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 cysteine 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.

©1983 The Biochemical Society
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. Phenylthiodyantoin-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 13 (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[2-14C]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[2-14C]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 $\mu$-Bondapak C$\text{18}$ 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 $\mu$mol) 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
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 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.

<table>
<thead>
<tr>
<th>Substrate group interacting</th>
<th>Position of contact</th>
<th>Predicted interacting residue</th>
<th>Corresponding identified residue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose: 3-OH &amp; 4-OH</td>
<td>188</td>
<td>Asn</td>
<td>Asp-188</td>
</tr>
<tr>
<td>3-OH</td>
<td>245</td>
<td>Asn</td>
<td>Glu-245</td>
</tr>
<tr>
<td>4-OH</td>
<td>215</td>
<td>Asn</td>
<td>Asn-215</td>
</tr>
<tr>
<td>4-OH &amp; 6-OH</td>
<td>189</td>
<td>Asp</td>
<td>Asp-189</td>
</tr>
<tr>
<td>Ribose 2'-OH</td>
<td>317</td>
<td>Asx</td>
<td>Asp-317</td>
</tr>
<tr>
<td>3'-OH</td>
<td>319</td>
<td>Ser</td>
<td>Ser-319</td>
</tr>
<tr>
<td>β and γ phosphates of ATP</td>
<td>212*</td>
<td>Ser</td>
<td>Thr-212</td>
</tr>
<tr>
<td></td>
<td>393</td>
<td>Ser</td>
<td>Not identified</td>
</tr>
</tbody>
</table>

* 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 [14C]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 1 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.

<table>
<thead>
<tr>
<th>241</th>
<th>252</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thiol I sequence: Ala Ile Asn Cys Glu Tyr Gly Ser Phe Asp Asn Glu</td>
<td>CM-</td>
</tr>
<tr>
<td>X-ray sequence: Eps Ile Cys Asn Glu Ser Ser Phe Arg Lys Ala</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>207</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slightly-labelled thiol: Gly Val Ile Phe Gly Thr Gly Val Asn Gly</td>
</tr>
<tr>
<td>X-ray sequence: Gly Ile Ile Phe Gly Ser Gly Val Asn Ala</td>
</tr>
</tbody>
</table>

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

Hunkapiller MW & Hood LE (1978) Biochemistry 17, 2124-2133.
Viola RE & Cleland WW (1978) Biochemistry 17, 4111-4117.