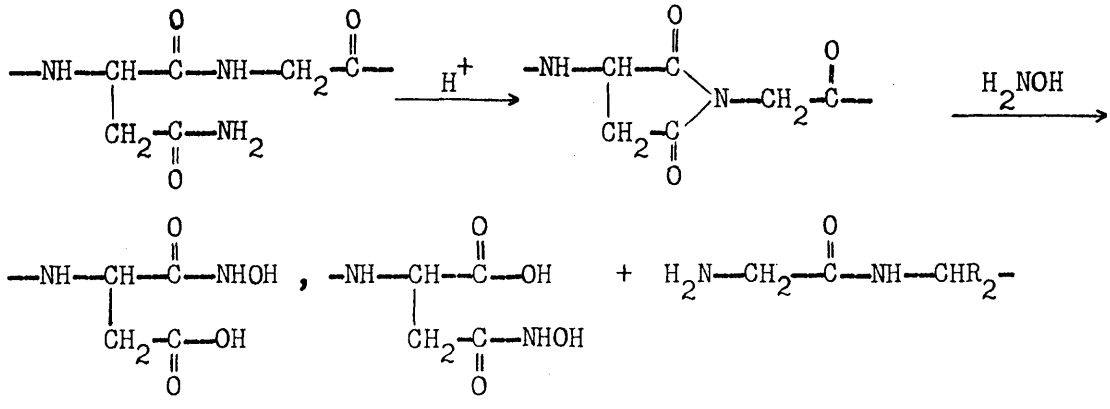
CHAPTER 6

CLEAVAGE OF HEXOKINASE WITH HYDROXYLAMINE

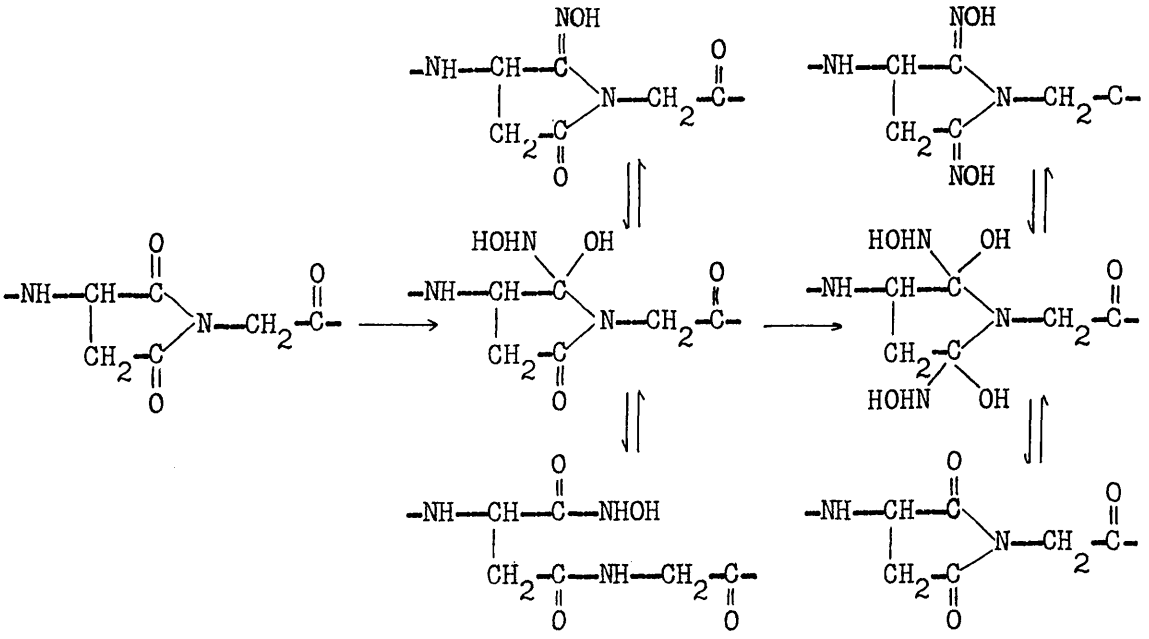
6.1 Introduction and rationale

If all the amino acids occurred in peptides in random arrangement of equal quantities, the sequence Asn-Gly should not occur with a probability exceeding 1 in 400. Because amino acids are not found to occur in equimolar quantities this probability of occurrence will not be strictly observed; nevertheless, a protein molecule the size of yeast hexokinase ($\sim$450 amino acid residues) should not possess more than two such units. Cleavage at this linkage by hydroxylamine thus offers the prospect of obtaining large fragments suitable both for Edman degradation from a new start point and sub-digestion by a means other than cyanogen bromide, and of thus obtaining the necessary overlaps of the peptides derived from a cleavage by cyanogen bromide.

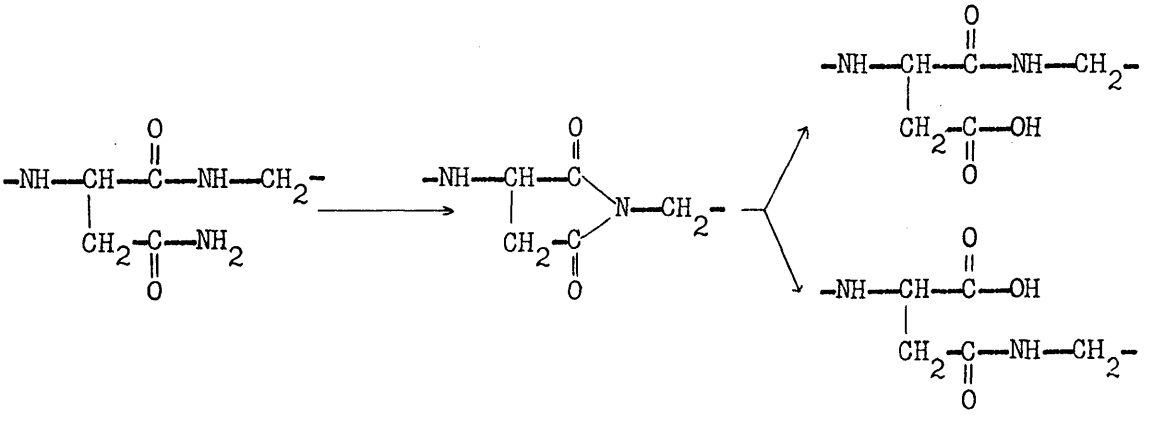
The mechanism of hydroxylamine cleavage is believed to involve formation of an internal cyclic imide of an Asn-Gly bond (Bornstein, 1970) which is then susceptible to nucleophilic attack by hydroxylamine. Formation of the imide is favoured by low pH, so any procedure prior to such cleavage of a protein should involve a stage of exposure to acidic conditions (pH 3.0).



The mechanism may involve formation of first the mono-, then the di- hydroxylamine adduct, each in equilibrium with its oxime. Rearrangement of the mono- adduct leads to formation of the mono-hydroxamate; of the di-adduct to the dihydroxamate with simultaneous release of a new peptide glycine N-terminal:



It is conceivable that base-catalysed cyclisation could exist, but the OH^- ion would then compete in hydrolysis of the cyclic imide to replace the amide with a carboxyl group:



α- and β-aspartyl glycyl peptides

The peptides produced upon digestion of hexokinase with cyanogen bromide are mostly too large to be sequenced completely by the automated Edman degradation with the chemicals available. This digestion is also very likely to be incomplete if any of the methionine residues have been oxidised beyond the sulphoxide. The reductant, dithioerythritol, employed in carboxymethylation, is insufficiently strong to reduce the methionine sulphone back to methionine. To attempt a complete sequence determination requires another digestion and fractionation.

The amino acid composition of hexokinase (Schmidt and Colowick, 1973a, and fig. 1.1) suggests that the use of proteolytic enzymes such as Trypsin, Chymotrypsin or staphylococcal protease would produce too many fragments for convenient separation.

The complexity of any proteolytic subdigestion can be reduced by first performing a digestive technique of high specificity to produce a small number of large fragments, which are subsequently employed as substrates for a less specific subdigestion.

It is known that under alkaline conditions hydroxylamine can cleave proteins at Asn-Gly linkages with a high degree of specificity (Bornstein, 1969). Two such sequences have been detected in the hexokinase molecule - one at residues 13 and 14 from the N-terminus, identified in the sequence obtained from the intact protein, and one at positions 215-6 of the X-ray sequence, identified in the sequence of peptide V-1 (see p. 75). from cyanogen bromide cleavage. It is highly unlikely that there is more than one more such sequence in the hexokinase molecule. Cleavage by hydroxylamine thus offers the prospect of obtaining peptide fragments of up to 200 residues in length

that are suitable for providing peptide overlaps both by Edman degradation from a new N-terminus and as a basis for subsequent subdigestion.

6.2 Materials and Methods

Materials:

Reduction: Guanidine hydrochloride was ultrapure grade from Schwarz-Mann. Dithioerythritol and Tris (Trizma base) were from Sigma Chemical Co.

Carboxymethylation: Radioactive iodoacetic acid was from the Radiochemical Centre Ltd. (now Amersham International) and was present in sealed vials containing 250 µCi at a specific activity of 54 µCi/µmol. Unlabelled iodoacetic acid was from Sigma Chemical Co. It was recrystallised twice from petroleum ether before use. Sodium hydroxide was from a 0.1N Volumetric solution from BDH.

Cleavage: Hydroxylamine hydrochloride was Grade I from Sigma. Lithium Hydroxide was Sepramar grade from BDH.

Gel Filtration: Urea was Aristar grade from BDH.

Gel Permeation: Propan-1-ol was HPLC grade from BDH.

Reduction

1 µmole of pure hexokinase B was dialysed into a solution of 6M guanidine-HCl/0.3M Tris-HCl pH 8.2 to a total volume of 10ml. The solution was made up using doubly distilled water that had been thoroughly deaerated. Dithioerythritol dissolved

in the same buffer was added to two-fold molar excess of reducing agent over protein thiol groups. The flask was flushed with nitrogen and was incubated for 3 hours at room temperature. Native hexokinase does not contain any disulphide bridges,which require a much longer incubation for cleavage. A short exposure to reducing conditions ensures that none can form.

Carboxymethylation

The reduced protein was carboxymethylated with sodium iodoacetate made radioactive to a specific activity of 1.5 μCi/μmol. Radioactive iodoacetic acid in slight molar excess over reducing thiol groups was dissolved in an equimolar quantity of 0.1N sodium hydroxide solution to yield the neutral salt. This aliquot was added to the reduced protein solution, which was incubated in the dark under nitrogen for 1 hour.

Reaction was terminated by addition of a slight excess of dithioerythritol. The reagents were removed from the protein by dialysis under reduced pressure, first into water and then into 10% formic acid. Treatment with acid promotes formation of internal cyclic imides as described. The solution was lyophilised.

Cleavage

Cleavage of hexokinase with hydroxylamine was carried out according to the procedure of Bornstein and Balian (1977). The lyophilised protein was dissolved in a solution of 6M guanidine-HCl/2M hydroxylamine-HCl adjusted to pH 9.0 with 4.5M (sat) lithium hydroxide. The solution was incubated at 45°C for 4 hours. The pH of the solution was monitored closely during the reaction and was readjusted to pH 9.0 with dropwise additions of 4.5M LiOH. Upon completion of the

time course reaction was terminated by acidification with formic acid to a pH of 3.5. The solution was stable and could be stored at 4°C.

The use of 6M guanidine in the reaction mixture ensures that the protein substrate is fully soluble and that all susceptible sites are exposed to reagent. Use of a pH below 10 reduces competing nucleophilic attack by hydroxide ion which would lead to opening of the cyclic imide without concomitant chain cleavage, to form the unsusceptible α- and β-aspartyl glycyl peptides, in which the amide is replaced by a carboxyl group.

Lithium hydroxide is employed as titrant because of the high solubility of lithium chloride.

Peptide chain separation

Gel filtration was carried out on a column (85 x 2.5cm) of Sephadex G-100 (Whatman Ltd., Maidstone, Kent, UK) equilibrated with 2M Urea/0.1M formic acid.

Ultra filtration was performed in an Amicon cell employing an Amicon UM-2 membrane, which is claimed to be impermeable to molecules whose molecular weight exceeds 2000.

Gel permeation chromatography was done on a double Waters I-125 protein analysis column (Waters Associates, Milford, Mass., U.S.A.) in 5% (v/v) acetic acid/10% (v/v) n-propanol. The instrumentation was as described in 4.3.3.

6.3 Purification of peptides

The acidified solution of the hydroxylamine digest was separated by gel filtration into two incompletely resolved pools (fig. 6.1). The volume of the unfractionated digest required four separate rooms to be made. The Sephadex column was reequilibrated and repacked between each run in order to prevent loss of resolution. The eluate was separated at the position of local absorption minimum, as indicated, into pools of high and low molecular weight material which are designated 'A' and 'B' respectively on fig. 6.1. The former was believed to contain principally undigested hexokinase, the latter the products of digestion. Each pool was concentrated by ultrafiltration.

Each pool was analysed by SDS-PAGE, together with a sample from the unfractionated digest. This latter showed three components, of which the largest and smallest were present in the high and low molecular weight pools, respectively, while the component intermediate in molecular weight was common to both.

The pool of high molecular weight material, pool A, was gel filtered once more under the same conditions (fig. 6.2(i)). It was thus resolved into two major pools, A-1 and A-2, separated at the local absorption minimum as indicated. Pool A-2 was concentrated by ultrafiltration and was added to pool B. This composite pool was also gel filtered under the same conditions (fig. 6.2(ii)). It was thus resolved into a smaller peak, probably enriched in undigested material, and a larger one, probably containing largely the products of digestion. This larger peak clearly contained most of the peptide and protein material produced by digestion. It was concentrated by ultrafiltration before being made the subject of another

Fig 6.1: Gel filtration of Hydroxylamine digest

Sephadex G-100 in 2M Urea, 0.1M Formic acid
Loading volume 6ml

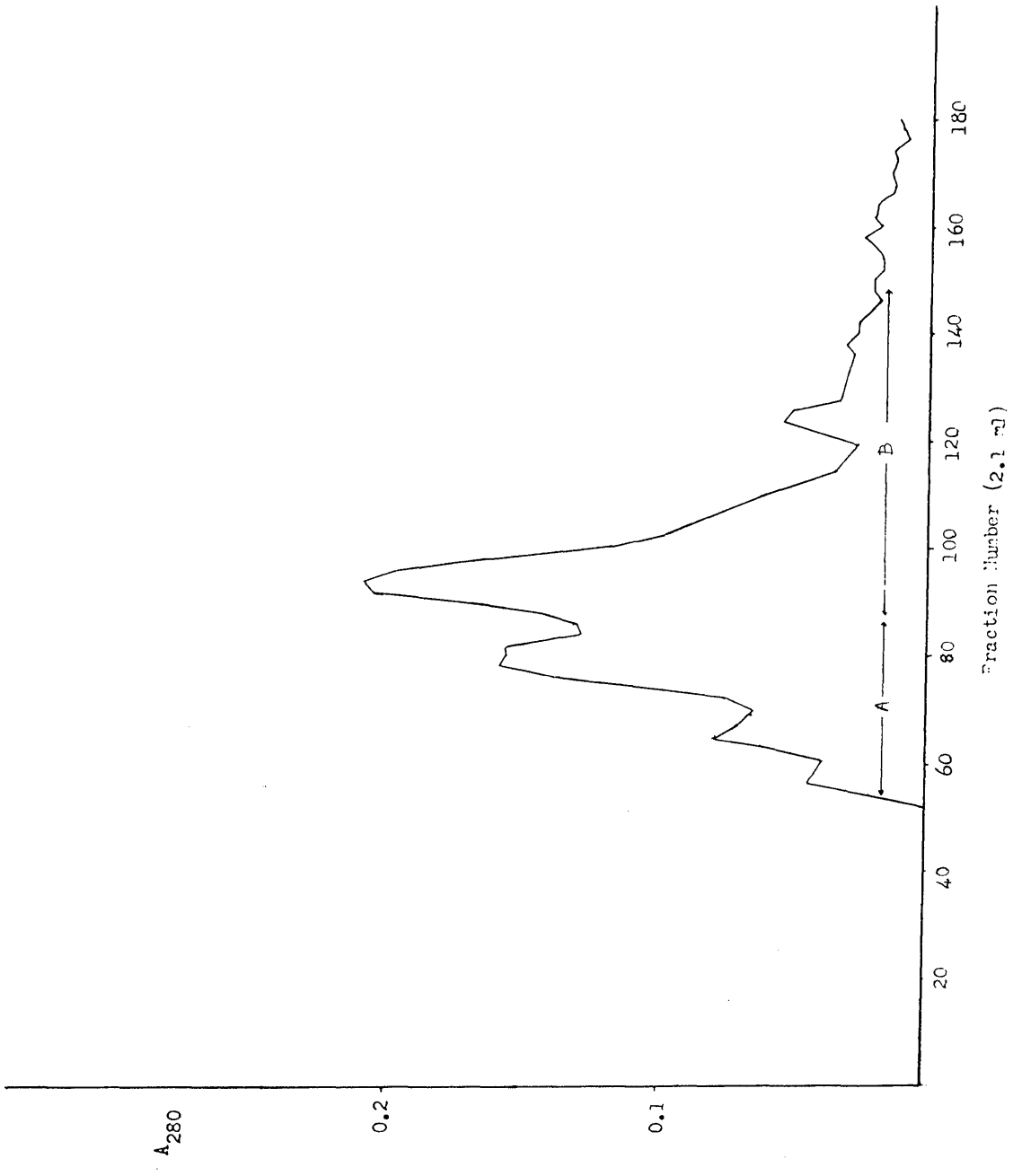
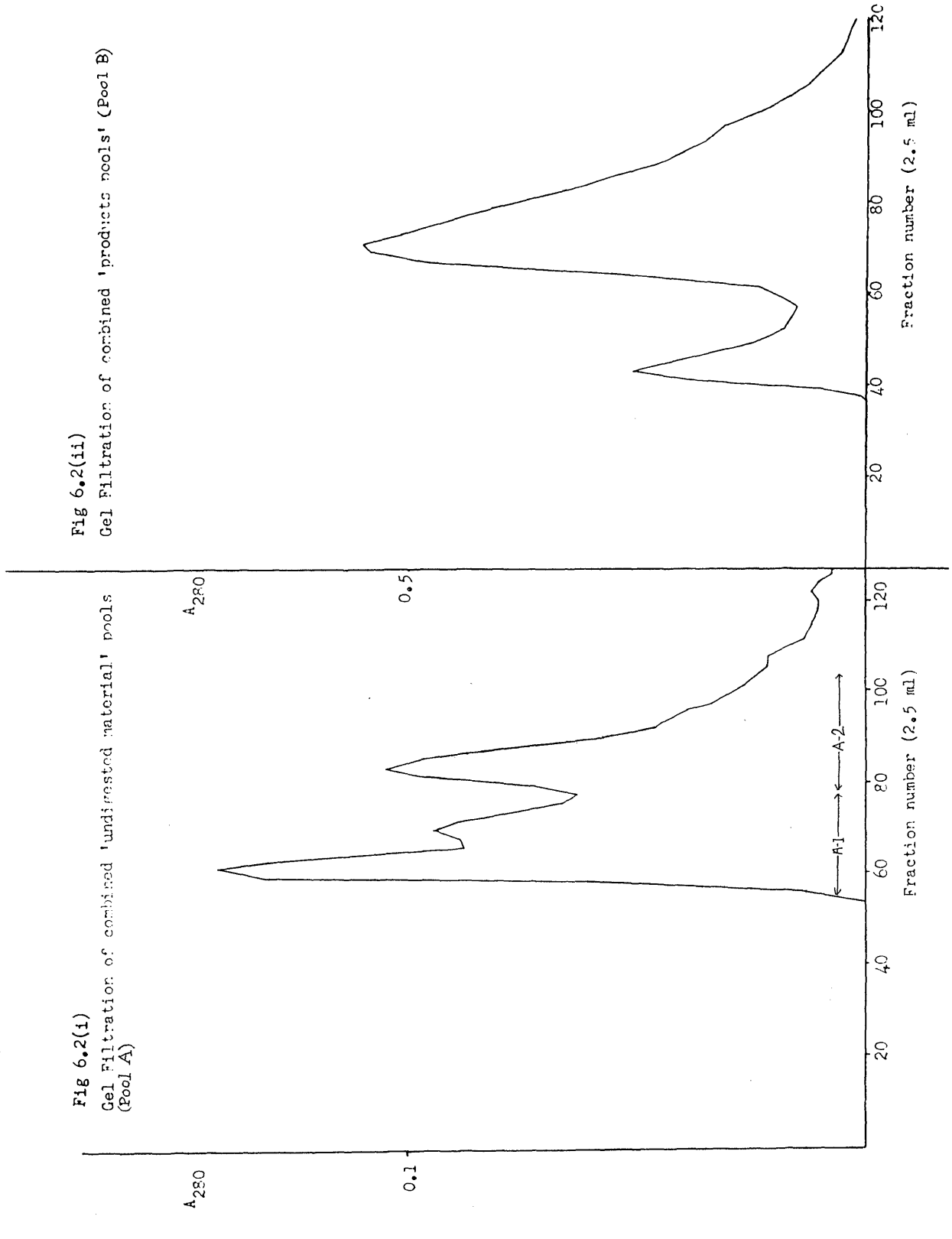


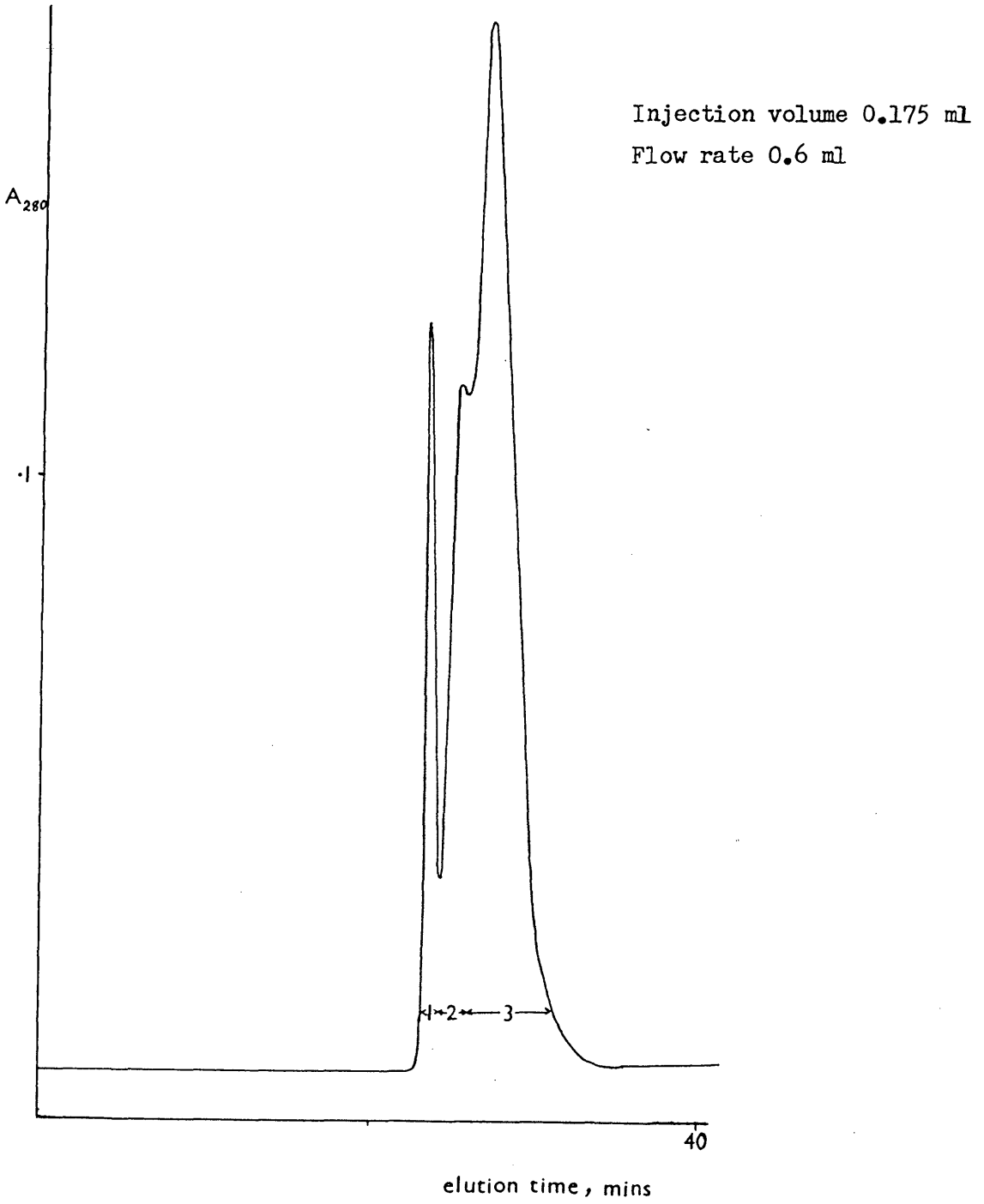
Fig 6.2: Gel filtration of Hydroxylamine digest



round of separation, on a basis of molecular size, by gel permeation chromatography on the Waters protein analysis column.

Upon gel permeation chromatography this pool was resolved into three components (fig. 6.3). The second and third of these were incompletely resolved. Pools were separated as indicated at the position of local absorption minimum, and were recycled through this process until each ran as a single homogeneous peak. These were termed pools 1, 2 and 3 in order of decreasing molecular weight. The contents of each were analysed by automated Edman degradation.

Fig 6.3 GEL PERMEATION CHROMATOGRAPHY OF LOW MOLECULAR WEIGHT MATERIAL



6.4 Peptide sequences obtained

Samples containing 50 nmoles of each of pools 1, 2 and 3 were taken for sequence determination. Such large quantities were used in an attempt to maximise the amount of sequence obtainable.

POOL 1

This pool proved to contain a single major peptide, from which 31 residues were determined:

Gly-Ser-Met-Ala-Asp-Val-Pro-Lys-Glu-Leu-
Met-Gln-Leu-Ile-Glu-Asn-Phe-Glu-Pro-Asp-
Ile-Thr-Val-Unk-Unk-Ile-Leu-Leu-Gln-Asp-
Val-

This peptide clearly overlaps the N-terminal sequence obtained from intact native hexokinase, and clarifies the sequence following N(13) of that sequence. It shows a complete hypothetical cyanogen bromide-cleavage-derived peptide running from positions 17-24 from the N-terminus and another from position 25 onwards.

The N-terminus of the X-ray sequence, known to be obtained from a truncated form of hexokinase, is proposed to fall within this peptide. The fit is not unequivocal. (See discussion for a complete description).

Since this pool contains only one peptide and no undigested material, it must correspond to all of the remaining 456 amino acid residues (using the figure of 472 residues from Schmidt and Colowick (1973a)) or to the core of the molecule from position 14 from the N-terminus to a position close to the C-terminus of the molecule.

POOL 2

This pool also contained one single major product, from which 44 residues were determined. It derived from cleavage at the site located within cyanogen bromide cleavage-derived peptide V-1, and both largely confirmed and completed the sequence of that peptide and overlapped with peptide IV-1A.

Pool 2 Gly-Ala-Tyr-Tyr-Asp-Val-Cys-Ser-Asp-Ile-
CNBr Gly-Ala-Tyr-Tyr-Asp-Val-Cys-Ser-Asp-Ile-

Pool 2 Glu-Lys-Unk-Gln-Gly-Lys-Leu-Ser-Asp-Asp-
CNBr Glu-Lys-Leu-Gln

Pool 2 Ile-Pro-Pro-Ser-Ala-Pro-Met-Ala-Ile-Asn-
CNBr Ala-Ile-Asn-

Pool 2 Cys-Glu-Tyr-Gly-Ser-Phe-Asp-Asn-Glu-Thr-
CNBr Cys-Glu-Tyr-Gly-Ser-Phe-Asp-Asn-Glu-Ser-

Pool 2 Val-Val-Leu-Thr
CNBr Val-Val-{Thr/Leu}-Pro

The sequence contains a dipeptide Gly-Lys at positions 15-16 which has no equivalent in the X-ray sequence.

When pool 3 was sequenced it proved to contain an equimolar mixture of two peptides which had the same N-terminal sequences as those of pools 1 and 2. These were termed 3A and 3B respectively.

6.5 Pattern of cleavage by hydroxylamine

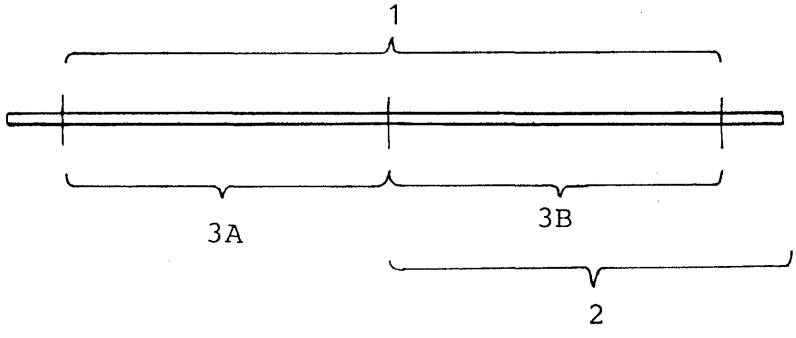
From the pattern of sequences obtained it is possible to determine the efficiency of cleavage at each site.

Pool 1 derives from either a single cleavage near the N-terminus or a cleavage here together with one near the C-terminus. Cleavage near the N-terminus proceeds with nearly 100% efficiency. This is an unprecedentedly high value for any hydroxylamine cleavage site. It may perhaps be artifactual due to selective precipitation at some stage in purification.

Pool 2 derives from cleavage at the internal site with an efficiency of about 60%, and failure to cleave at the C-terminal site.

Pool 3 derives from cleavage at all three sites. The efficiency of cleavage at each site must be about 50%.

Because Pool 2 is resolved from Pool 3B by sizing and yet contains one of the sequence present in that pool, it must have a greater chain length. This can only be so if a third and unidentified site for cleavage by hydroxylamine exists near the C-terminus of hexokinase:



The short N-terminal and prospective C-terminal peptides were nowhere observed and are presumably either lost during ultrafiltration or dialysis, or contain no aromatic residues and were not detected upon sizing.

CHAPTER 7

CLEAVAGE OF HEXOKINASE WITH ELASTASE

7.1 Introduction and rationale

In an attempt to provide sequence data rapidly, hexokinase (2μmole) was digested with porcine elastase to provide peptides suitable for sequence determination by electron-impact mass spectrometry.

Elastase shows some size specificity as well as its specificity for amino acid types, and produces few peptides that exceed 6 residues in length. The digest can be conveniently fractionated by a single chromatographic step to produce peptide mixtures of limited complexity suitable for mass spectrometric analysis. Such a single-step purification minimises losses, which helps to counteract the relative insensitivity of the electron-impact method of ionisation.

Peptide analysis by electron-impact mass spectrometry is not without disadvantages:

> The method is relatively insensitive, since the ionisation efficiency of the source is less than 1%. Samples must therefore be relatively large.
>
> The intensity of a spectrum falls off rapidly with increasing mass, so sequence ions become hard to distinguish from background.
>
> The requirement for peptides to be volatile limits their length, and for practical purposes the maximum length of a sequenceable peptide does not exceed 10 residues.

The use of elastase as protease minimises the third of these detriments. Signals hard to assign may be emphasised by derivatisation with deuterated reagents.

7.2 Materials and Methods

Proteolytic digestion with Elastase

100mg of hexokinase were dialysed against 50mM ammonium bicarbonate pH 8.5. The dialysed solution was made up to 20ml. Porcine elastase (the gift of Dr. D.M. Shotton), 1mg, in the same buffer, was added and digestion carried out at 37°C for 3 hours. Reaction was terminated by lyophilisation. Repeated lyophilisation removed most of the salt.

Ion-exchange chromatography

The lyophilised digest was taken up in 20ml of 0.025M pyridine/acetic acid buffer pH 2.1 and was applied to a water-jacketed column (30 x 1cm) of Locarte cation-exchange resin. Chromatography was carried out at 60°C with a nine-chamber gradient system of pyridine-acetic acid buffers, each including 20% n-propanol, with a pH gradient from 2.1 to 7.3 incorporating a pyridine molarity gradient from 0.025 to 8.0. Details of the buffer compositions are given overleaf. The flow rate of the column was maintained between 10 and 20 ml/hr by a Milton Roy chromatography minipump, operating at a pressure of 60 psi ($\sim 4 \times 10^5$ Pa). Each reservoir of the gradient held 60ml of buffer; 240 fractions each of 2ml were collected.

Aliquots ($\sim$1%) from each fraction were examined by high-voltage paper electrophoresis at pH 6.5 and 2.1, as described by Ambler (1963). Peptides were visualised by staining with fluorescamine (1% (v/v) in acetone containing 1% pyridine).

Buffer 1	pH 2.45 (0.025M Pyridine)		Buffer 6	pH 5.3 (3.75M)	
	Water	149 ml		Water	210 m
	n-Propanol	100 ml		n-Propanol	100 m
	Acetic Acid	250 ml		Acetic Acid	40 m
	Pyridine	1 ml		Pyridine	150 m
Buffer 2	pH 2.90 (0.075M)		Buffer 7	pH 5.9 (5.0M)	
	Water	197 ml		Water	190 m
	n-Propanol	100 ml		n-Propanol	100 m
	Acetic Acid	200 ml		Acetic Acid	10 m
	Pyridine	3 ml		Pyridine	200 m
Buffer 3	pH 3.3 (0.15M)		Buffer 8	pH 6.9 (6.5M)	
	Water	279 ml		Water	150 m
	n-Propanol	100 ml		n-Propanol	100 m
	Acetic Acid	115 ml		Acetic Acid	0.5 m
	Pyridine	6 ml		Pyridine	250 m
Buffer 4	pH 4.1 (0.75M)		Buffer 9	pH 7.3 (8.0M)	
	Water	270 ml		Water	80 m
	n-Propanol	100 ml		n-Propanol	100 m
	Acetic Acid	100 ml		Acetic Acid	0.5 m
	Pyridine	30 ml		Pyridine	320 m
Buffer 5	pH 4.8 (1.6M)				
	Water	279 ml			
	n-Propanol	100 ml			
	Acetic Acid	57 ml			
	Pyridine	64 ml			

I am grateful to Mr. A.T. Etienne for providing these buffers.

A number of peptide pools were made on the basis of these results.

Solvent was removed from peptide pools by rotary evaporation followed by dilution with water and freeze drying. Because the peptide pools were found to be contaminated with non-volatile salt it was necessary to desalt each by gel filtration on a column (90x1 cm) of Sephadex G-10 equilibrated with 10% (v/v) formic acid. Following gel filtration the peptides that eluted ahead of the salt were pooled and a sample taken for amino acid analysis. The presence of arginine or lysine in significant quantity indicated the conditions required for conversion to a volatile derivative:hydrazinolysis to convert arginine to ornithine residues; prolonged acetylation for peptides with a high lysine content.

Losses upon gel filtration were significant because many peptides eluted with the salt. Further pools were made that contained peptides in suitable quantity for mass spectrometry.

Peptide derivatisation

Hydrazinolysis

All the amino acids with the single exception of arginine are amenable to derivative formation by acetylation and permethylation alone. Arginine may be modified by these methods after its conversion into ornithine by a mild treatment with aqueous hydrazine (Morris et al, 1973). Peptide mixtures were incubated at 80°C with aqueous hydrazine for 12 minutes. Reaction was terminated by dilution with water and the peptide mixtures were freeze-dried repeatedly.

Acetylation

Peptide mixtures were dissolved in 1 drop of water and were incubated with 1ml of acetic anhydride/methanol:1/3 (v/v). The α-amino group of an N-terminal amino acid is completely acetylated within 5 minutes under these conditions, but the ε-amino group of lysine and the δ-amino group of ornithine require much longer. Samples shown by amino acid analysis to be rich in either or both of arginine and lysine were incubated for 3 hours at room temperature (Thomas et al, 1968).

Permethylation

The short permethylation technique of Morris et al (1973) was employed, in which the base formed by incubation of dimethylsulphoxide with sodium hydride is employed to abstract acidic protons, with methyl iodide as methyl group donor. Methylation was terminated by dilution with water after 65 seconds and the reaction products extracted with chloroform. The chloroform extract was washed three times with water before removing the solvent under a stream of nitrogen.

Peptide analysis

Suitably derivatised peptide mixtures were analysed on an AE1 MS50 mass spectrometer employing the electron-impact method of ionisation at a resolution of 2000. The ionisation current was 8mA and the accelerating voltage 8000. I am indebted to Dr. Anne Dell for performing this analysis.

Perdeuterolabelling

Because peptide sequence ions are separated on a basis of mass it is conceivable that sequence ions from different peptides in one mixture could share the same mass. This will introduce an element of ambiguity into sequence assignment if

the peptides remain unresolved by the temperature gradient method of mixture analysis. Such signals may be resolved by performing mass spectral analysis after derivatisation with either or both perdeuteroacetic anhydride or perdeuteromethyl iodide. Sequence ions are shifted in mass relative to their conventionally-labelled equivalents in multiples of 3 mass units, to an extent depending on the number of replaceable groups, and thus on their composition. Thus peptide sequence ions initially coincident may be resolved. This method is also useful in confirming single sequences from mixtures of multiple components.

The experimental procedure for perdeutero labelling is as for acetylation and permethylation but on a smaller scale, to conserve expensive reagents. I am indebted to Dr. Anne Dell for perdeuterolabelling.

7.3 Results

A large number of peptide pools were analysed in this manner. Many short (typically n = 3 or 4) peptides were sequenced. The sequences obtained from each pool are listed in fig. 7.1. Sequences or residues shown in parentheses were not assignable to positions or equivalent to residues in the chemically-determined sequences. The figures in parenthesis following assignable sequences indicate their positions relative to the X-ray sequence.

The mass spectrometer does not normally resolve leucine and isoleucine. Both these residues appear as leucine in the sequences shown in fig. 7.1. The ambiguity is resolved upon their comparison with the sequences determined by automated Edman degradation (fig. 7.2).

Although the majority of the peptides sequenced by mass-spectrometry could be aligned with regions of the chemically-determined sequences, this was not possible for all because of the incomplete nature of the latter. The peptides that could be fitted thus are shown relative to the corresponding chemically determined sequences in fig. 7.2.

By production of sequence information from a second digest and peptide fractionation, some small ambiguities and allocations in the chemically-derived sequences were resolved; thus residue 127 is shown to be leucine rather than glycine, and residue 256, unidentified in the chemically-derived sequences, to be leucine also. It is not possible to employ a tripeptide as the basis of resolution of sequence ambiguities, since the assignment of position based on matching two residues alone is insufficiently accurate.

None of the sequences derived from cleavage by cyanogen bromide is truly extensively confirmed but the majority possess

Fig 7.1

Sequences obtained from massspectrometric sequencing of Hexokinase

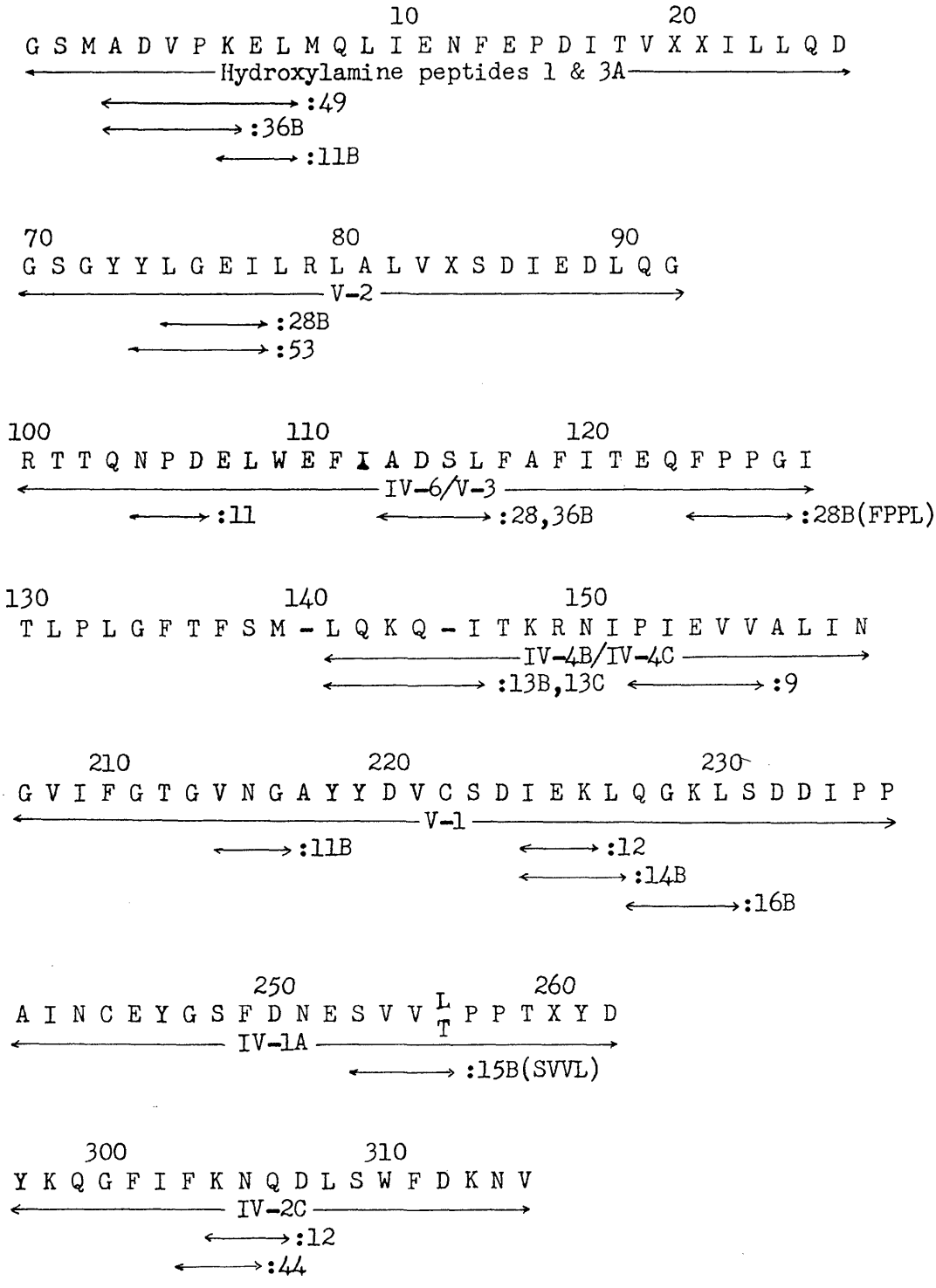
Peptide pool:	Sequences obtained
9	Pro Leu Glu Val Val(152-6)
11	Asn Pro Asp(104-7)
11B	Lys Glu Leu(3-5) Val Asn Gly(214-6)
	(Asn Leu Pro Leu) (Glu Lys Ala Ala) (Lys Tyr Arg Leu)
12	Leu Glu Lys(225-7) Lys Asn Glu(304-6)
13B	Leu Gln Lys Gln Leu(141-6) (Leu Val Lys)
	(Lys Asn Lys Glu Leu)
13C(CD_3I)	Leu Gln Lys Gln Leu(141-6)
14B	Leu Glu Lys Leu(225-8) (Phe Glu Asn)
15B	Ser Val Val Leu(253-6)
16B	Gln Gly Lys Leu(229-30)
21C(CD_3I)	(Tyr Tyr Thr)
25B,C(CD_3I)	(Lys Tyr Arg Leu)
28	Ala Asp Ser Leu(113-6) (Leu Val Ser Leu)
28B	Leu Glu Gly Leu(75-8) Phe Pro Pro Leu (124-7)
31B	Lys Glu Leu(3-5)
36B	Asp Val Pro Lys(-1-4) Ala Asp Ser Leu(113-6)
41	(Leu Asn Lys)
42	(Lys / (Gly Leu) } Leu Gly)
44	Phe Lys Asn(303-5)
49	Asp Val Pro Lys Glu Leu(-1-5)
53	Tyr Leu Gly Glu Leu(74-8) (Phe Thr Ala)

Leucine and Isoleucine are unresolved by the mass spectrometer. Arginine appears as Ornithine (2,5-diaminopentanoic acid) but is here described correctly.

The sequences shown in parenthesis were unassignable to positions in the chemically-determined sequences.
The numbers in parenthesis following some of the sequences indicate the positions of these peptides relative to the primary structure of hexokinase. The numbers are the positions of these residues in the X-ray sequence. Peptide 49 falls outside this sequence, in the N-terminal region of the hexokinase molecule.

Fig 7.2

Alignment with chemically-derived sequences of peptides sequenced by mass spectrometry



The numbering is that of the X-ray sequence of Anderson et al (1978b)

a region or regions sequenced by mass spectrometry. Both peptides IV-4C and V-2, which share few residues with the X-ray sequence (see discussion) are corroborated. This result suggests that they are truly derived from hexokinase and are neither contaminants nor artifacts.

To sequence a medium-sized protein by electron-impact mass spectrometry alone would be a forbiddingly difficult task, since with the short peptides required by the method it is difficult to achieve sufficient overlap between peptides. Even with longer peptides the task would be only slightly easier, since ionisation yields would fall. However, as a means of obtaining extensive sequence information rapidly to corroborate or confirm unproven sequences, to provide sequences at blocked N-termini and to remove ambiguities, it is a technique of remarkable versatility and utility.

With the development of sample ionisation by bombardment with fast neutral atoms, this shortcoming in mass spectrometric technique of poor efficiency of ionisation has been largely resolved. Allied with the use of magnetic field strengths suitable for analysis of large (mwt > 10^3) molecules the range and compass of mass spectrometry has been greatly extended.

CHAPTER 8

DISCUSSION

8.1 Pattern of cleavage by cyanogen bromide

The amino acid analysis of hexokinase B by Schmidt and Colowick (1973a and fig. 1.1) indicated the presence of 11 residues of methionine per molecule of monomeric hexokinase, a figure found also, in this department, by Barnard (unpubl.). No more than twelve peptides should therefore be produced by digestion of hexokinase with cyanogen bromide. However, N-terminal analysis of a whole cyanogen bromide digest indicated the presence not of 12 but of 13 peptides, and identified the relative amounts of their N-termini (fig. 8.1).

Of these 13 peptides, 12 were characterised by either total or partial sequencing. In practice the digestion was not complete and these peptides were not present in equimolar quantities. Some were present only as minor components of the peptide pools produced by the application of HPLC, and some were identified only by their positions within the peptides derived from cleavage by hydroxylamine.

The number and nature of the N-termini of these characterised cyanogen bromide cleavage-derived peptides corresponds well to that of the whole, unfractionated digest (fig. 8.1).

One of the two peptides proposed to have either glutamic acid or glutamine as its N-terminal residue was not observed at all. Investigation of the pattern of cleavage by cyanogen bromide in the C-terminal region of hexokinase suggests that residue 351, proposed by the X-ray sequence to be methionine, may be correctly identified (vide infra) even though no peptide corresponding to cleavage by cyanogen bromide at that point was

Fig 8.1

N-terminal residue found	Number of peptides predicted	Number of peptides observed	Identity of peptide
D	2	2	III-2 & IV-4A
E(Q)	2	1	H_2NOH peptide 1
G	2	2	V-1 & V-2
A	2	2	IV-1A & IV-6 (minor)
Y	1	1	IV-2C
R	1	1	IV-6 (major)
V	1	1	Protein N-terminus
L	1	1	IV-4C (major)
I	1	1	IV-4C (minor)

Pattern of cleavage of Yeast Hexokinase B: the peptides predicted to occur in a cyanogen bromide digestion and those eventually characterised.

observed.

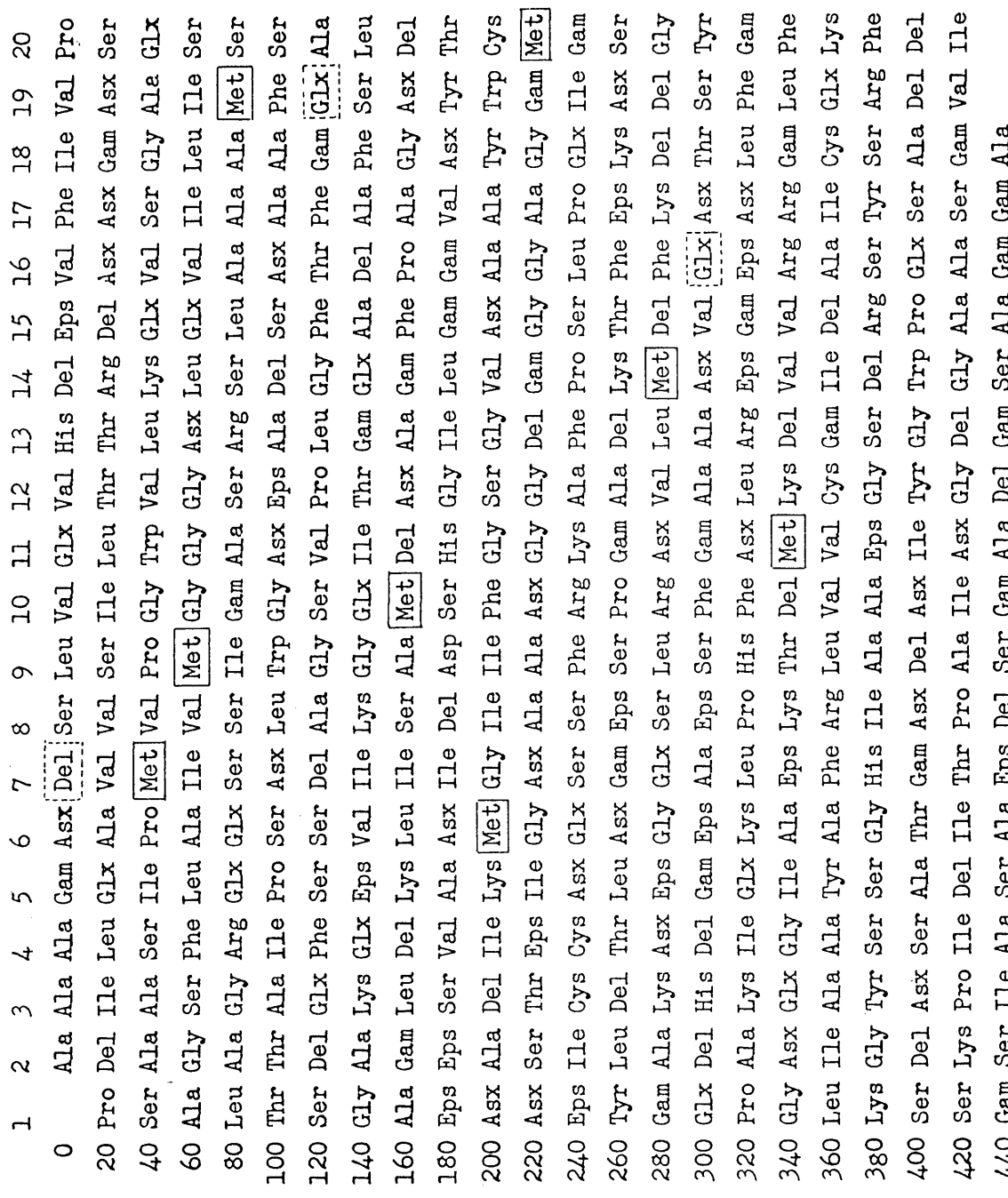
The X-ray sequence of Anderson et al (1978b) identifies only 8 of the 11 methionine residues believed to be present per hexokinase subunit (Schmidt and Colowick, 1973a). Peptides corresponding to cleavages at all these proposed methionine residues except that at 351 were observed and characterised. Two of the remaining methionines are in the N-terminal peptide that is removed from the apoenzyme by proteolysis, and thus do not appear in the X-ray sequence.

The remaining methionines were misidentified by the X-ray sequence and were located by matching sequenced cyanogen bromide cleavage-derived peptides to positions in the X-ray sequence. The locations of all these methionine residues in the X-ray sequence are shown in fig. 8.2.

The existence of the methionine proposed by Anderson et al (1978b) to occupy position 351 of their sequence, but not identified from any peptide, is the probable explanation for the pattern of cleavage of peptides IV-2C and IV-2D. These peptides gave only 19 identifiable residues which derived from a cleavage at methionine-295. They are unlikely to terminate at the well-characterised methionine -317 but probably extend beyond this point to another, uncharacterised, methionine, since not only does a total length of 23 residues seem far too short to merit inclusion in pool IV but also their amino acid analyses show them to contain methionine, which could derive from an internal residue but not a terminal one. Amino acid analysis of peptides II-2 and II-3, which derive from cleavage after methionine-317 and are proposed to contain all of the C-terminal 148 residues also show the presence of methionine, which must be due to an internal residue.

No peptide of length corresponding to cleavage at

Fig 8.2: Pattern of cleavage by Cyanogen Bromide of Yeast Hexokinase

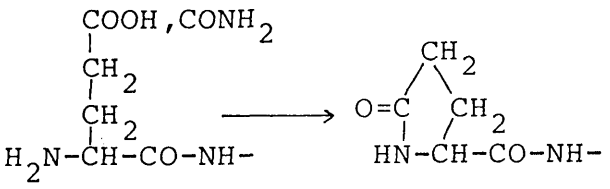


Methionine residues identified by the X-ray sequence are shown in boxes. Residue 351 was not identified chemically but all the others are believed to be correct Residues misidentified by the X-ray sequence and subsequently shown to be methionine are hatched in.

The sequence is the X-ray sequence from Anderson et al.(1978b)

methionines -317 and -351 was found. All the peptides which had Asp-318 as N-terminal residue contained the entire C-terminal region of the hexokinase molecule.

Pools III-1 and III-2 proved resistant to sequencing. Such resistance may be due to possession of N-terminal glutamate or glutamine, in the form of the internally-cyclised pyrollidone:



These peptides could correspond to that derived from cleavage at the partially characterised Met (-7) extending as far as Met-99, or at the uncharacterised Met-351. Without the capacity to remove such a blocked N-terminus these peptides remain,of necessity,uncharacterised.

With the exception of methionines 316 and 351, cleavage by cyanogen bromide seems to have been fairly complete. No N-terminal sequence corresponding to one of the shortest ($n < 30$) peptides expected from such a digestion was observed in high yield, so cleavage at the corresponding methionine residues seems to have been complete. Such short peptides were presumably all located in Pool VI which could not be investigated due to the practical difficulty presented by its heavy contamination with salt.

8.2 Alignment of chemically-sequenced peptides with the X-ray sequence

Because the majority of sequence data came from peptides produced by cleavage with cyanogen bromide, and cleavage by hydroxylamine provided only one overlap between such independently sequenced peptides, assignment of the non-overlapped peptides to their correct positions in the primary structure of hexokinase must be made by reference to the X-ray sequence alone.

The initial basis of alignment of chemically sequenced peptides with the X-ray sequence was on sequence identity alone. Peptides were aligned with that region of the X-ray sequence that provided the greatest number of correctly-matched residues. On this basis, peptides III-3B, IV-4A, IV-2C/2D, IV-4A/4B, the minor component of IV-4C, the major component of IV-6 and V-1 were readily aligned with corresponding regions of the X-ray sequence (fig. 8.3). The overlap between IV-1A and V-1 provided by peptides 2 and 3B derived from the cleavage by hydroxylamine confirmed their positions and completed the sequence of V-1.

Also incorporated at this point and on this basis was a 17-residue peptide isolated by Kuromizu et al (1979) from a tryptic digest of hexokinase A that had been inactivated by autophosphorylation by ATP at a serine residue. This peptide was aligned with residues 129-145 (fig. 8.5). Although nine residues of this peptide aligned very well with the X-ray sequence, the N- and C-terminal regions were less accurately matched; nor was either lysine or arginine found at position 128.

It was seen from these peptides that their presumptive N-terminal methionines had, with the exception of III-3B, been identified correctly, and that their aromatic residues had been identified as such, if not actually correctly by the X-ray

Fig 8.3(i) Chemically-derived sequences readily assignable to positions in the X-ray sequence

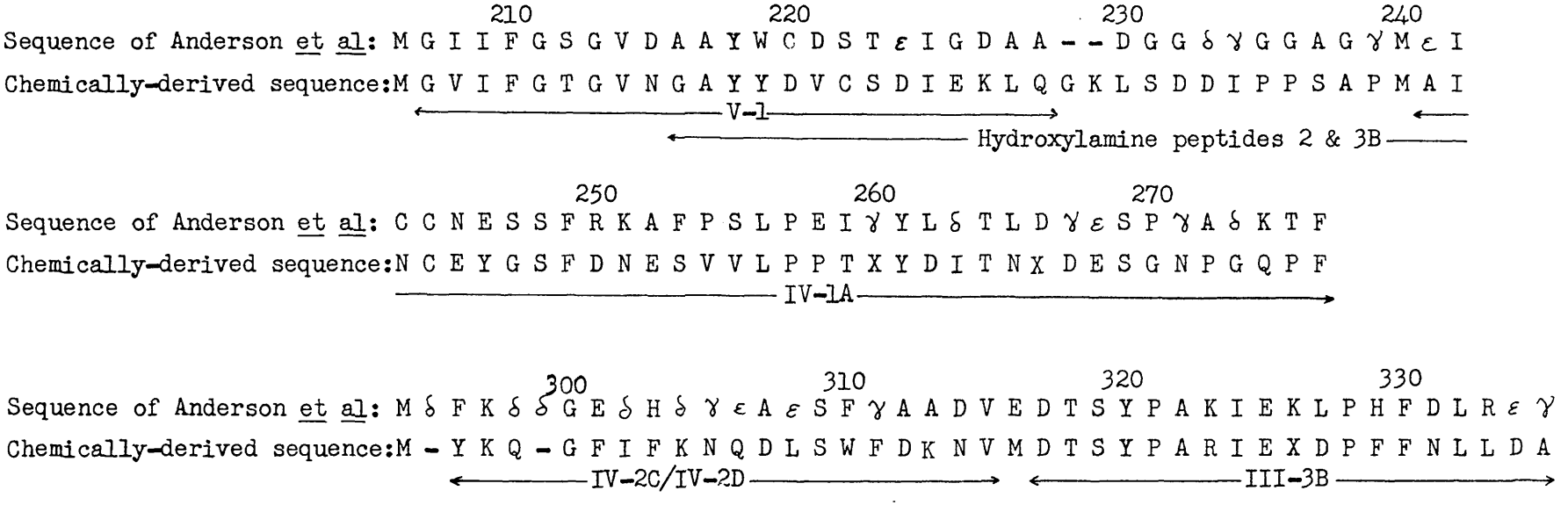
```
                                    50                  60
Sequence of Anderson et al: M V P G W V L K E V S G A E A G S F L
Chemically-derived sequence:M I D G Y V E K E I N D L E D V N F G
                            ←————————IV-4C————————→
                                (minor component)

                          100                 110                 120
Sequence of Anderson et al: M S T T A I P S D L W G D ε A δ S D A A F S S δ E F S S δ A
Chemically-derived sequence:M R T T Q N P D E L W E F I A D S L F A F I T E Q F P P G I
                            ←————————————————IV-6/V-3————————————————→

                                130                 140
Sequence of Anderson et al: G S V P L G F T F γ E A G A K E ε
Sequence of Kuromizu et al: D T L P L G F T F S Y P A S N Q K

                        170                 180                 190
Sequence of Anderson et al: M δ D A γ F P A G D δ ε ε S V A D I N D S H G
Chemically-derived sequence:M - - - D F P T G E E S X S F L X I D D G G T
                                ←————————IV-4A'/IV-4B/IV-4B'————————→
```

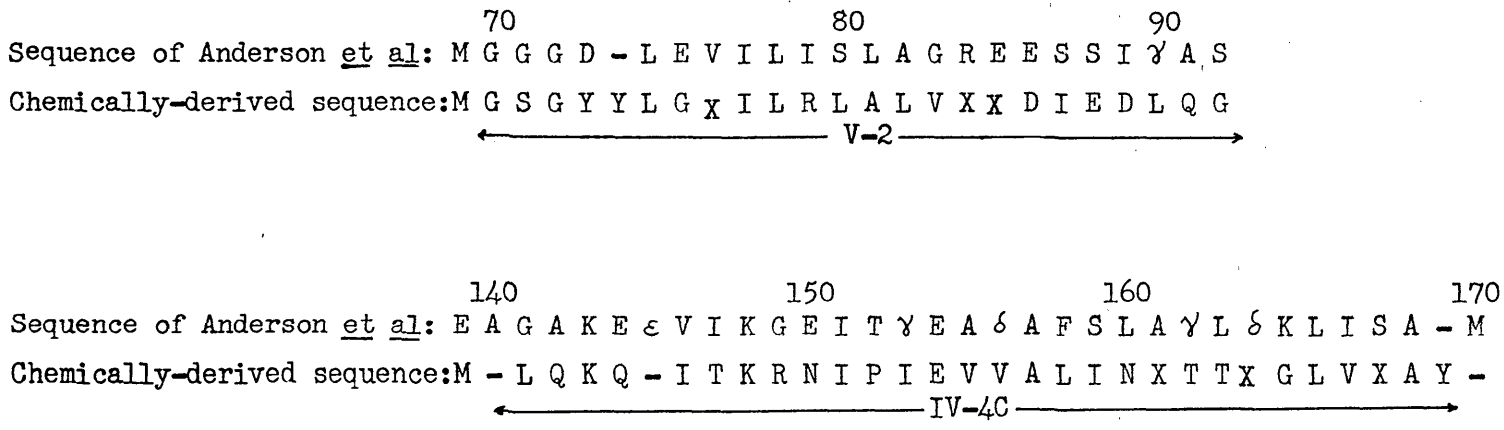
Fig 8.3(ii)


210 220 230 240
Sequence of Anderson et al: M G I I F G S G V D A A Y W C D S T ε I G D A A - - D G G δ γ G G A G γ M ε I
Chemically-derived sequence: M G V I F G T G V N G A Y Y D V C S D I E K L Q G K L S D D I P P S A P M A I
V-1
Hydroxylamine peptides 2 & 3B
250 260 270
Sequence of Anderson et al: C C N E S S F R K A F P S L P E I γ Y L δ T L D γ ε S P γ A δ K T F
Chemically-derived sequence: N C E Y G S F D N E S V V L P P T X Y D I T N X D E S G N P G Q P F
IV-1A
300 310 320 330
Sequence of Anderson et al: M δ F K δ δ G E δ H δ γ ε A ε S F γ A A D V E D T S Y P A K I E K L P H F D L R ε γ
Chemically-derived sequence: M - Y K Q - G F I F K N Q D L S W F D K N V M D T S Y P A R I E X D P F F N L L D A
IV-2C/IV-2D
III-3B

sequence. With the exception of III-3B, where the figure is remarkably higher, the X-ray and chemically derived amino acid sequences show about 30% of identical residues.

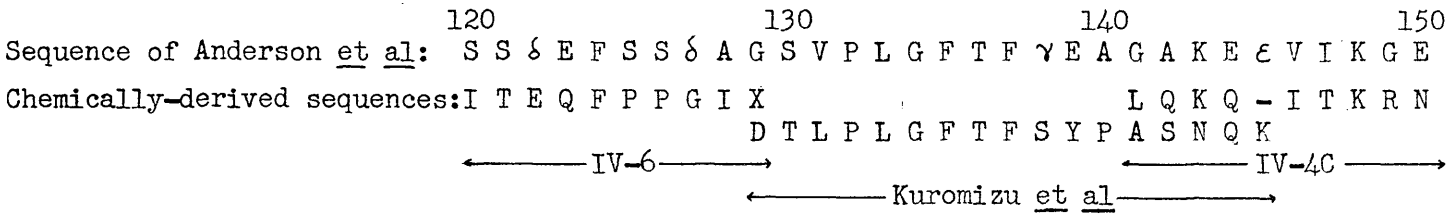
Without the benefit of known overlaps, the other peptides could not readily be assigned a position on a basis of sequence identity alone. Because the X-ray sequence seemed to be accurate in its assignments of methionines and aromatic residues, an attempt was made to align the peptides by matching one or more of these features. By this means, peptide IV-4C was aligned with residues 141-169 and peptide V-2 with residues 70-99 (fig. 8. In both cases the extent of sequence identity was poor and the peptides bore less obvious steric resemblance to the X-ray sequence than those that were readily assigned. It is especially surprising that the Tyr-Tyr dimer of peptide V-2 has no equivalent in the X-ray sequence - neither correctly identified nor as a dimer of aromatic or cyclic (histidine) residues. Konigsberg has claimed in Anderson et al (1978b) that residues 70-72 are in fact Gly-Gly-Ala, and in Steitz et al (1981) that they are Gly-Gly-Thr. This result was obtained from a tryptic peptide, the full sequence of which remains unpublished. Similarly the allocation of peptide IV-4C to positions 140-69 might seem at first to be circumscribed by the overlap with the position of the sequence of Kuromizu et al (1979) (fig. 8.5), but this sequence derives from hexokinase A and it is known (ibid) that the corresponding peptide from hexokinase B differs both in length, being of 24 residues, and composition. If the assignment of peptide IV-4C to positions 141-169 is correct, then the equivalent in hexokinase B of the sequence of Kuromizu et al (1979) might derive from cleavage after residues Unk-118 and lysine-143. That peptide IV-4C is not a contaminant is shown by the fact that its N-terminal pentapeptide has been identified

Fig 8.4



Peptide sequences not readily assignable to positions in the X-ray sequence due to inadequate sequence homology with it

Fig 8.5



Chemically-derived sequences proposed to occupy positions 120-150 of the X-ray sequence

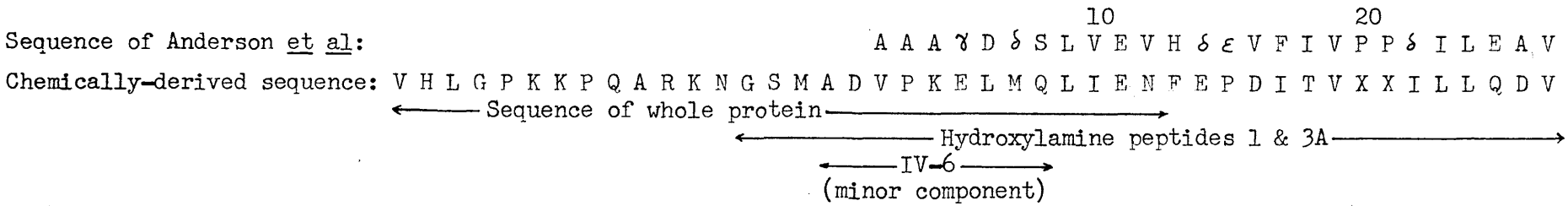
both by this writer and Dr. Maria Blair in two separate elastase digests of hexokinase that were analysed by mass-spectrometric sequencing, and by the appearance of this peptide (in low yield) in a cyanogen bromide peptide pool analysed by Dr. Nicola Barron. Each of these analyses made use of hexokinase from a different purification. It is unfortunate that sequencing by mass spectrometry provides no overlap of this peptide with any other.

The sequence obtained from the intact whole molecule and that obtained from peptides 1 and 3A obtained from the cleavage by hydroxylamine contain a total of 44 residues. The sequence includes two of the hypothetical peptides expected from the cyanogen bromide digest and the N-terminal sequence of another.

The crystalline hexokinase B from which the X-ray sequence was derived is known to be a lytic product that has lost its N-terminal 11 residues. The true N-terminus must therefore lie at least this number of residues away from the N-terminus of the X-ray sequence. The polypeptide chain may be mobile in the crystal and therefore provide only a poorly-resolved structure in this region. Certainly no major structural feature is accentuated by the X-ray sequence here. The alignment proposed for the chemically determined N-terminal sequence of the protein (fig. 8.6) is comparatively poor in sequence identity. Although the two sequences show some slight structural resemblance it is not particularly well defined. The true N-terminal is set 18 residues downstream of Ala-2 which seems to be the N-terminus of the X-ray sequence. Eleven residues of the 18 are due to the missing N-terminal peptide. The other 7 are proposed to be disordered and mobile and thus not identifiable

Fig 8.6

N-Terminal sequence of intact Yeast Hexokinase B



in the crystal structure.

Bennett and Steitz (1980a) have stated that there are about 15 amino acid residues missing from the N-termini of the crystalline forms prepared from both cleaved and intact hexokinase, so the amino-terminal peptide may in fact be disordered in the high-salt environment necessary for crystal growth.

It was clear from attempts to find the best homology between chemically-sequenced peptides and the proposed X-ray sequence that the two sequences do not run in continuous and equivalent parallel. The X-ray sequence contains both deletions and insertions relative to the chemically-derived sequences obtained. It clearly omits the Gly-Lys dimer found between residues 229 and 230, and hypothesises residues at positions 171-3, 295 and 299 that have no equivalent in the appropriate chemically-derived sequence. Positions such as those where one sequence is not continuous relative to the other are indicated in figures 8.3 and 8.4 by dashes. It must be emphasised that these are not positions at which a residue was present but identifiable; such positions are designated by the symbol X or Unk.

Positions 229-30, 295 and 299 all fall within regions where the polypeptide chain both has a random coil configuration and lies on the surface of the molecule. These regions of the chain may be mobile and disordered in the crystalline enzyme. Residues 171-3 are proposed to form one strand (αG) of the small, 2-stranded β-sheet in the small lobe and ought to be restricted in their movement, and be well defined in the crystalline structure. Residue 174 is the N-terminus

of the cyanogen bromide cleavage-derived peptide IV-4A' which shows good homology with the X-ray sequence at this point. It is curious that the methionine residue which must occupy position 173 is proposed to occupy position 170.

Conceivably residues 170-2 - which could not,unfortunately, be identified - are not in fact missing but form the C-terminus of the peptide preceding IV-4A'. Although the positions of regions of secondary structure are approximate and thus the β-strands may not occupy precisely the positions indicated, it would be highly surprising to find no amino acid sequence at all for this feature.

8.3 The accuracy and validity of the X-ray sequence model

Anderson et al (1978b) claim that their X-ray sequence is correct in approximately 60% of its assignments. This claim is based on the alignment, by computer, on a basis of matching the maximum number of residues, of two nonapeptides. One is part of a peptide from Hexokinase A that was sequenced by Kuromizu et al (1979 & fig. 8.5), the other a peptide from hexokinase B that incorporated an S-carboxymethyl cysteine residue, sequenced by Blair and Barnard (unpubl.). Anderson et al (ibid) have claimed that on the basis of these results it should be possible for peptides as short as 10 residues in length to be fitted to the X-ray sequence. Thus total sequencing of all the peptides from a single digest (assuming that their lengths were not less than this figure) would produce the complete primary structure, without any need to establish or define peptide overlaps.

The figure of 60% accuracy claimed for the X-ray sequence is based on a very small sample - less than 5% of the total sequence. It is clear from the results described herein,

obtained from a much larger sample, that homology as extensive as that claimed is the exception rather than the rule, and is found only in short stretches at a few sites. A figure of 60% homology is thus not representative and a more correct figure for the net sequence homology would not exceed 30%.

It is thus highly unlikely that the complete primary structure of hexokinase could be determined by sequencing and aligning short ($n < 10$) peptides, since there would almost always remain some uncertainty as to their allotted positions. This would particularly be the case in regions of the X-ray sequence that are ill defined and possess a high proportion of unassigned residues. Inspection of the X-ray sequence reveals that the C-terminal 50 residues comprise such a region, with few large residues (Tyr, Phe, Met, Lys) that might be expected to be identified correctly by the X-ray sequence and to provide the basis for aligning a sequence by matching its gross features alone. Even when the chemically-sequenced peptides are much longer than 10 residues, such as in peptides derived from cleavage by cyanogen bromide, the correct and unequivocal positions of such peptides have not always been apparent.

8.3.1 Alignment of peptides on a basis of steric homology

A more credible claim made for the accuracy of the X-ray sequence by its authors is to have minimised the discrepancy between correctly and incorrectly assigned residues to as few as two carbon atoms of the side-chain. If this is so, then the relative sizes of the amino acid sidechains along one sequence should be parallelled by those of the other.

Two sequences can be said to possess 'steric homology'

when the residues of similar size occupy corresponding positions. Features in one sequence, such as large (and thus easily-identifiable) sidechains, or runs of the same residue correspond to similar features in the other. Such steric homology could be used to align peptide sequences which possessed few correctly-identified amino acids.

If we assume as a (fairly restrictive) definition of structural homology a difference between residues of one C, N or O atom, and likewise all residues having aromatic or cyclic sidechains, then the following mismatches between the chemically-derived and the X-ray sequences are observed, which can be considered to show acceptable steric homology:

V:I	T:γ	D(N):δ	H:F
L:I	S:γ	E(Q):ε	W:Y
I:T	A:γ	E(Q):D(N)	Y:F
V:L	S:T		
V:γ	S:A		
L:δ	A:G		
I:δ			

γ, δ and ε represent unbranched aliphatic sidechains of 2, 3 and 4 carbon atoms respectively.

When this scheme is employed the net homology (identity and steric) between the X-ray sequence and the chemically-derived sequences is 43%.

8.3.2 Optimisation of electron density data

A more realistic means of making the best use of electron density data available from high-resolution X-ray crystallography is to fit sequenced peptides directly, without attempting to identify residues from the electron density data. In effect, a model of the electron density of each chemically sequenced peptide is constructed, and the model is matched by computer with the known electron density along the polypeptide

backbone. This was done for yeast phosphoglycerate kinase by Banks et al (1979) who were able to sequence completely the cyanogen bromide peptides of that enzyme. These authors aligned the peptides by a series of criteria of steadily increasing discrimination: matching first of plausible methionines, then of residues of large and well-defined side chain structure (such as aromatic residues): by attempting to equate cysteine residues with the known binding sites of mercurials, and equating short hydrophobic stretches with regions of β-sheet. Peptides fulfilling none of these criteria were simply fitted into the gaps in the sequence until the shape of the sequenced peptide matched the electron density. This approach is only possible where sequence and electron density data are available together, which was unfortunately not the case in the present work.

Fitting non-overlapping chemically-sequenced peptides into an X-ray sequence is likely to produce a poorer fit than if they were fitted directly to electron density data. An X-ray sequence is unlikely to be correct in all its positions unless the resolution of the original X-ray crystallography has been extraordinarily high (<1.8A). It will thus be intrinsically inaccurate in its sequence allocations. The chemically-derived sequences, which match the electron density data more closely than they do the X-ray sequence, cannot be aligned with the X-ray sequence with an accuracy greater than the accuracy of the X-ray sequence. This is because a 'correct match' of a residue in a chemically-derived sequence to one in the X-ray sequence indicates also a correct match of the same residue in the X-ray sequence with the electron density data. No more residues of a chemically-derived peptide sequence can be matched with residues in the X-ray

sequence than the X-ray sequence identifies correctly.

Thus only those correctly identified residues provide a sensible basis for aligning non-overlapping chemically-sequenced peptides into a complete primary structure. The accuracy (or lack of it) in their assignment will reduce the potential of the X-ray sequence in turn to act as a basis of alignment. The error involved in formation of the X-ray sequence will be reflected in the error involved in aligning peptides. The cumulative effect of such errors may be such as to reduce the resemblance of (particularly short) chemically-derived sequences to the implicit electron density data. The structural information hidden in the electron density data is not used as fully as it could be, and the grosser structural features it possesses are not as manifest in the double alignment - of X-ray sequence to electron density data and of chemically-derived sequence to X-ray sequence - as they would be in a single alignment. Because peptide alignment in the absence of a highly accurate X-ray sequence is made on a basis of matching gross structural features, that are comparatively easy to identify from electron density data, it is essential that these be emphasised as fully as possible.

By the same principles, the steric resemblance of chemically-sequenced peptides to electron density data will be reduced through the degradation of the electron density data. In searching for steric homology, the features sought are those in which there is a large difference, in length or shape, between the sidechains of adjacent amino acids. If this difference is less obvious in the X-ray sequence than in the electron density data, then a structural prominence is not usefully available.

If, as seems to be the case in the present work, the X-ray sequence is correct on average in only 30% of its allocations, and perhaps 50% in stretches, then short peptides cannot be aligned on a basis of matching residues. It becomes essential that the pattern of the shape of the molecules, read along the polypeptide backbone, be as correctly interpreted as possible and that the shape of the molecule of the X-ray sequence should at least mimic accurately this actual shape. In view of the fact that the sum of the extent of acceptable steric homology with the extent of sequence identity is 43% it is possible to be reasonably confident that the information available in the electron density data is well enough expressed in the X-ray sequence for this latter to present a reasonable means of aligning and overlapping peptide sequences.

8.4 The nature of the substrate-binding site

The refined model of the sugar-binding site (Anderson et al, 1978c) derived from electron density mapping of the complex of glucose with the inhibitor o-toluoyl glucosamine shows that glucose coordinates to hexokinase via its 1,3,4 and 6-hydroxyl groups. The natures of the coordinating groups were surmised from the electron density data. Thesedata do not distinguish between free acid and amide, and since it is unlikely that all the coordinating groups in the active site are free carboxylate groups, residues 188, 215 and 245 were assumed to be amidated. Residue 189 is necessarily the free acid, so it can act as a general base catalyst.

Similarly, Shoham and Steitz (1980) derived from the electron density data of the complex of hexokinase with 8-Br-AMP, and by model building of ATP into the active site, a model of the nucleotide binding site and the binding of the β- and γ-phosphates of ATP.

The residues implicated in coordination have been identified in the chemically-derived sequences and are shown on the same diagram, fig. 8.7. Although in some cases the identity of the residue is not that proposed, in all cases it is consistent with its postulated role.

8.5 The thiol groups of hexokinase

Otieno et al. (1975, 1977) have discovered from affinity labelling of hexokinase that one of its four thiol groups is

Fig 8.7

The residues predicted to make substrate-binding interactions and their chemical identification

Substrate group interacting	Position of contact	Predicted residue	Corresponding identified residue
Glucose: 1-OH	277	Gln	Not identified
3-OH & 4-OH	188	Asn	Asp-188
3-OH	245	Asn	Glu-245
4-OH	215	Asn	Asn-215
4-OH & 6-OH	189	Asp	Asp-189
Adenine ring	397	Tyr	Not identified
Ribose: 2'-OH	317	Asx	Asp-317
3'-OH	319	Ser	Ser-319
β and γ phosphates	212 *	Ser	Thr-212
of ATP	393	Ser	Not identified

The predictions are by Anderson et al (1978c) and Shoham & Steitz (1980) on the basis of electron-density maps and a proposed interaction scheme.
* Deduced to be binding to the backbone nitrogen, not to the side-chain

essential for activity, since blockage with even a small neutral group such as -CN causes irreversible inhibition, whereas derivatisation of the other three thiols does not affect the activity of the enzyme. This thiol is presumed to be buried in the hydrophobic interior of the protein since it is unreactive at temperatures below 31°C.

The X-ray sequence (Anderson et al, 1978b) proposed two cysteine residues to be situated close to the active site, as residues 243 and 244. Cys-243 was proposed by these authors to be the essential thiol, since it was situated close (5Å) to the glucose-binding site and made extensive hydrogen-bonded interactions with residues in that site, to Asx-188, Asx-215 and Asx-245. In fact, the sequence around this region reveals only one thiol, at position 244, whilst position 243 is actually asparagine. This latter assignment agrees better with the hydrogen-bonding scheme. It is unclear as to what interactions with the rest of the molecule position 244 makes. It is situated 10Å from the active site and must be buried in the protein. Presumably inactivation proceeds by disruption of the active site by conformational change.

8.6 Discussion of the conformational change

Upon binding glucose, hexokinase undergoes a large conformational change in which the two domains of the monomer move towards each other. The binding of glucose becomes significantly tighter. This suggests that additional contacts to the protein are formed, either with the substrate or between the two lobes. Bennett and Steitz (1980b) have proposed from crystallographic studies of the hexokinase A-glucose complex, which crystallises in the 'closed' conformation, that two regions of the small lobe occupy positions that cannot be

accounted for by the assumption that this lobe makes a simple gross movement upon glucose binding. These regions are proposed to be around residues 139 and 155. The γ-oxygen of Ser-138 is proposed to make an additional interaction with the glucose 6'-hydroxyl group. This serine is identified only from the sequence of Kuromizu et al (1979) and we have no unequivocal evidence for the identity of this group in hexokinase B. Residue 155 may make an additional interaction with the glucose 2'-hydroxyl group, a group whose coordination is not essential (Sols et al, 1958) yet whose blockage by N-acylation is inhibitory (Maley and Lardy, 1955). This residue is proposed by this author to be valine. The X-ray sequence here proposes alanine. The interaction with the hydroxyl group of glucose would presumably occur to a group on the backbone of the hexokinase molecule.

A lysine residue at position 283, which is in helix αF of the large lobe, may make contact with the small lobe in the closed structure. This residue remains unidentified.

No observable changes in the large lobe of the hexokinase-glucose complex bring the catalytic ATP-binding site closer to the glucose-binding site. The model is unable to suggest any additional contacts made by ATP in this site relative to the model of the apoenzyme. It therefore remains unclear what are the causes of the conformational changes that take place during catalysis - both the glucose-induced conformational change of the ATP-binding site and the ATP-induced tightening of the glucose-binding site. Presumably the elucidation of these changes awaits both the complete amino acid sequence and a knowledge of all the contacts - hydrogen bonds, salt bridges and hydrophobic bonds-that characterise the binary and tertiary hexokinase substrate complexes.

REFERENCES

Abbott, S.J., Jones, S.R., Weinman, S.A. and Knowles, J.R. (1978): J. Amer. Chem. Soc. 100, 2558

Ambler, R.P. (1963): Biochem. J. 89, 349-378

Anderson, W.F., Fletterick, R.J. and Steitz, T.A. (1974): J. Mol. Biol. 86, 261-9

Anderson, W.F. and Steitz, T.A. (1975): J. Mol. Biol. 92, 279-287

Anderson, C.M., McDonald, R.C. and Steitz, T.A. (1978a): J. Mol. Biol. 123, 1-13

Anderson, C.M., Stenkamp, R.E. and Steitz, T.A. (1978b): J. Mol. Biol. 123, 15-34

Anderson, C.M., Stenkamp, R.E. McDonald, R.C. and Steitz, T.A. (1978c): J. Mol. Biol. 123, 207-219

Banks, R.D., Blake, C.C.F., Evans, P.R., Haser, R., Rice, D.W., Hardy, G.W., Merrett, M. and Phillips, A.W. (1979): Nature (London) 279, 773-777

Barnard, E.A. (1975): Methods in Enzymol. 42, 6-20

Bennett, W.S. Jr. and Steitz, T.A. (1978): Proc. Natl. Acad. Sci. U.S.A. 75, 4848-4852

Bennett, W.S. Jr. and Steitz, T.A. (1980a): J. Mol. Biol. 140, 183-209

Bennett, W.S. Jr. and Steitz, T.A. (1980b): J. Mol. Biol. 140, 211-230

Blakley, R. (1977): Methods in Enzymol. 47, 260-263

Blättler, W.A. and Knowles, J.R. (1979): J. Amer. Chem. Soc. 101:2, 510-511

Bohnensack, R. and Hoffmann, E. (1969): Eur. J. Biochem. 9, 534-541

Bornstein, P. (1969): Biochem. Biophys. Res. Commun. 36, 957-964

Bornstein, P. and Balian, G. (1977): Methods in Enzymol. 47, 132-145

Braunitzer, G., Schrank, B. and Ruhfus, A. (1970): Hoppe-Seyler's Z. Physiol. Chem. 351, 1589-1593

Britton, H.G. and Clarke, J.B. (1972): Biochem. J. 128, 104P

Cleland, W.W. (1977): Adv. in Enzymol. 45, 273-387

Cohn, M. (1963): Biochemistry 2, 623-629

Colowick, S.P., Womack, F.C. and Nielsen, J. (1969): In 'The role of nucleotides for the function and conformation of enzymes', pub. Munksgaard, Copenhagen

Colowick, S.P. and Womack, F.C. (1969): J. Biol. Chem. 244, 774-776

Cornelius, R.D., Hart, P.A. and Cleland, W.W. (1977): Inorg. Chem. 16, 2799-2805

Cornelius, R.D. and Cleland, W.W. (1978): Biochemistry 17, 3279-3285

Cullis, P.M. and Lowe, G. (1978): J. Chem. Soc. Chem. Commun. 512-513

Danenberg, K.D. and Cleland, W.W. (1975): Biochemistry 14, 28-39

Danenberg, P.V. and Danenberg, K.D. (1977): Biochem. Biophys. Acta 480, 351-356

Darrow, R.A. and Colowick, S.P. (1962): Methods in Enzymol. 5, 226- 235

Derechin, M., Rustum, Y.M. and Barnard, E.A. (1972): Biochemistry 11, 1793-1797

Dunaway-Mariano, D. and Cleland, W.W. (1980a): Biochemistry 19, 1496-1505

Dunaway-Mariano, D. and Cleland, W.W. (1980b): Biochemistry 19, 1506-1515

Feldman, I. and Kramp, D.C. (1978): Biochemistry 17, 1541-1547

Fletterick, R.J., Bates, D.J. and Steitz, T.A. (1975): Proc. Natl. Acad. Sci. U.S.A. 72, 38-42

Fromm, H.J. (1969): Eur. J. Biochem. 7, 385-392

Fromm, H.J., Silverstein, E. and Boyer, P.D. (1964): J. Biol. Chem. 239, 3645-3652

Fromm, H.J. and Zewe, V. (1962): J. Biol. Chem. 237, 3027-3032

de la Fuente, G., Lagunas, A. and Sols, A. (1970): Eur. J. Biochem. 16, 226-233

de la Fuente, G. and Sols, A. (1970): Eur. J. Biochem. 16, 234-23

Gancedo, J-M., Clifton, D. and Fraenkel, D.H. (1977): J. Biol. Chem. 252, 4443-4444

Gross, E. and Witkop, B. (1961): J. Amer. Chem. Soc. 83, 1510

Hess, B., Boiteux, A. and Kruger, J. (1969): Adv. Enz. Reguln. 7, 149-167

Hoggett, J.G. and Kellett, G.L. (1976a): Eur. J. Biochem. 66, 65-

Hoggett, J.G. and Kellett, G.L. (1976b): Eur. J. Biochem. 68, 347-353

Hohnadel, D.C. and Cooper, C. (1972): Eur. J. Biochem. 31, 180-185

Hunkapiller, M.W. and Hood, L.E. (1978): Biochemistry 17, 2124-2133

Jacrot, B., Cusack, S., Dianoux, A.J. and Engelman, D.M. (1982): Nature (London) 300, 85-86

Jaffe, E.K. and Cohn, M. (1979): J. Biol. Chem. 254, 19839-10845

Janson, C.A. and Cleland, W.W. (1974): J. Biol. Chem. 249, 2572-2574

Jarvest, R.L., Lowe, G. and Potter, B.V.L. (1980): J. Chem. Soc. Chem. Commun. 1142-1145

Kaji, A. and Colowick, S.P. (1965): J. Biol. Chem. 240, 4454-4462

Kaji, A., Trayser, K.A. and Colowick, S.P. (1961): Fed. Proc. Fed. Amer. Soc. Exp. Biol. 20, 233

Kenkare, U.W. and Colowick, S.P. (1965): J. Biol. Chem. 240, 4570-4576

Kopperschlager, G. and Hoffmann, E. (1969): Eur. J. Biochem. 9, 419-423

Kosow, D.P. and Rose, I.A. (1970): J. Biol. Chem. 245, 198-204

Kosow, D.P. and Rose, I.A. (1971): J. Biol. Chem. 246, 2618-2625

Kuromizu, K., Cheng, L-Y., Takahashi, N., Fletcher, P., Inagami, T. and Colowick, S.P. (1979): Fed. Proc. Fed. Amer. Soc. Exp. Biol. 38, 324

Laemmli, U.K. (1970) Nature (London) 227, 672-676

Lazarus, N.R., Ramel, A.H., Rustum, Y.R. and Barnard, E.A. (1966): Biochemistry 5, 4003-4016

Lowe, G. (1981): Biochem. Soc. Trans. 9.3, 207-208

Lowe, G. and Potter, B.V.L. (1981): Biochem. J. 199, 227-233

Maitra, P.K. (1975): Methods in Enzymol. 42, 26-30

Maley, F. and Lardy, H.A. (1955): J. Biol. Chem. 214, 765-773

Mayes, E.L., Hoggett, J.G. and Kellett, G.L. (1982): Biochem. J. 203, 523-525

McDonald, R.C., Steitz, T.A. and Engelman, D.M. (1979): Biochemistry 18, 338-342

Merritt, E.A., Sundaralingam, M., Cornelius, R.D. and Cleland, W.W. (1978): Biochemistry 17, 3274-3278

Midelfort, C.F. and Rose, I.A. (1976): J. Biol. Chem. 251, 5881-5887

Mildvan, A.S. (1981): Phil. Trans. Roy. Soc. Lond. Ser. B 293, 65-74

Morris, H.R., Dickinson, R.J. and Williams, D.H. (1973): Biochem. Biophys. Res. Commun. 51, 247-255

Nageswara Rao, B.D., Buttlaire, D.H. and Cohn, M. (1976): J. Biol. Chem. 251, 6981-6986

Nageswara Rao, B.D., Cohn, M. and Noda, L. (1978): J. Biol. Chem. 253, 1149-1158

Nageswara Rao, B.D., Kayne, F.J. and Cohn, M. (1979): J. Biol. Chem. 254, 2689-2695

Nageswara Rao, B.D. and Cohn, M. (1981): J. Biol. Chem. 256, 1716-1721

Newsholme, E.A. and Start, C. (1973): 'Regulation in Metabolism', pub. John Wiley and Sons, London

Noat, G., Ricard, J., Borel, M. and Got, C. (1969): Eur. J. Biochem. 11, 106-112

Orr, G.A., Simon, J., Jones, S.R., Chin, G.J. and Knowles, J.R. (1978): Proc. Natl. Acad. Sci. U.S.A. 75, 2230-2233

Otieno, S., Bhargava, A.K., Barnard, E.A. and Ramel, A.H. (1975): Biochemistry 14, 2403-2410

Otieno, S., Bhargava, A.K., Serelis, D. and Barnard, E.A. (1977): Biochemistry 16, 4249-4255

de Pamphilis, M.L. and Cleland, W.W. (1973): Biochemistry 12, 3714-3723

Peters, B.A. and Neet, K.E. (1977): J. Biol. Chem. 252, 5345-5349

Peters, B.A. and Neet, K.E. (1978): J. Biol. Chem. 253, 6826-6831

Pickover, C.A., McKay, D.B., Engelman, D.M. and Steitz, T.A. (1979): J. Biol. Chem. 254, 11323-11329

Pliura, D.H., Schomburg, D., Richard, J.P., Frey, P.A. and Knowles, J.R. (1980): Biochemistry 19, 325-329

Purich, D.L. and Allison, R. (1980): Methods in Enzymol. 64, 1-46

Purich, D.L. and Fromm, H.J. (1972): Arch. Biochem. Biophys. 149, 307-315

Ramel, A.H., Rustum, Y.M., Jones, G.J. and Barnard, E.A. (1971): Biochemistry 10, 3499-3508

Robbins, E.A. and Boyer, P.D. (1957): J. Biol. Chem. 224, 121-135

Rose, I.A. (1979): Adv. in Enzymol. 50, 361-398

Rose, I.A. (1980): Biochem. Biophys. Res. Commun. 94, 573-578

Rose, I.A., O'Connell, E.L., Litwin, S. and Bar-Tana, J. (1974): J. Biol. Chem. 249, 5163-5168

Roustan, C., Brevet, A., Pradel, L-A. and van Thoai, N. (1974): Eur. J. Biochem. 44, 353-358

Rudolph, F.B. and Fromm, H.J. (1970): Biochemistry 9, 4660-4664

Rudolph, F.B. and Fromm, H.J. (1971): J. Biol. Chem. 246, 2104-2110

Rustum, Y.M., Massaro, E.T. and Barnard, E.A. (1971): Biochemistry 10, 3509-3516

Schmidt, J.J. and Colowick, S.P. (1973a): Arch. Biochem. Biophys. 158, 458-470

Schmidt, J.J. and Colowick, S.P. (1973b): Arch. Biochem. Biophys. 158, 471-477

Schroeder, W.A., Shelton, J.B. and Shelton, J.R. (1969): Arch. Biochem. Biophys. 130, 551-556

Schulze, I.T. and Colowick, S.P. (1969): J. Biol. Chem. 244, 2306-2316

Shill, J.P. and Neet, K.E. (1971): Biochem. J. 123, 283-285

Shill, J.P. and Neet, K.E. (1975): J. Biol. Chem. 250, 2259-2268

Shoham, M. and Steitz, T.A. (1980): J. Mol. Biol. 140, 1-14

Shoham, M. and Steitz, T.A. (1982): Biochem. Biophys. Acta 705, 380-384

Solomon, F. and Rose, I.A. (1971): Arch. Biochem. Biophys. 147, 349-350

Sols, A., de la Fuente, G., Villar-Palasi, C. and Asensio, C. (1958): Biochem. Biophys. Acta 30, 92-101

Steitz, T.A. (1971): J. Mol. Biol. 61, 695-700

Steitz, T.A., Fletterick, R.J. and Hwang, K.J. (1973): J. Mol. Biol. 78, 551-561

Steitz, T.A., Fletterick, R.J., Anderson, W.F. and Anderson, C.M. (1976): J. Mol. Biol. 104, 197-222

Steitz, T.A., Anderson, W.F., Fletterick, R.J. and Anderson, C.M. (1977): J. Biol. Chem. 252, 4494-4500

Steitz, T.A. Shoham, M. and Bennett, W.S. Jr., (1981): Phil. Trans. Roy. Soc. Ser. B 293, 43-52

Tanford, C. (1961): 'Physical Chemistry of Macromolecules', Wiley, New York, N.Y.

Tarr, G., Beecher, J.F., Bell, M. and McKean, D.J. (1978): Anal. Biochem. 84, 622-627

Thomas, D.W., Das, B.C., Gero, S.D. and Lederer, E. (1968): Biochem. Biophys. Res. Commun. 32, 519-525

Trayser, K.A. and Colowick, S.P. (1961): Arch. Biochem. Biophys. 94, 161-168

Viola, R.E. and Cleland, W.W. (1978): Biochemistry 17, 4111-4117

Viola, R.E., Raushel, F.M., Rendina, A.R. and Cleland, W.W. (1982): Biochemistry 21, 1295-1302

Wilkinson, K.D. and Rose, I.A. (1979): J. Biol. Chem. 254, 12567-12572

Wilkinson, K.D. and Rose, I.A. (1980): J. Biol. Chem. 255, 7569-7574

Wilkinson, K.D. and Rose, I.A. (1981): J. Biol. Chem. 256, 9890-9894

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

Womack, F.C., Welch, M.K., Nielsen, J. and Colowick, S.P. (1973): Arch. Biochem. Biophys. 158, 451-457

Womack, F.C. and Colowick, S.P. (1979): Proc. Natl. Acad. Sci. U.S.A. 76, 5080-5084

Wurster, B. and Hess, B. (1974): FEBS Lett. 40, S112-118

Zewe, V., Fromm, H.J. and Fabiano, R. (1964): J. Biol. Chem. 239, 1625-1634.

Bioscience Reports **3**, *963-971 (1983)*
Printed in Great Britain

Location in the yeast hexokinase structure of residues related to the enzyme activity

Andrew J. W. GRAY, Nicola M. BARRON & Eric A. BARNARD*

Department of Biochemistry,
Imperial College of Science & Technology, London SW7 2AZ, U.K.

(Received 31 August 1983)

Seven residues implicated as acting directly in substrate binding in yeast hexokinase B have been identified in the crystallographic structure by chemical sequencing. The cysteine which is regarded as a residue critically maintaining the active conformation of yeast hexokinase has been selectively labelled and likewise located in the structure. In some parts of the amino acid sequence predicted from the high-resolution electron density map it is found that alignments of chemically sequenced peptides can be made unambiguously; however, the extent of matching to the predicted sequence varies considerably along the chain.

Since the amino acid sequence of neither of the two isoenzymes of yeast hexokinase (EC 2.7.1.1), A and B (Lazarus et al., 1966), has been determined, the structural basis for its catalytic mechanism has not been clearly established. Crystallographic analyses at 2.1-Å resolution have determined the arrangement of the polypeptide chains and led to a postulated amino acid sequence, suggested to be probably about 60% correct (Anderson et al., 1978a,b). This 'X-ray sequence' has an amino acid composition which shows considerable discrepancy from the reported composition for over half of the amino acid types and has 17% of its total in unidentified residues. Also missing from it is at least an N-terminal segment of 11 residues that has been chemically sequenced by Schmidt and Colowick (1973).

The enzyme, a homo-dimer, possesses 4 cysteines per subunit of M_r 51 000, and no disulphides (Lazarus et al., 1968; Schmidt & Colowick, 1973; Jones et al., 1975). One of these thiols is apparently essential for the enzyme to be active, insofar as any substitution there, even by a small uncharged group such as -CN, inactivates completely. Moreover, a hexose-based affinity reagent alkylates this cyteine only (Jones et al., 1975; Otieno et al., 1977). It was concluded that this thiol (thiol I) is close to the active site in the enzyme-hexose complex in solution above 31°C, but that there was no evidence that it is directly involved in catalysis or substrate binding. The other 3 thiols have been shown to be definitely non-essential (Otieno et al., 1977). We have now identified these various cysteines in the X-ray sequence.

* *To whom correspondence should be addressed.*

By means of difference electron density maps based on the hexokinase, hexokinase-glucose, and hexokinase-adenine nucleotide structures and model-building, Bennett and Steitz (1980) and Shoham and Steitz (1980) found hydrogen-bonding to the substrates to occur at a number of positions in the crystallographic structure, defining the active site. We have obtained chemical evidence for the identities of these active-site amino acids. The fitting of peptides to the X-ray sequence is also considered.

Materials and Methods

Peptide sequencing was performed (except where stated) on a Beckman sequencer fitted with the modified vacuum system and cold-trap of Hunkapiller and Hood (1978). Polybrene (3 mg) was added to all samples, and peptides with a significant lysine content were coupled with 4-sulphophenylisothiocyanate in the sequencer cup, to prevent premature washout (Blakley, 1977). A modified Beckman 0.1 M Quadrol program with double coupling, cleavage, and extraction steps was employed. The repetitive yield was about 95%. Reagents for sequencing were repurified according to Hunkapiller and Hood (1978) before sequencing the long peptides derived from hydroxylamine cleavage. Phenylthiohydantoin-amino acids were analysed on a DuPont Zorbax column on a Varian 5000 liquid chromatograph fitted with the Waters 'Intelligent Sample Processor', in a 0.033 M sodium acetate (pH 5.3)/ethanol system.

Hexokinase B (EC 2.7.1.1) was purified from fresh pressed bakers' yeast (United Yeast Co. Ltd.) by a scaled-up version of method A of Barnard (1975). The final specific activity was about 600 units/mg, at 25°C. The protein showed a single band in SDS/PAGE of apparent M_r 51 000. N-Terminal sequencing of the purified protein showed that it was intact and possessed the 11 N-terminal residues lacking in commercial samples of this enzyme.

Hexokinase (5 mg/ml) was fully S-carboxymethylated by treatment with dithiothreitol (2 mM) and alkylation under N_2 with 10 mM iodo[2-^{14}C]acetic acid, in 6 M guanidine/Tris (pH 8.5) medium. The carboxymethylated protein was cleaved with either CNBr (in 100-fold molar excess over methionines) in 70% formic acid or with hydroxylamine (Bornstein & Balian, 1977). The peptides produced were separated on a Waters Liquid Chromatograph, firstly by molecular sizing on a dual Waters I-125 protein analysis column in 5% acetic acid / 10% n-propanol, and then by reverse-phase high-performance liquid chromatography (h.p.l.c.) on a Waters μ-Bondapak phenyl column in a gradient of acetonitrile in 0.1% trifluoroacetic acid.

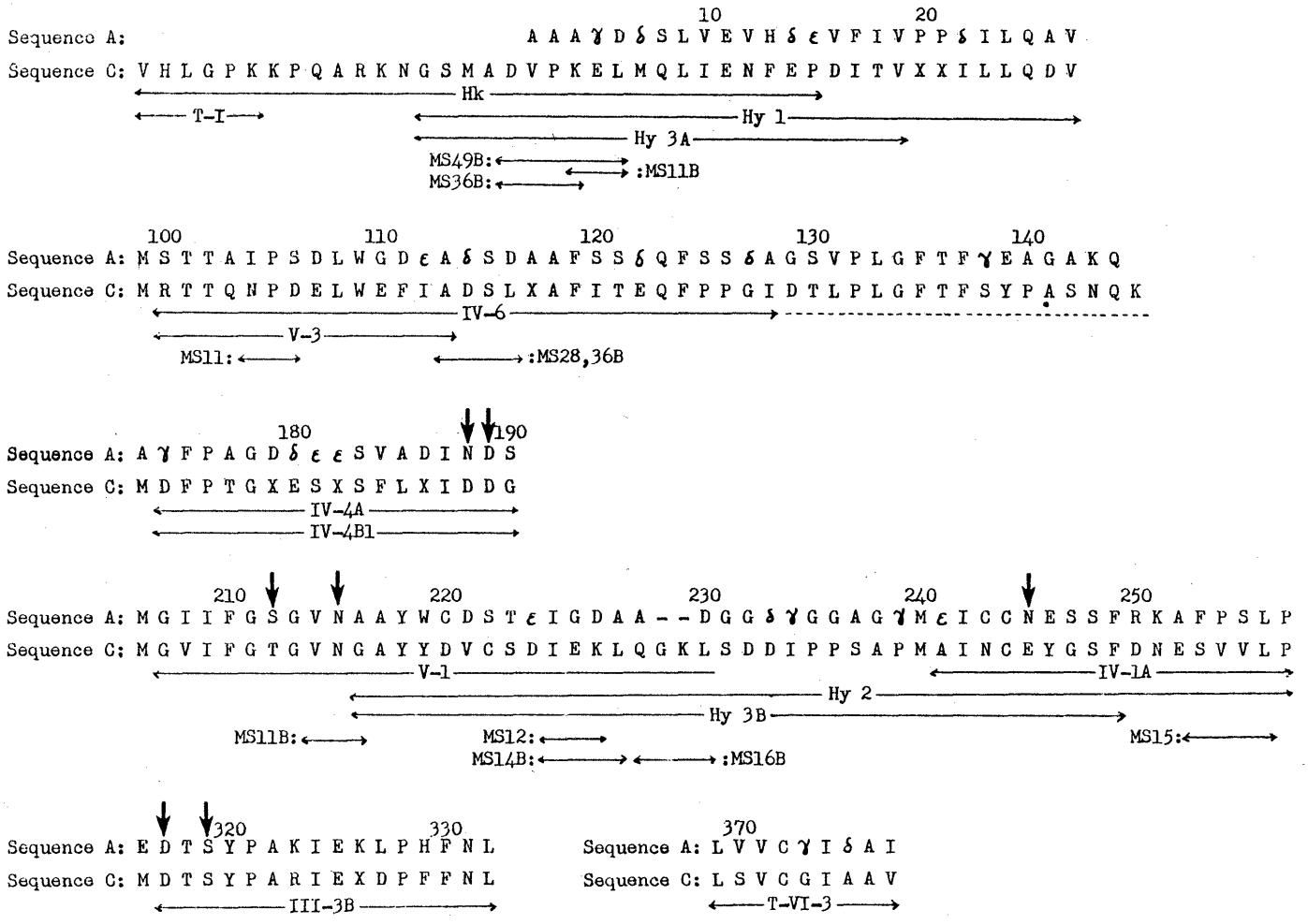
Hexokinase radiolabelled selectively at thiol I was prepared after pre-alklylation of the enzyme (2 mg/ml) with iodoacetic acid (7 mM) in the presence of mannose (6 mM) in 0.05 M glycine / 0.1 M NaCl, pH 8.6 at 35°C for 3 h. Reagents were fully removed by dialysis against the latter buffer and iodo[2-^{14}C]acetic acid was added to 7 mM. Reaction was for 80 min at 35°C, maintaining the pH with solid Tris base. The product was then fully S-carboxymethylated (with unlabelled iodoacetate) in guanidine as above. After CNBr cleavage as above, the peptides were separated on a column (2.5 x 80 cm) of Biogel P-30 (-400 mesh) in 20% formic acid. Reverse-phase h.p.l.c.

was applied to the material in each of the 3 labelled peaks obtained, using a Waters μ-Bondapak C_{18} column in 5% acetic acid / n-propanol. The label in the P-30 void-volume, peak I, was in incompletely cleaved large peptides. Peak II, centred at 250 ml, was resolved by the h.p.l.c. into 3 peptides, of which only the least polar (BII-3) contained ^{14}C. From labelled peak III, centred at 278 ml, the label, much weaker, was in only one peptide (BIII-1).

To derive short peptides suitable for confirmatory sequencing, native hexokinase (2 μmol) was digested with elastase (1:50) in 50 mM ammonium bicarbonate, pH 8.5 at 37°C for 3 h. After freeze-drying, the digest was fractionated on a column (30 x 1 cm) of Locarte cation-exchange resin at 60°C with a nine-chamber system of pyridine/acetic acid buffers (pH 2.15 to 6.5; pyridine gradient 0.025 to 8.0 M). Suitable peptide fractions were acetylated, permethylated, and analysed by electron-impact mass spectrometry on a Kratos MS50 instrument, using the procedures of Morris and Dell (1975). Alternatively, carboxymethylated protein samples were cleaved with chymotrypsin-free trypsin (1:100) and the peptides separated by gel filtration on Sephadex G-50 or Biogel P-4 followed by sequential electrophoresis and chromatography on paper. Peptides were analysed by the dansyl-Edman reaction up to 7 cycles. Those peptides which were well resolved were also analysed by mass spectrometry as above. Cysteine was confirmed by the radioactivity in all procedures. Iodoacetic acid and iodo[2-^{14}C]acetic acid (Amersham; diluted to 2 Ci/mol) were recrystallized before use (Jones et al., 1975). All peptides throughout were sequenced twice. Methods not specified were as in Shotton and Hartley (1973) or Jones et al. (1975).

Results and Discussion

It was found necessary, for reasons which will emerge later, to seek peptides as long as possible for fitting against the X-ray sequence. Some of the products of CNBr cleavage were suitable. After such cleavage of hexokinase B, [^{14}C]carboxymethylated at all its cysteines, the peptides were separated by high-pressure gel permeation chromatography into six pools of descending molecular size, I to VI, and peptides of suitable sizes were further resolved by reverse-phase h.p.l.c. Peptides shown to be pure at this stage were used for automated sequencing. These mostly had lengths of 20-50 residues and were in pool IV, while three were in pool V. Pool III was shown to contain many of the sequences of pools IV and V combined in longer peptides. To obtain some longer peptides, the very restricted cleavage by hydroxylamine (at Asn-Gly bonds only) was applied to the protein. This gave three pools upon high-pressure gel permeation (plus a small N-terminal fragment, which was not isolated); their peptides were analysed by h.p.l.c. and partly sequenced. The pool-1 peptide ran from a glycine at a position -5 with reference to the X-ray sequence (as discussed below) to the C-terminus, while pools 2 and 3 contained the peptides generated from the former and one or two other such cleavage sites present in the protein. The larger peptides purified from the CNBr cleavage and the sequences obtained from the long hydroxylamine peptides could be aligned to the X-ray sequence without ambiguity (Fig. 1). In the regions of interest here, confirmations were obtained by screening the many small peptides derived from an


10 20
Sequence A: A A A γ D δ S L V E V H δ ε V F I V P P δ I L Q A V
Sequence C: V H L G P K K P Q A R K N G S M A D V P K E L M Q L I E N F E P D I T V X X I L L Q D V
Hk
T-I
Hy 1
Hy 3A
MS49B:
:MS11B
MS36B:
100 110 120 130 140
Sequence A: M S T T A I P S D L W G D ε A δ S D A A F S S δ Q F S S δ A G S V P L G F T F γ E A G A K Q
Sequence C: M R T T Q N P D E L W E F I A D S L X A F I T E Q F P P G I D T L P L G F T F S Y P A S N Q K
IV-6
V-3
MS11:
:MS28,36B
180 190
Sequence A: A γ F P A G D δ ε ε S V A D I N D S
Sequence C: M D F P T G X E S X S F L X I D D G
IV-4A
IV-4B1
210 220 230 240 250
Sequence A: M G I I F G S G V N A A Y W C D S T ε I G D A A - - D G G δ γ G G A G γ M ε I C C N E S S F R K A F P S L P
Sequence C: M G V I F G T G V N G A Y Y D V C S D I E K L Q G K L S D D I P P S A P M A I N C E Y G S F D N E S V V L P
V-1
IV-1A
Hy 2
Hy 3B
MS11B:
MS12:
MS15:
MS14B:
:MS16B
320 330
Sequence A: E D T S Y P A K I E K L P H F N L
Sequence C: M D T S Y P A R I E X D P F F N L
III-3B
370
Sequence A: L V V C γ I δ A I
Sequence C: L S V C G I A A V
T-VI-3

elastase digest of hexokinase and sequencing suitable ones by mass spectrometry (Fig. 1).

The residues which had been found crystallographically to be bonded to groups on the substrate (Anderson et al., 1978c; Bennett & Steitz, 1980; Shoham & Steitz, 1980) were almost all identified in the sequence (Table 1). Some corrections have to be made to the X-ray assignments, but all of the residues found at the positions which had been implicated are consistent with their postulated roles. It is interesting that the residue at position 189, seen as Asx in the X-ray structure, has been proposed to be Asp in a catalytic scheme, in which a buried carboxyl activates the 6-OH of glucose (Anderson et al., 1978c; Viola & Cleland, 1978). It is indeed found to be Asp.

To identify the 'apparently essential' cysteine, hexokinase B was labelled selectively at thiol I (Otieno et al., 1977) by pre-blockade with iodoacetate in the presence of the substrate mannose, without activity loss, followed by reaction with [^{14}C]iodoacetate at 35°C in the absence of substrate. After 90% inactivation, the incorporation of label was 0.95 mol per subunit, all as [^{14}C]carboxymethyl-cysteine. This product was cleaved by CNBr and the labelled peptides separated by gel filtration and h.p.l.c. One strongly radioactive peptide was obtained (BII-3). This contained one residue of [^{14}C]carboxymethyl-cysteine, and gave a sequence of 13 amino acids around thiol I (Table 2).

Another cysteine was found to be lightly labelled. The other [^{14}C]peptide (BIII-1) separated in the h.p.l.c. had 10-20% of the specific radioactivity of BII-3 and gave a 19-residue sequence around a second cysteine (Table 2). This thiol is presumed to be one of the two (thiols III and IV) which react slowly with iodoacetate, even

Fig. 1. Comparison of parts of the X-ray sequence (sequence A) and the corresponding chemically derived sequence (sequence C). The numbering is that of Anderson et al. (1978b) for their X-ray sequence. In the latter the pairs Asp and Asn or Glu and Gln were not distinguished, but we have changed the X-ray sequence to show the acid or amide where the chemical sequence gives it, and otherwise have put the acid. γ, δ, and ε are side-chains which could not be identified in the X-ray sequence but were approximated by chains of 2, 3, or 4 carbons respectively. Peptide alignments are shown which gave the maximum degree of fit. **X** denotes a residue present but whose identity was unconfirmed in the chemical sequence. The broken line signifies a peptide sequenced by Kuromizu et al. (1979) after [^{32}P]phosphorylation (at Ser-138) by ATP in the presence of lyxose; note that this sequence came from hexokinase A. Note the possible evidence for duplication at 138-145 and 319-326.
Hk = intact hexokinase; **T** = tryptic; **Hy** = hydroxylamine cleavage; **MS** = determined by mass spectrometry; Roman numerals, by CNBr. Vertical arrows show the identified active-site residues.

Table 1. The residues predicted to make substrate-binding interactions and their chemical identification

The predictions are by Anderson et al. (1978c) and Shoham and Steitz (1980) on the basis of electron density maps and a proposed interaction scheme.

Substrate group interacting		Position of contact	Predicted residue	Corresponding identified residue
Glucose:	3-OH & 4-OH	188	Asn	Asp-188
	3-OH	245	Asn	Glu-245
	4-OH	215	Asn	Asn-215
	4-OH & 6-OH	189	Asp	Asp-189
Ribose	2'-OH	317	Asx	Asp-317
	3'-OH	319	Ser	Ser-319
β and γ phosphates of ATP		212*	Ser	Thr-212
		393	Ser	Not identified

* Deduced to be binding to the backbone nitrogen, not to the side-chain.

above the transition temperature of 31°C (Jones et al., 1975), and are not accessible in the enzyme-hexose complex (Otieno et al., 1977). It is seen that the labelled peptides BII-3 and BIII-1 fit definitively around Cys-244 and position 222 of the X-ray sequence (Table 2). The chemical sequence in that region (Fig. 1) was thus independently confirmed.

Careful amino acid analysis on the preparation used here confirmed the previous evidence (see Introduction) that there are only 4 cysteines per subunit of hexokinase, whereas the X-ray sequence shows 5. Two of the latter are adjacent, at positions 243 and 244 (Anderson et al., 1978b), but there we find only a single cysteine, at 244, and no other anywhere in that region (Fig. 1). The first cysteine in the X-ray sequence, at 220, corresponds to a cysteine which is close to that position (222) in the chemical sequence, although there are considerable differences in this region (Table 2). A further cysteine was present in a labelled peptide purified from a tryptic digest of the fully [^{14}C]carboxymethylated hexokinase. The sequence of its first 9 residues (confirmed by mass spectrometry) corresponded well to that around Cys-372 in the X-ray sequence (Fig. 1). We have thus located 3 of the 4 cysteines present. The other cysteine in the X-ray sequence is at position 378 therein. The tryptic peptide containing Cys-372 in fact contained another cysteine near 378 but the C-terminal sequence of that peptide could not be unambiguously determined. The identification of thiol I as being on residue 244 is compatible with the situation of that residue, which is on the same face of the protein as the glucose-binding site and adjacent to an active-site (glucose-binding) residue (Glu-245; Table 1), whereas, in the complex of hexokinase with an inhibitory hexose derivative (toluoylglucosamine) the sulphur of Cys-244 is about 11 Å from the sugar hydroxyls (Anderson et al., 1978c); this distance and the

Table 2. [^{14}C]Carboxymethyl-cysteine-containing peptides isolated from selectively-labelled hexokinase

CM = Carboxymethyl. Eps is a residue not identified in the X-ray sequence but having a side-chain density similar to a 4-carbon chain. The numbering is that used in the X-ray sequence by Anderson et al. (1978b) and does not represent the true number of residues from the N-terminus of the protein (see text). In a meeting report (Barnard et al., 1981) a seven-residue sequence at 241 was given incorrectly. The peptides did not terminate at the last residue sequenced.

	241			CM-								252
Thiol I sequence:	Ala	Ile	Asn	Cys	Glu	Tyr	Gly	Ser	Phe	Asp	Asn	Glu
X-ray sequence:	Eps	Ile	Cys	Cys	Asn	Glu	Ser	Ser	Phe	Arg	Lys	Ala
			207									
Slightly-labelled thiol:	Gly	Val	Ile	Phe	Gly	Thr	Gly	Val	Asn	Gly		
X-ray sequence:	Gly	Ile	Ile	Phe	Gly	Ser	Gly	Val	Asn	Ala		
							CM-			225		
	Ala	Tyr	Tyr	Asp	Val	Cys	Ser	Asp	Ile			
	Ala	Tyr	Trp	Cys	Asp	Ser	Thr	Eps	Ile			

accessibility of the thiol are assumed to change in the transition at 31°C above which thiol I is reactive (Otieno et al., 1977).

Anderson et al. (1978c) proposed a cysteine at position 243 to be thiol I, since that residue makes extensive hydrogen-bonded contacts with three active-site residues, at positions 188, 215, and 245. We find this residue in fact to be asparagine, which fits better to that hydrogen-bonding scheme. Modification of the sulphur of position 244 must also disrupt the active-site structure.

The X-ray sequence begins at a position numbered 2 by Anderson et al. (1978b), which we find to be located some way in from the true N-terminus (Fig. 1). The alignment made with the start of the X-ray sequence is tentative, because of the poor initial agreement between the sequences. N-terminal sequencing of the intact protein showed an initial 11-residue segment to be present that is identical with the N-terminal peptide found by Schmidt and Colowick (1973) to be readily released by tryptic cleavage of native hexokinase B. Following this is a further stretch which is found from both the N-terminal protein sequencing and peptides isolated after cleavages, before any matching can be made with the start of the X-ray sequence. The structure used for the latter is known to be of an autolytic product and to have some of the residues in the new N-terminal region disordered (Anderson et al., 1978a), explaining the gap and the poor fit to the X-ray sequence there. It is clear that, due to this and to internal gaps (shown by dashes in Fig. 1) in the X-ray sequence relative to the chemical sequence, the true sequence will need to be substantially renumbered.

The degree of ordering of the structure of hexokinase B in the crystal varies in different regions of the chain (Anderson et al., 1978a,b), and good agreement can be found between the X-ray and

chemical sequences in a few regions tested here but agreement is poor in a number of others. The regions of poor agreement could only be fitted without ambiguity by using long peptides which run on to one of the short stretches of nearly correct prediction by X-ray or give good overlaps with such peptides. It was, therefore, not possible to use all of the peptides generated, to fit in an unambiguous scheme and chemical sequence to the X-ray sequence, due to the necessarily imprecise character of the latter. This limited the extent of sequence obtained purely by reference to the X-ray prediction. If (assuming the gene is unavailable) one wishes to avoid full chemical sequencing (employing secondary cleavages and overlaps) of a protein because its high-resolution electron density map is available, then it may be feasible in some cases to fit all of a set of peptides by matching electron densities directly, as was accomplished successfully for phosphoglycerate kinase by Banks et al. (1979). However, the alternative strategy of predicting an approximate X-ray sequence and then looking for peptides in a single digest which match in sequence to it did not provide a rapid solution here. The prediction approach both makes intrinsic any errors in the X-ray identification and fails to optimize the actual electron density data. It may in general be ambiguous for such chains, of the order of 50 000 mol.wt. or above. On the other hand, that strategy was the most convenient for chemically identifying the active-site residues by reference to the X-ray structure, which is well defined at those points by difference mapping.

Acknowledgements

This work was supported by grants from the Science and Engineering Research Council and the Medical Research Council. We are grateful to Dr. Anne Dell for interpretation of the mass spectra, to Dr. Maria Blair for isolation of radioactive tryptic peptides, and to Dr. Rudolf Shipolini for expert help with the automated sequencing. N.M.B. held a University of London Postgraduate Studentship.

References

Anderson CM, McDonald RC & Steitz TA (1978a) J. Mol. Biol. **123**, 1-13.

Anderson CM, Stenkamp RE & Steitz TA (1978b) J. Mol. Biol. **123**, 15-24.

Anderson CM, Stenkamp RE, McDonald RC & Steitz TA (1978c) J. Mol. Biol. **123**, 207-219.

Banks RD, Blake CCF, Evans PR, Haser R, Rice DW, Hardy GW, Merrett M & Phillips AW (1979) Nature **279**, 773-777.

Barnard EA (1975) Methods Enzymol. **42**, 6-20.

Barnard EA, Barron NM, Dell A & Gray AJW (1981) Biochem. Soc. Trans. **9**, 209-212.

Bennett WS & Steitz TA (1980) J. Mol. Biol. **140**, 211-230.

Blakley RL (1977) Methods Enzymol. **47**, 260-263.

Bornstein P & Balian G (1977) Methods Enzymol. **47**, 132-138.

Hunkapiller MW & Hood LE (1978) Biochemistry **17**, 2124-2133.

Jones JG, Otieno S, Barnard EA & Bhargava AK (1975) Biochemistry **14**, 2396-2403.
Kuromizu K, Cheng L-Y, Takahashi N, Fletcher P, Inagami T & Colowick SP (1979) Fed. Proc. Fed. Am. Soc. Exp. Biol. **38**, 324.
Lazarus NR, Ramel AH, Rustum YM & Barnard EA (1966) Biochemistry **5**, 4003-4016.
Lazarus NR, Derechin M & Barnard EA (1968) Biochemistry **7**, 2390-2400.
Morris HR & Dell A (1975) in Instrumentation in Amino Acid Sequence Analysis (Perham RN, ed.), pp 147-191, Academic Press, New York.
Otieno S, Bhargava AK, Serelis D & Barnard EA (1977) Biochemistry **16**, 4249-4255.
Schmidt JJ & Colowick SP (1973) Arch. Biochem. Biophys. **158**, 458-470.
Shoham M & Steitz TA (1980) J. Mol. Biol. **140**, 1-14.
Shotton DM & Hartley BS (1973) Biochem. J. **131**, 643-675.
Viola RE & Cleland WW (1978) Biochemistry **17**, 4111-4117.