Journal of Computer-Aided Molecular Design, 6 (1992) 1–18
ESCOM

J-CAMD 147

A prediction of the three-dimensional structure of maize $NADP^+$-dependent malate dehydrogenase which explains aspects of light-dependent regulation unique to plant enzymes

Richard M. Jackson*, Richard B. Sessions and J. John Holbrook

Molecular Recognition Centre and Department of Biochemistry, University of Bristol School of Medical Sciences, University Walk, Bristol BS8 1TD, U.K.

Received 20 July 1991
Accepted 19 August 1991

Key words: NADP-malate dehydrogenase; Molecular model; C-terminal cystine loop; C-terminal 'internal inhibitor'; Active site cysteine

SUMMARY

A model has been built for the plant NADP-malate dehydrogenase from *Zea mays*, a key enzyme in photosynthesis, which undergoes light-dependent regulation. The model was based on sequence and presumed structural homology to the known three-dimensional structure of mammalian porcine cytosolic NAD-malate dehydrogenase. A cystine-loop present in an extended C-terminal region of plant NADP-malate dehydrogenases was modelled using molecular mechanics and computer graphical methods, based on the assumption that a disulphide bridge exists in the inactive form of the enzyme between Cys^{351} and Cys^{363}. The predicted conformation of the intact C-terminal cystine-loop suggests that the extended polypeptide will bind in the active centre and inhibit enzyme activity. Another ionizable cysteine residue in the active site is predicted to control the charge of the catalytic His^{215} and might be responsible for the uniquely tight binding of the positively charged nicotinamide ring of $NADP^+$ in this and other C4 and C3 plant NADP-malate dehydrogenases.

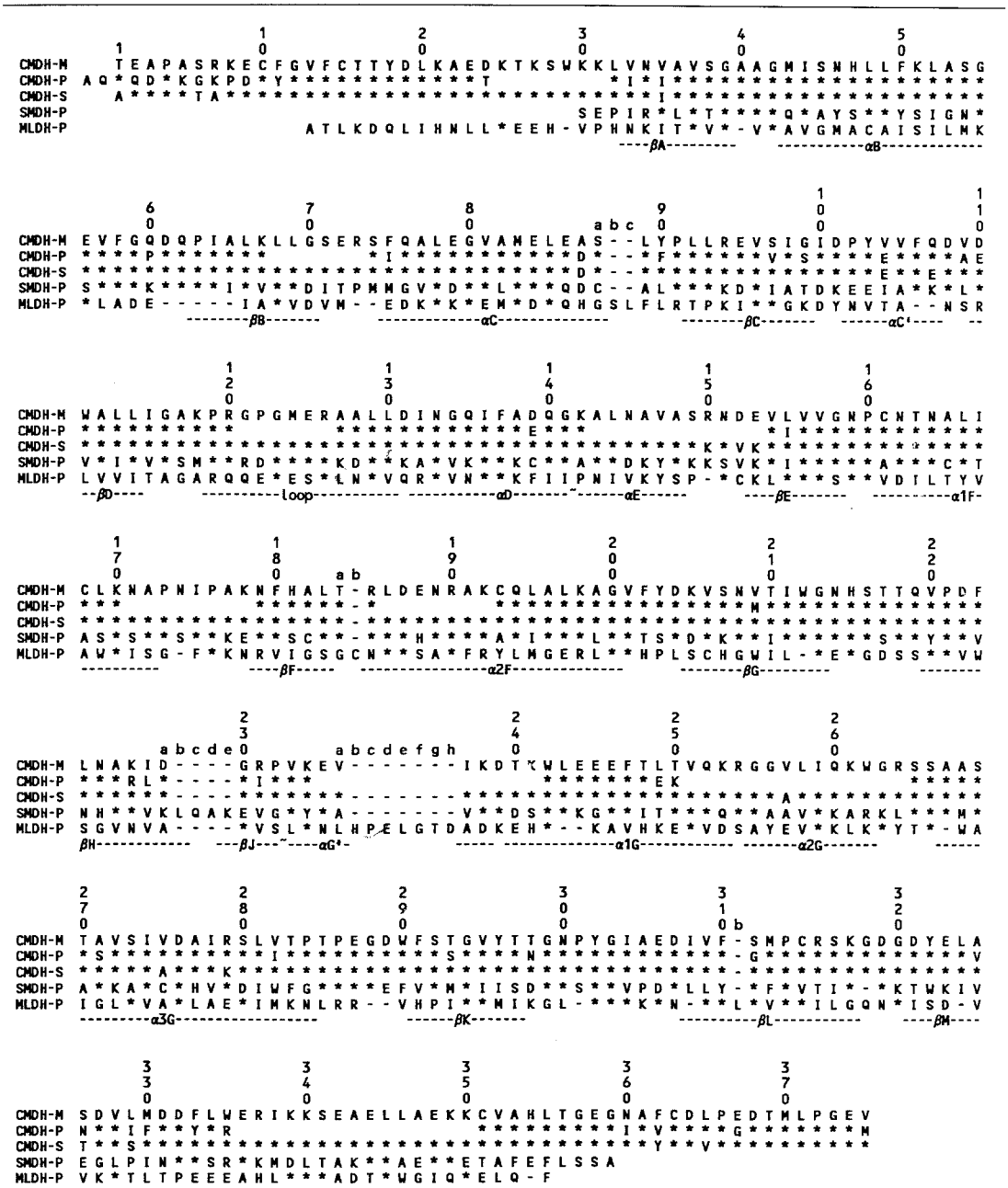
INTRODUCTION

NADP-malate dehydrogenase is a key enzyme in the plant C4 photosynthetic pathway, being involved in the reduction of oxaloacetate (OAA), the primary product of CO_2 fixation, in the mesophyll cell chloroplasts [1]. Unlike the functionally different NAD-specific malate dehydrogenases found in the mitochondria and cytosol of most cells, the plant NADP-malate dehydrogenase is regulated by its redox state, undergoing light-dependent activation in the leaf chloroplast and

* To whom correspondence should be addressed.

TABLE I

ALIGNMENT OF *Z. mays* MDH (CMDH-M), PEA MDH (CMDH-P), *S. vulgare* MDH (CMDH-S), PIG CYTOSOL MDH (SMDH-P) AND PIG MUSCLE LDH (MLDH-P)*. THE SECONDARY STRUCTURE ELEMENTS OF PIG CYTOSOL MDH ARE INDICATED BELOW THE ALIGNED SEQUENCES



*Identities to CMDH-M are represented by an asterisk.

dark-mediated inactivation, via disulphide reduction and formation, respectively [2]. In vivo this chloroplast enzyme is regulated by the photosynthetic electron transfer chain via the ferredoxin-thioredoxin system [3]. Activation can be obtained in vitro by reducing agents such as reduced thioredoxin, dithiothreitol (DTT) or other mercaptans. Similarly inactivation can be achieved by oxidising agents such as oxygen or oxidised thioredoxin [4]. A similar regulatory mechanism has also been proposed for NADP-malate dehydrogenase from C3 plants and several other chloroplast enzymes are known to be regulated via redox control by the ferredoxin-thioredoxin system [5].

The known primary structures of the C4 NADP-malate dehydrogenases of *Zea mays* [6], *Sorghum vulgare* [7], and the partial sequence of the C3 NADP-malate dehydrogenase from pea [8] verify the considerable sequence identity within the plant NADP-malate dehydrogenase family. *Z. mays* and *S. vulgare* are 95% identical in their complete sequence. Pea and *Z. mays* or *S. vulgare* are approximately 84% identical in those regions where sequence is available (78% of the pea sequence is presently available). However, there is also very significant homology between NAD-malate dehydrogenases and NADP-malate dehydrogenases. *T. flavus* and porcine cytosolic malate dehydrogenase have 49% and 43% identity with *Z. mays*, in comparable regions [6] (see Table 1 for sequence alignment). The plant NADP-linked malate dehydrogenases are approximately 40–50 residues longer than their NAD-linked counterparts, due to the presence of both N- and C-terminal polypeptide extensions. The locations of 4 cysteines, two in the N-terminus and two in the C-terminus are identical in the 3 sequences isolated so far. It has been proposed that these residues may be the thioredoxin-regulated thiols [9], and evidence has been presented that the N-terminus disulphide bond between Cys^{10} and Cys^{15} is the regulatory site of the enzyme [10]. It is reduced in going from the inactive to the active form of the enzyme.

A second regulatory mechanism also operates in vivo, in addition to the thioredoxin-mediated redox control. Experiments in reconstituted chloroplast systems indicate that the NADPH/$NADP^+$ ratio may be critical in determining the level of NADP-malate dehydrogenase activity [11,12], since $NADP^+$ inhibits reductive activation of the enzyme. NAD-specific dehydrogenases bind the reduced form of the coenzyme (NADH) much more tightly than the oxidised form (NAD^+), for example in porcine cytosolic NAD-malate dehydrogenase the ratio K_d NADH to K_d NAD^+ is 2×10^{-3} whereas in *Z. mays* NADP-malate dehydrogenase this ratio (K_d NADPH to K_d $NADP^+$) is about 1 in the active form of the enzyme and increases to 100 in the inactive oxidised enzyme [13]. Since the reductive activation of the enzyme is inhibited by over 95% in the presence of $NADP^+$, only at high NADPH/$NADP^+$ ratios will a large proportion of the enzyme be in the active reduced state. Hence the NADPH/$NADP^+$ operates as a fine-control of enzyme activity. The molecular basis for the marked difference in specificity for the redox state of the nicotinamide ring is presently not understood. Therefore the objectives of this study were two-fold:

(i) To use a structure-homology relationship, starting from a suitable known structure to predict the effects of changes in the coenzyme binding domain and active site. Thereby gaining insight into the enzymes' requirement for NADP as opposed to NAD in other malate and lactate dehydrogenases. Also, perhaps more importantly, to understand the basis for the uniquely tight binding of the positively charged nicotinamide ring of $NADP^+$.

(ii) To predict as far as was possible the structure of the N- and C-terminal polypeptide extensions. In particular to investigate, using molecular mechanics, whether a disulphide bond between Cys^{351} in a structurally conserved C-terminal helix (α-H) and Cys^{363} (which is 7 resi-

dues into the C-terminal extension) is sterically and energetically feasible, given the constraints of the remaining enzyme model.

MATERIALS AND METHODS

Coordinates of porcine cytosolic malate dehydrogenase were taken from the Brookhaven Data Base - Resolution 2.5 Å with the correct amino acid sequence [14]. The model was constructed for the dimeric form of the enzyme. The programs INSIGHT (using the Silicon Graphics personal IRIS graphics workstation), MOLEDT and DISCOVER (calculations were performed on an IBM3090-150s) were used to view, build and perform molecular mechanics calculations, respectively (Biosym Technologies Inc.).

Sequence alignment

The amino acid sequence alignment between *Z. mays* and porcine cytosol malate dehydrogenase is essentially the same as that of Fickenscher et al. [9] and Metzler et al. [6]. The following procedures were undertaken in order to assess the correctness of the alignment and find positions of insertions and deletions.

(i) Using a least-squares refinement process, the Cα coordinates of the secondary structural elements of porcine cytosol malate dehydrogenase were superimposed on each of *B. stearothermophilus* lactate dehydrogenase [15], porcine muscle lactate dehydrogenase [16] and dogfish muscle lactate dehydrogenase [17], based on the sequence alignment of 9 malate dehydrogenase and 20 lactate dehydrogenase sequences (Holbrook, J.J., data not shown). This defined the structurally conserved regions of the L-lactate/malate dehydrogenase framework and also variable length loop regions.

(ii) Conserved residues between maize and porcine cytosol malate dehydrogenases were viewed (whilst maintaining the backbone geometry of porcine malate dehydrogenase) to verify that the packing of the model was correct, and hence identify any frameshift errors made in the alignment.

The sequence alignment of the malate dehydrogenases pig cytosol, *Z. mays*, *S. vulgare* and pea as well as pig muscle lactate dehydrogenase is given in Table 1. The only difference from the alignment of that of Metzler et al. [6] occurs in a surface loop between β-H and β-J; this alteration was based on the above least-squares refinement process.

Replacement of side chains

Side chains that differed between the two aligned sequences were replaced using MOLEDT. Torsional angles for new side chains were maintained whenever possible. If no guide atoms existed idealized torsion angles were used for the extended regions. If heavy atoms of a side chain came within 2.5 Å of another heavy atom, torsion angles of the new side chain were adjusted to relieve bad contacts. It became evident during this work that packing of the secondary structure elements in the hydrophobic core of the two structures was remarkably well preserved and sequence identity was highest in this region. In instances where side chains differed they were invariably spatially adjacent to one another, changes thus occurred in pairs, whereby the two new side chains occupied the same volume as their precursors.

Loop building

Only in two cases did the number of surface loop residues vary between the two structures. A surface loop in pig cytosol at positions 229b,c,d and e represents a 4-residue insertion with respect to *Z. mays*. This loop was replaced by the shorter loop of pig muscle lactate dehydrogenase, which has the same loop length as *Z. mays*. Another surface loop in *Z. mays* at position 318 represents a one-residue insertion with respect to pig cytosol; again this corresponds to a similarly longer loop in pig muscle lactate dehydrogenase at this position, which was used for the loop fragment. Fusing these loop fragments was achieved by superimposing the Cα backbone of 4-residue splicers at each end of the loop fragment with the existing pig cytosol framework.

Construction of the C-terminal cystine loop model

The C-terminal region of NAD-linked lactate dehydrogenase and malate dehydrogenases is highly sequence variable (but has a conserved leucine residue in all the sequences), however retains a high degree of secondary structure consensus, as expressed by hydropathy plots and secondary structure prediction methods as implemented by Gribskov et al. [18] and is invariably a long helix (α-H) in the structures of the 6 distinct lactate dehydrogenases and the malate dehydrogenase solved by X-ray crystallography. Hydropathy profiles and secondary structure prediction would suggest retention of the α-H helix in the maize, *S. vulgare* and pea NADP-malate dehydrogenases, followed by a turn Thr-Gly-Glu-Gly and a 6-residue β-strand. The remaining 10 residues of the C-terminus have no identifiable structure.

Cys^{351} is near the C-terminal end of the helix α-H in a solvent-exposed position and Cys^{363} is in the predicted β-strand. A property that is consistent with this model is that residues that link closely spaced cysteines involved in a disulphide bridge, often have high β-turn potential [19]. In accordance with this, high β-turn potential exists approximately equidistant from the two cysteine residues of the model. Secondary structure prediction is of limited value without an assessment of resultant tertiary interactions. Therefore the main reason for utilizing a molecular mechanics approach was to establish if steric interactions or conformational energy precluded the formation of a disulphide bond between Cys^{351} on α-H and Cys^{363} on the antiparallel β-strand. The precise nature of the turn is unimportant but the ability to form a reverse turn is a necessary part of the model. Hence, a β-II type turn (where $i+1=Thr^{356}$ and $i+2=Gly^{357}$) and a conformationally idealized 8-residue β-strand were constructed immediately after helix α-H. This positioned the β-strand antiparallel to α-D and α-H. It could be seen that the two cysteine sulphurs could be brought into bonding distance by docking the β-strand against α-H and α-D. This could be achieved by rotating about the ψ-angle of Gly^{359}. The β-strand which was initially in a solvent-exposed position (see Fig. 4) packed against α-H and α-D on a solvent-exposed face of the molecule.

Energy minimization

Energy minimization was performed using DISCOVER v.2.4 with the all atom Consistent Valence Force Field (CVFF) [20]. A neutral charge group parameterization was used in order to compensate for the lack of solvent in the simulation. The neighbour list was updated every 30 steps using a nonbonded cutoff distance of 10 Å, with a switching function applied between 8 and 10 Å. Residues in the α-H and the C-terminal construct were allowed full conformational freedom, as were residues in the adjacent face of α-D, otherwise all residues were fixed to their mod-

elled coordinates. A progressive distance constraint was placed between the two cysteine sulphur atoms as follows:

Distance forcing between the cysteine sulphur atoms was implemented slowly, in order to avoid large strains in the system. The target distance was progressively decreased using a harmonic restraint of the form $K(r-r_o)^2$, K being the force constant in kcal mol^{-1} Å^{-2}, r_o and r being the desired equilibrium and current atom-atom distances in Ångstroms, respectively. The force constant K was 15.0 kcal mol^{-1} and the initial atom–atom distance was 16.0 Å with a final target distance of 3.0 Å, r_o was decreased in steps of 0.5 Å with each step involving 500 steps of conjugate gradients minimization. Four different models were generated for comparison.

(i) Model E_1: The initial construct prior to distance forcing was minimized.
(ii) Model E_2: Distance forcing was carried out as described above, followed by unconstrained minimization.
(iii) Model E_3: Distance forcing was carried out as described above, followed by constrained minimization with a distance constraint between the two sulphur atoms of 3.0 Å and a force constant K of 15.0 kcal mol^{-1}.
(iv) Model E_4: Distance forcing was carried out as described above, followed by a disulphide bond being introduced between the two sulphur atoms and unconstrained minimization.

All the above were minimized to a maximum r.m.s. derivative of 0.002 kcal mol^{-1}Å^{-2} before comparison.

The remaining 10 residues of the C-terminus extension had not been considered thus far. However, it was evident from the model in which a disulphide bond is formed (model E_4), that these residues can dock into the active site of the enzyme, since they point towards the active site. The docking could be achieved by using sterically allowed backbone $\varphi\psi$-angles. Preferred side-chain torsion angles (χ_1) for a given backbone $\varphi\psi$-angle were taken from the analysis of MacGregor et al. [21]. Manual orientation of other side-chain torsion angles (χ_2,χ_3, etc.) ensured that there were no unacceptable steric clashes with the remaining protein model, whilst standard geometries were maintained as far as was possible. This model was then minimized with the C-terminus residues Lys^{350} to Val^{374} unconstrained during the calculation. Residues within 6.0 Å of these residues were fixed at their initial positions, except for side chains in contact with the C-terminus which were constrained to their initial positions with a force constant of 1000.0 kcal mol^{-1}Å^{-2}.

The construction of the N-terminal polypeptide extension

In order to access the importance of the N-terminal polypeptide extension in the regulatory mechanism of the enzyme, a tentative model for residues 13–29 has been constructed. The backbone conformation was built to correspond with that of mammalian lactate dehydrogenases [16], based on the following criteria:

(i) Mammalian lactate dehydrogenases are unique in having a 16-residue N-terminal extension with respect to other known NAD-linked bacterial lactate dehydrogenase and malate dehydrogenase sequences. In these enzymes this peptide extension constitutes an 'arm' which projects away from the main body of the subunit making contacts with both the dimeric and tetrameric related subunits [16]. The first 16 amino acids of the plant NADP-malate dehydrogenase N-terminal extension have low sequence identity with their mammalian counterparts, however, the presence of charged and hydrophilic amino acids in the first 8 amino acids of the polypeptide extension (which spans between dimerically related subunits) is

common to both the plant and mammalian sequences. Hydropathy plots and surface probabilities have significant correlation in the N-terminal extended region (data not shown).

(ii) The only sizeable change in the surface loops between maize and pig comes at a point at which the N-terminal polypeptide makes contact with the dimeric-related subunit. The larger loop of pig malate dehydrogenase sterically interferes with the N-terminal 'arm' when pig muscle lactate dehydrogenase and pig cytosol malate dehydrogenase are superimposed. The smaller loop of *Z. mays* has the same length as that of pig muscle lactate dehydrogenase and is sterically compatible with the conformation of the N-terminal 'arm'.

However, given the basic constraints of the rest of the protein in its dimeric form, a more direct route of the N-terminal polypeptide extension towards the active site can be built. Therefore a greater extent of interaction between the C- and N-terminal extensions could be envisaged.

RESULTS AND DISCUSSION

A ribbon representation of one subunit of the maize model is presented in Fig. 1. The N-terminal extension in this figure is of the dimeric-related subunit (the rest of this subunit is not shown). In accordance with other model-building studies, the majority of charged side chains are solvent exposed as defined by the solvent-accessible surface [22], this was particularly evident where an uncharged side chain in pig cytosol MDH was substituted for a charged side chain in the maize

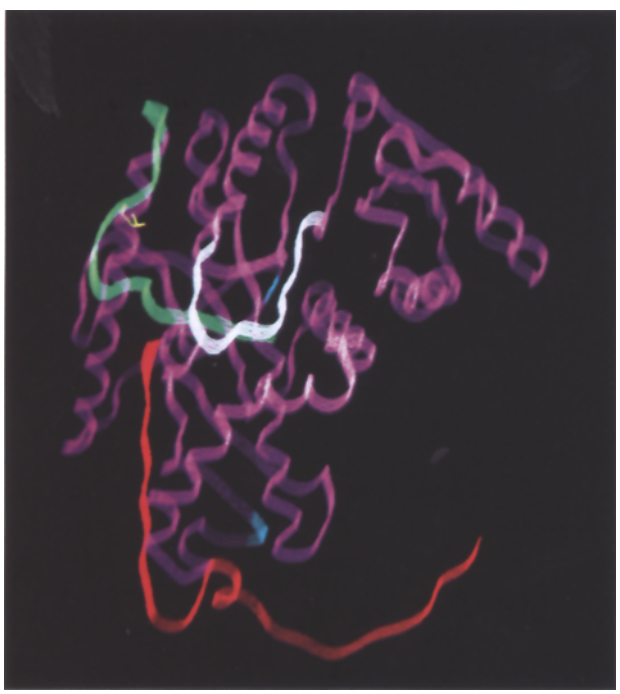


Fig. 1. Ribbon representation of one subunit of *Z. mays* with the N-terminal extension of the dimeric-related subunit in red. Magenta: structurally conserved backbone; white: active site mobile loop; green: C-terminal extension; yellow: C-terminal disulphide bridge; cyan: variable-length surface loops.

MDH and vice versa. There were 55 such charge changes, and only in two cases were the charge groups not partially accessible to solvent. One of these residues, Asp^{87}, is involved in an internal ion-pair with Arg^{254} and Arg^{190} in the pig cytosol structure and is also an aspartate in the *S. vulgare* sequence. However, residue 87 is an alanine in the maize sequence. Given that the aspartate can hydrogen bond with two arginine residues, both of which are conserved in all 3 sequences, and one of which is the completely conserved active-site residue Arg^{190} (which binds the substrate C1 carboxyl group in L-lactate and L-malate dehydrogenases) it would strongly suggest that Ala^{87} is a sequencing error. Similarly the other solvent-inaccessible side chain Asp^{152} in the maize sequence is also probably a sequencing error. This residue is a valine in the pig cytosol and *S. vulgare* sequence, and is part of the hydrophobic core of the protein, the side chain being in Van der Waals contact with the side chains of 5 other hydrophobic residues (Ala^{112}, Leu^{144}, Ala^{148}, Val^{154} and Ile^{175}) which are conserved in all 3 sequences. Apart from the steric incompatibility of the aspartate side chain, the environment seems an unlikely location for a charged residue.

Of the 8 cysteines in the maize structure, excluding the 4 cysteines in the N- and C-termini there is no indication from spatial considerations that any of the remaining 4 cysteines could participate in a disulphide bond (in either the dimer or generated tetramer) without sizeable secondary structure perturbation. Cys^{168} and Cys^{314} are in buried hydrophobic environments, whilst the side chain of Cys^{193} is partially solvent accessible as defined by its Connolly surface; that of Cys^{161} is even less so. Cys^{193} is near the dimer interface on α-2F. Cys^{161} projects into the active site on a loop connecting β-E and α-1F and is largely buried. This model agrees with the findings of Scheibe et al. [8] that there are two solvent-accessible thiols per subunit in the inactive oxidised form of the enzyme, both of which have relatively low reactivity. The experimental observations combined with the results of modelling imply that all other cysteines are inaccessible to solvent in this state.

Active site

Comparison of the sequence that makes up the active-site pocket of the maize and pig enzymes reveals that within a 5.0 Å sphere about the Nε of the active site His^{215} all but one of the residues are identical. The important catalytic residue Asp^{187}, which hydrogen bonds with His^{215} the proton donor/acceptor in catalysis, and the substrate-binding residue Arg^{190}, which forms an ion-pair interaction with the C1 carboxyl group of the substrate, are all invariant features of all known L-lactate and L-malate dehydrogenase enzymes. The residue which differs is Ala^{161}-Cys, the residue names are of the pig and maize structures, respectively. Cys^{161} is conserved in the 3 known plant NADP-malate dehydrogenases, and might thus have important consequences for the activity of the enzyme. It is adjacent to the catalytic His^{215} and the coenzyme nicotinamide ring whose nearest neighbour contacts with the Sγ of Cys^{161} are 3.6 Å and 3.2 Å, respectively. With His^{215} and Asp^{187} the cysteine could form an active-site triad reminiscent of that of sulphydryl proteases such as papain (see Fig. 2). The importance of this type of ion-pair interaction in papain is well understood, the cysteine being responsible for an increase of 4.8 units in the histidine pK_a [23]. By analogy it is predicted to have the same effect in plant NADP-malate dehydrogenases. The location of the cysteine as an integral component of the active site is consistent with the fact that thiol-selective reagents almost completely inhibit maize activity [13].

In NAD-linked malate and lactate dehydrogenases the binding of ketoacid demands that the His^{215} is protonated, and binding of the hydroxyacid requires this group must be unprotonated

[24]. In fact this is the only amino acid pK_a that is reflected in the pH dependence of substrate binding. It is noteworthy then that the pH dependence for maize and pea oxaloacetate activity is approximately bell-shaped in character, showing an optimum at pH 8.5 (at nonsaturating substrate concentrations), and only in very basic conditions (pH > 10) is significant malate activity observed. By direct analogy with NAD-malate dehydrogenases, this indicates a His^{215} pK_a of approximately 9 which is an increment of 2.1 units from that of the structurally similar pig cytosol malate dehydrogenase. A simple interpretation of the pH profile of activity might reflect a dependence on the protonation state of two amino acids, these being Cys^{161} and His^{215}. Furthermore enzyme affinity for $NADP^+$ would inevitably be increased by the ionization of the cysteine sulphur, some 3.2 Å distant from the positively charged nicotinamide ring.

Coenzyme binding site

Many features of the characteristic dehydrogenase βαβαβ coenzyme binding motif (or Rossmann fold) remain constant between the pig and maize structures, including the β-sheet region β-A which traverses the underside of the adenine-ribose diphosphate moiety (see Fig. 3B). One important change is located in the adjacent face of helix α-B which makes contact with the underside of β-A via a sharp turn connecting the two secondary structural elements. The change of Ala^{45} in porcine malate dehydrogenase to that of the sterically larger serine in *Z. mays* must have consequences for the local conformation of β-A. In fact serine could not be accommodated in the modelled structure with the pig malate dehydrogenase backbone conformation, without unacceptable steric clashes. This position has previously been implicated in discrimination between NAD and NADP, via site-directed mutagenesis in the coenzyme binding domain of glutathione reductase [25]. This study revealed that replacement of alanine by the sterically smaller glycine discriminated against NADP and for NAD, presumably by its effect on the local conformation of β-A which in turn interacts with the adenosine-ribose diphosphate moiety of the coenzyme.

The binding site for the 2′-phosphate group in *Z. mays* has 3 important charge changes with respect to porcine malate dehydrogenase, in addition to two other side chains that can hydrogen bond to the 2′-phosphate (see Figs. 3A, B).

(i) Lys^{118} is located in β-D and is in close proximity to the 2′-phosphate and the diphosphate groups of NADP (modelled using the NAD coordinates) and may provide electrostatic stabilization of negative charge. This residue is a methionine in the porcine cytosol structure.

(ii) Most immediately evident is the substitution Pro^{73} to Arg^{73} on a turn connecting β-B and α-C. The guanidinium group can clearly make close contact with the ribose 2′-phosphate.

(iii) The maize structure has a glycine at position 70 whilst pig malate dehydrogenase and indeed all NAD-linked malate dehydrogenase and lactate dehydrogenases have an aspartate at this position which hydrogen bonds to the 2′- and 3′-hydroxyl groups of the adenine-ribose. The proximity of a negatively charged residue at this position would clearly destabilize 2′-phosphate binding in NADP. This hypothesis has recently been confirmed by site-directed mutagenesis [26].

Ser^{71} on the turn connecting β-B and α-C and Ser^{74} on α-C were modelled into the side-chain conformations of the sterically larger residues isoleucine and methionine, respectively. Both have the potential to form hydrogen bonds with separate 2′-phosphate oxygens.

Another interesting feature of the maize model is the presence of Arg^{264} adjacent to the coen-

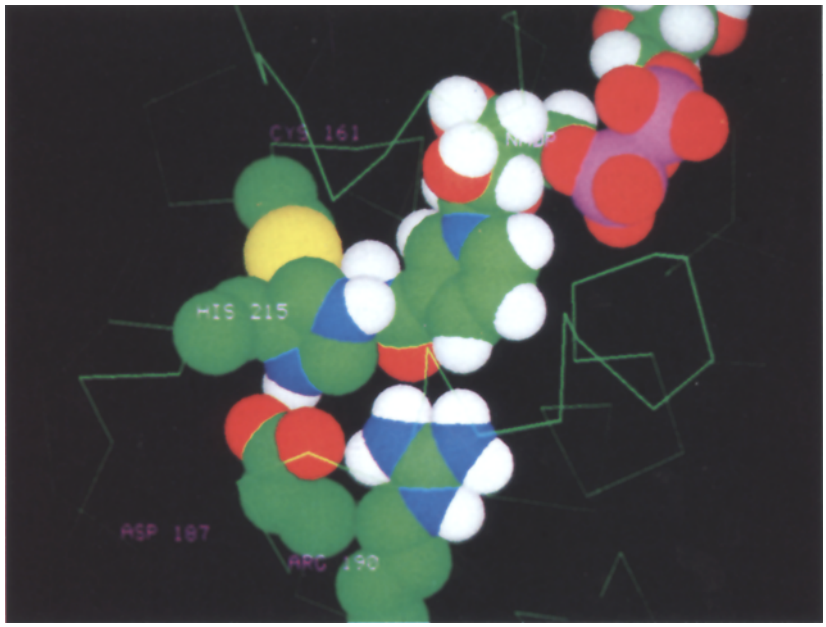


Fig. 2. A CPK representation of the *Z. mays* active-site residues, with a vector representation of the α-carbon backbone in green. Atom colors are green: carbon; red: oxygen; blue: nitrogen; magenta: phosphorous; yellow: sulphur and white: polar hydrogens.

zyme diphosphate moiety. The NAD-linked enzymes have no specific charge interaction with this group; it only seems evident with their NADP-linked counterparts.

Hence it would appear that selectivity of NADP-specific dehydrogenases is achieved by adversely affecting the adenine-ribose diphosphate interaction in order to weaken NAD binding, in addition to providing basic residues for a 2′-phosphate binding site.

C-terminal cystine loop

Energy minimization carried out on the helix-turn β-strand constructs are of significance only on a comparative basis. The difference in the energies and structures of the models after progressive distance constraint minimization are compared with the minimized starting structure (model E_1) in Table 2 and Fig. 4, respectively.

The model in which the sulphur-sulphur distance was constrained to 3.0 Å throughout energy minimization (E_3 in Table 2) has slightly less favourable coulombic and van der Waals interactions than model E_2 (in which distance forcing to 3.0 Å was followed by unconstrained minimization). However, this difference is primarily due to the distance constraint forcing the nonbonded thiol groups together in model E_3. Analysis of individual residue energies in the two systems shows an increase in angle (2.5 kcal) and van der Waals (1.3 kcal) energies of the two cysteine residues (Cys^{351} and Cys^{363}) of model E_3 as compared to E_2. This accounts for the difference in their total energy (3.4 kcal). Model E_4 (in which a disulphide bond is present between Cys^{351} and Cys^{363}) is included in the analysis even though it is not chemically identical to the other models.

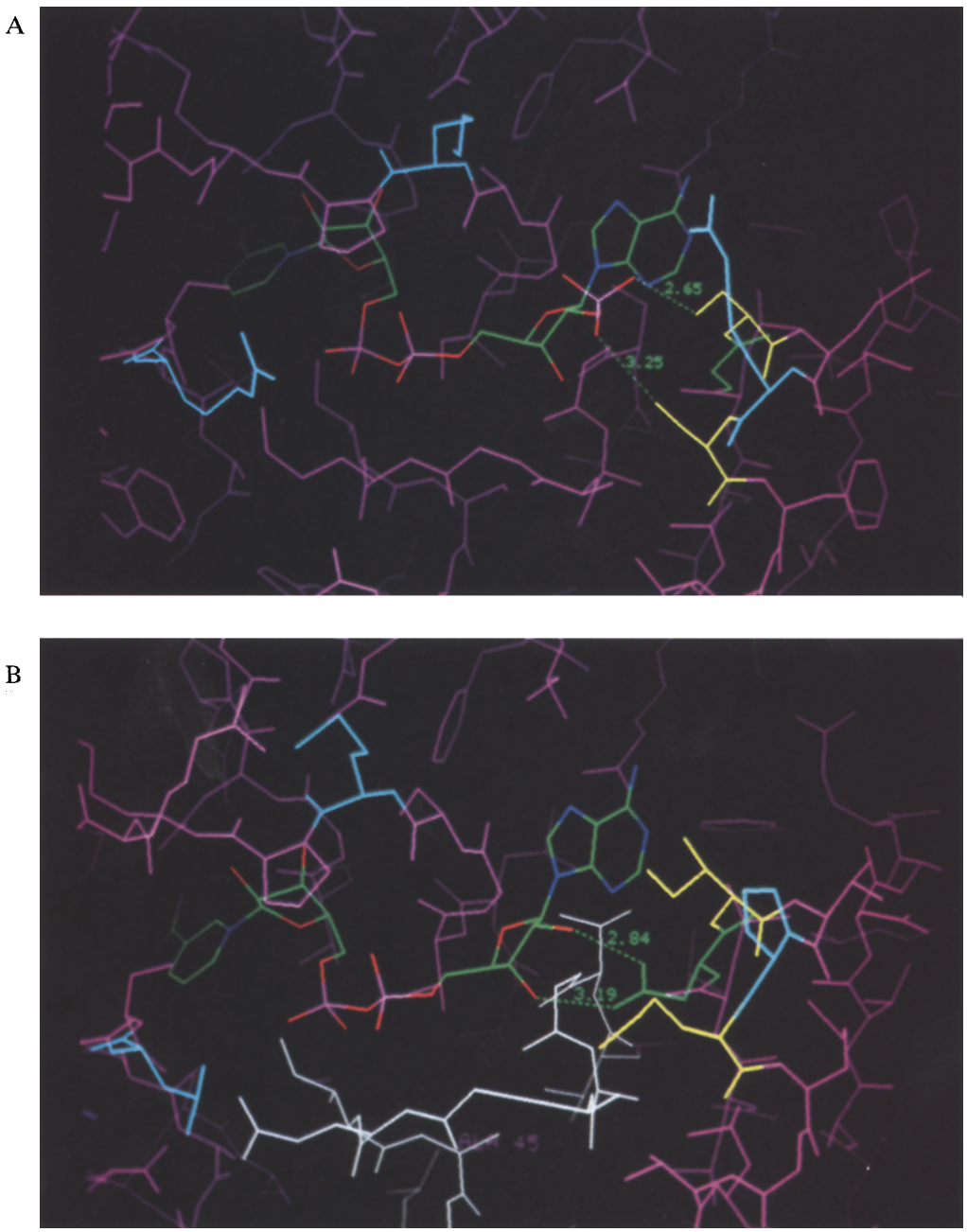


Fig. 3. (A) View of *Z. mays* NADPH binding site (the 2′-phosphate was modeled on the 2′-hydroxyl group of the adenine ribose ring). Arg^{73}, Lys^{118} and Arg^{264} are coloured cyan. Ser^{71} and Ser^{74} are coloured yellow. Gly^{70} is coloured green. The NADP is coloured according to atom type as in Fig. 2. (B) View of porcine cytosol NAD binding site. Pro^{73}, Met^{118} and Leu^{264} are coloured cyan. Ile^{71} and Met^{74} are coloured yellow. Asp^{70} is coloured green. β-B and α-C are coloured white with Ala^{45} labelled. NAD is coloured according to atom type.

TABLE 2
DIFFERENCES BETWEEN THE MINIMIZED ENERGY OF THE STARTING STRUCTURE E_1 AND THE ENERGIES OF STRUCTURES MINIMIZED AFTER DISTANCE FORCING TO 3.0 Å[a]

	Valence energy[b]				Nonbonded energy[b]		Total
	Bond	Angle	Torsion	X-term	VdW	Coulomb	
(E_2-E_1)	−0.92	−1.68	−3.93	−0.07	−14.23	4.49	−16.34
(E_3-E_1)	−1.09	−5.35	−1.09	0.69	−12.51	6.39	−12.96
(E_4-E_1)	0.17	−3.29	7.56	1.39	−8.39	−1.63	−4.24

[a] E_2 represents the energy of the minimized structure with the distance constraint removed before minimization, E_3 that in which the distance constraint was maintained at 3.0 Å and E_4 that in which a disulphide bond was introduced between the two cysteine residues.
[b] All energies in units of kcal mol^{-1}.

The overall nonbonded energy is more favourable than in other models, due to a more favourable coulombic energy component (of which Cys^{351} and Cys^{363} contribute 2.1 kcal when compared with model E_1), however, there is a large increase in torsion energy when compared with other models. Of this increase 2.3 kcal can be attributed to residues Cys^{351} and Cys^{363} which form the disulphide (when compared with model E_1), the remainder being the result of slight increases in torsion energy in all residues of the cystine loop. However, the total energy is still more favourable than that of the minimized starting structure E_1, due to favourable nonbonded energy compensating for the increase in torsion energy. Hence there would appear to be no energetic reason why Cys^{351} and Cys^{363} should not form a disulphide bond.

Most crystallographically determined disulphide bonds adhere to relatively strict stereochemical criteria [27]. Hence it is worth noting that the conformation of the disulphide in model E_4 is stereochemically ideal. The side-chain torsion angles (χ_1) of Cys^{351} and Cys^{363} are *trans* and *gauche*$^-$, respectively, and bond angles differ from ideality by no more than 2°. It has been noted that the χ_{ss} dihedral angle (Cβ1-Sγ1-Sγ2-Cβ2) has a very strong preference for the values ±90° [28], hence the observed value for model E_4 of −91.2° can be considered ideal.

In conclusion then, there would appear to be no energetic or steric factors which preclude the formation of the C-terminal cystine loop. More detailed examination of the tertiary structural interactions in this region explain why only the plant sequence for helix α-D could be compatible with a cystine loop in the C-terminal region. The outer face of α-D is solvent exposed in the structures of the NAD-linked lactate and malate dehydrogenases. Residues of α-D in the maize enzyme model are in contact with the β-strand of the cystine loop and have side chains which are small and hydrophobic (the residues are identical in all 3 plant malate dehydrogenases at this point), replacement of these side chains by the larger polar side chains of NAD-lactate/malate dehydrogenases would result in steric clashes between α-D and the modelled β-strand, particularly at positions 127 and 128 which are both alanines in the plant enzymes (see Table 1). Furthermore Fig. 5 illustrates that the contact region between the β-strand and α-D is essentially hydrophobic in nature forming a tightly packed interface. This includes van der Waals contacts of hydrophobic residues between α-D and the C-terminus (α-H and the constructed β-strand), of Leu^{130} (α-D) with Cys^{351} (α-H), Cys^{363} and Phe^{362} (β-strand), Ile^{132} (α-D) with Phe^{362}, and Ala^{127} (α-D) with

Leu^{365} (β-strand). In addition, there is a strong hydrogen bond (2.65 Å) between Thr^{356} (α-H) and Gln^{135} (α-D) both of which are conserved residues in the 3 plant sequences. Furthermore there are hydrophobic contacts between α-H and the β-strand, Val^{352} and Leu^{355} (α-H) are in contact with Phe^{362} (β-strand), together with Cys^{351} and Cys^{363}, these residues form the inner face of the C-terminal loop in contact with α-D.

These interactions might have consequences for the conformational change induced on substrate binding, in which the active-site loop and α-D move relative to α-H. The C-terminal helix α-H of NAD-malate/lactate dehydrogenase makes important interactions with residues that form part of the active-site vacuole. In particular, β-G and β-H between which the catalytic His^{215} residue resides, and with the active-site loop which carries Arg^{126}, the mobile arginine essential for catalysis [29,30]. Following substrate binding, helix α-H is known to undergo a sizeable conformational change which is coupled with active-site loop closure and a shift in α-D. This conformational change is a necessary requirement for catalysis, involving removal of bulk solvent from the active site and bringing the catalytic Arg^{126} into a position whereby it polarizes the substrate carbonyl. The presence of a cystine loop may by virtue of its contact with α-D and α-H, prevent this structural rearrangement.

However, a more direct form of inhibition may result due to the remainder of the C-terminal peptide docking into the active-site cleft. A consequence of the disulphide bridge is that the C-terminus of the β-strand points towards the active site. Since the remaining peptide is of just the right length, it could be docked into the active site of the enzyme in the presence of NADP (see Materials and Methods section) whilst retaining reasonable stereochemical criteria for the backbone ($\varphi\psi$) and side chain (χ_i) torsion angles (see Fig. 6 for a Ramachandran plot of the C-terminus construct, residues Gly^{357} to Val^{375}, after partial optimization by energy minimization). The C-terminal carboxyl group of Val^{375} forms an ion-pair interaction with Arg^{190} (which normally binds the substrate C1 carboxyl). The carboxyl of the penultimate residue, a conserved glutamate (Glu^{374}) can also interact with Arg^{120}, which binds the substrate C3 carboxylate in a malate dehydrogenase synthesized on the lactate dehydrogenase framework [31] and which is present in all known malate dehydrogenase sequences (see Fig. 7). Hence the acidic C-terminus may mimic the acidic substrate both sterically and electrostatically via its double negative charge.

In this context, it is interesting to note Scheibe's results on the effect of low concentrations of different denaturants on the nonreductive activation of pea NADP-malate dehydrogenase, under conditions where no measurable structural perturbations are observed via fluorescence or circular dichroism [31]. Guanidine and arginine hydrochloride have a large stimulatory effect on activity, whilst ionic strength (NaCl) has only a minor effect and the nonionic denaturant, urea, has a very small effect compared with the above. These results are in agreement with our model, which predicts that ionic interactions are of primary importance in inhibition of enzyme activity, and that a cosolvent that effectively competes with the basic residues of the substrate binding site can activate the oxidised enzyme.

Also consistent with the model is the observation that the reductive activation of NADP-malate dehydrogenase is inhibited by over 95% in the presence of $NADP^+$, whilst other substrates (NADPH, malate and oxaloacetate) have no effect. Since NADPH reverses the inhibitory effect of $NADP^+$ via competition for the same site [11], we must assume this is the coenzyme binding site. Given that the regulatory disulphide bonds are distinct from the active site a specific interaction between the coenzyme nicotinamide ring and the protein would appear to be responsible for

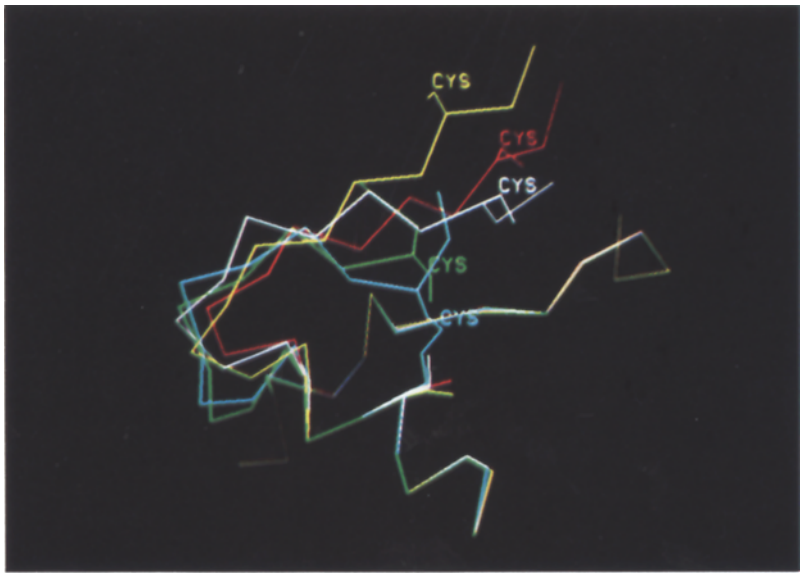


Fig. 4. The loop structures, superimposed using the fixed atoms in the system (see Materials and Methods section). Only the α-carbon backbone of α-D and α-H with the first 9 residues of the modeled C-terminal extension are displayed. The original starting structure is represented in yellow prior to energy minimization; model E_1: red; model E_2: white; model E_3: green; model E_4: cyan.

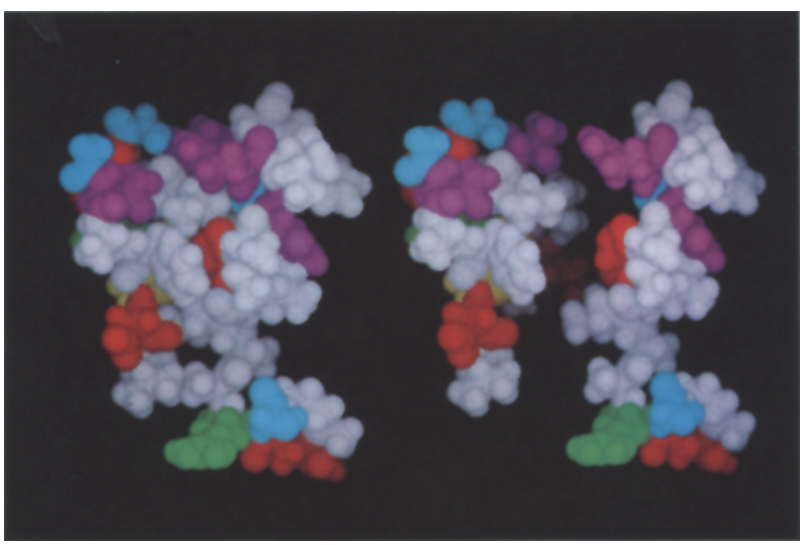


Fig. 5. The interface between α-D and the modeled β-strand of the cystine loop. Residues alanine, valine, isoleucine, leucine and phenylalanine are depicted in white, aspartate and glutamate in red, arginine and histidine in green, serine, threonine, glutamine and asparagine in magenta, glycine in cyan and cystine in yellow. The picture on the left displays the actual packing at the interface whilst on the right the cystine loop construct has been translated away from α-D (which is on the extreme right of the picture) in order to clearly define the interface.

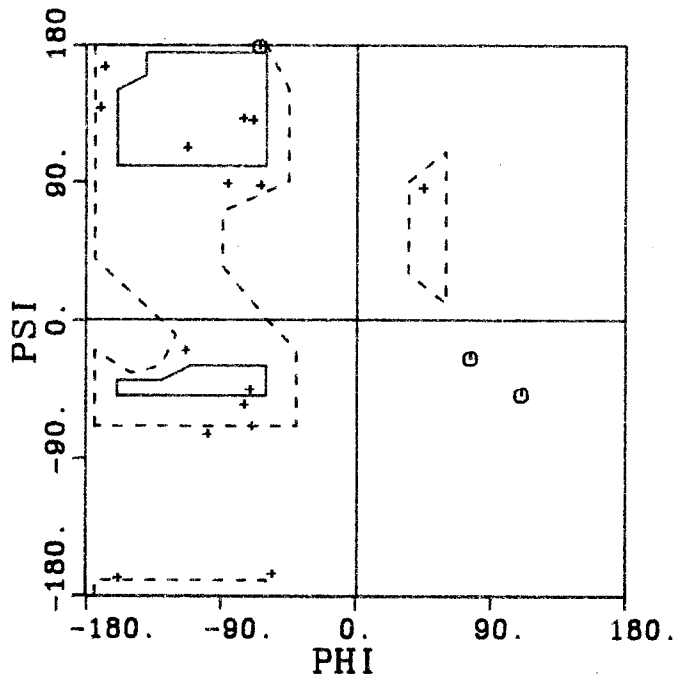


Fig. 6. Ramachandran plot of torsion angles ϕ and ψ, for the C-terminus residues Gly^{357} to Val^{375}. Glycine residues are represented as circles.

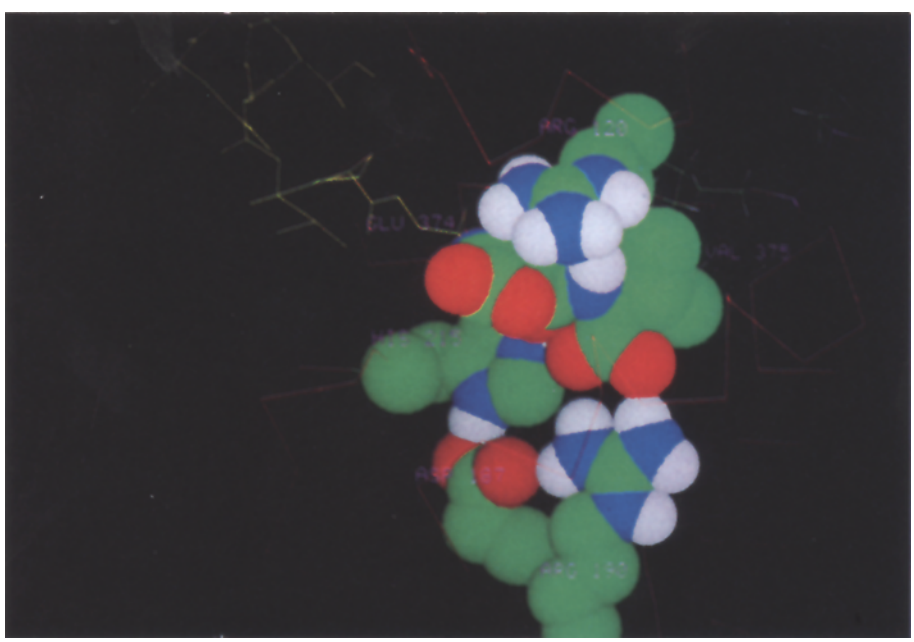


Fig. 7. CPK representation of the C-terminus residues Glu^{374} and Val^{375} docked into the active site of the *Z. mays* model. Residues are coloured by atom as in Fig. 2. The α-carbon backbone of *Z. mays* is in red. The heavy-atom representation of the C-terminal extension is in yellow.

the reactivity of a disulphide at a comparatively remote site on the enzyme. The C-terminus of the C-terminal polypeptide fulfills this role by making a stronger interaction with the nicotinamide ring of $NADP^+$ than that of NADPH in the active site of the enzyme, and so stabilizing the inactive conformation of the enzyme.

The N-terminal polypeptide conformation

The functional importance of the N-terminal 'arm' of mammalian lactate dehydrogenases is that of stabilizing the tetrameric state of the enzyme [16]. The proteolytic removal of 10–11 amino acids from the N-terminus of pig muscle lactate dehydrogenase leads to the formation of the active dimer [33,34], and the *B. stearothermophilus* lactate dehydrogenase which lacks an N-terminal 'arm' is easily able to dissociate to active dimers [35]. This might also be the functional role of the N-terminal extension in the plant NADP-malate dehydrogenases although a check against experiment is difficult since maize, pea and spinach NADP-malate dehydrogenase have been variously described as both dimers and tetramers under gel filtration conditions.

The conformation of the first 13 amino acids of the N-terminal extension cannot be predicted with any confidence as there is no experimental constraint or structure homology relationship to indicate its location. However, a relevant question to ask is, can it make direct contact with the active-site region or the C-terminal polypeptide extension? The predicted conformation of the last 17 amino acids of the *Z. mays* N-terminal extension (residues 13–29) could allow the first 13 amino acids to project away from the dimer-dimer interface antiparallel to helix α-H and might interact directly with the C-terminal extension if the N-terminal had a fully extended β-strand conformation. However, it is too short to interact directly with the active site, if the conformation of the last 17 amino acids of the N-terminal extension is analogous to that of mammalian lactate dehydrogenases. This construct did not give rise to a disulphide bridge between the known disulphide-forming residues Cys^{10} and Cys^{15} of the N-terminus, and therefore must be viewed with some caution.

Experimental results on the number of disulphide bonds reduced in the activation of the enzyme vary between one [8,10], two [36] and three [37], the model suggests there are no more than two disulphide bonds. However, from our results it cannot be stated with any certainty whether one or both of the two disulphide bonds are critical in activation of the enzyme, since the conformation of the N-terminal polypeptide extension of the dimeric-related subunit may have consequences for that of the C-terminal extension and vice versa.

However, it is useful to interpret our results in terms of available experimental results. In particular, an 8-residue deletion at the C-terminus of the inactive, oxidised wild-type C3 NADP-malate dehydrogenase from pea chloroplasts restores activity to 50% of that of the active reduced wild-type enzyme, and there is a considerably reduced rate of the specific thioredoxin-mediated activation of the truncated enzyme. Furthermore, the rate of activation by the nonspecific mercaptan DTT increases in the truncated enzyme and would suggest the modification results in the loss of a thioredoxin binding site, but increases the solvent accessibility of a disulphide bond [38]. This situation could arise from the loss of a thioredoxin binding site at the C-terminus or equally likely through the loss of a thioredoxin binding site at the N-terminus, via modification of the interactions between the N- and the C-terminal extensions. Given the evidence of the results of Decottignies et al. [10] that the N-terminal disulphide between Cys^{10} and Cys^{15} is the regulatory site of the

enzyme, the latter of these hypotheses seems most likely. However, if this is the case then it seems certain that the N- and C-termini must interact with each other.

Experimental support for the model

The model provides an explanation for many of the unique experimental features of the plant NADP-malate dehydrogenases.

(1) The alkaline pH optimum for oxaloacetate activity, the high affinity of the enzyme for $NADP^+$ and the complete inhibition of enzyme activity by thiol-selective reagents can be explained by the unique occurrence of the ionizable cysteine residue at the enzyme active site.

(2) The enzymes' requirement for NADP (as opposed to NAD in other L-malate/lactate dehydrogenases) is consistent with the model. The specificity for NADP appears to be the result of a modification of the interaction of the protein with the adenine-ribose-diphosphate moiety of the coenzyme in addition to the presence of a ribose 2′-phosphate binding site.

(3) The presence of two reactive cysteine residues in the inactive oxidised enzyme, is consistent with the solvent accessibility of two thiol groups in the model (Cys^{161} and Cys^{193}). This implies that the two cysteine residues in the C-terminus are inaccessible to solvent in the oxidised enzyme.

(4) The appreciable activity of the oxidised enzyme in the presence of the denaturants guanidine and arginine hydrochloride can be explained by the displacement of the acidic C-terminal residues Glu^{374} and Val^{375} (carboxyl terminus) from the active site of the enzyme.

(5) An 8-residue deletion at the C-terminus of the inactive oxidised C3 NADP-malate dehydrogenase from pea chloroplasts restores activity to 50% of that of the active reduced wild-type enzyme. This observation is consistent with the C-terminus acting as an 'internal inhibitor'.

(6) The ability of bound $NADP^+$ to inhibit the activation of the inactive oxidised enzyme, coupled with the knowledge from the model that both N- or C-terminal disulphide bonds are remote from the nicotinamide binding site, indicates that communication between these sites must be facilitated by a differential response of the enzyme to the redox state of the nicotinamide ring. The presence of the bound C-terminus peptide in the active site provides a model whereby this could occur.

ACKNOWLEDGEMENTS

This work was supported by the SERC grants GR/F7398.4 and GR/F1955.1 (to J.J.H.) and by a studentship (to R.M.J.).

REFERENCES

1 Perrot-Rechenmann, C., Jacquot, J.-P., Gladal, P., Weedon, N.F., Cseke, C. and Buchanan, B.B., Plant Sci. Lett., 30 (1983) 219.

2 Droux, M., Miginiac-Maslow, M., Jacquot, J.-P., Gadal, P., Crawford, N.A., Kosower, N.S. and Buchanan, B.B., Arch. Biochem. Biophys., 256 (1987) 372.

3 Buchanan, B.B., Wolosink, R.A. and Shurmann, P., Trends Biochem. Sci., 4 (1979) 93.

4 Scheibe, R., Fickenscher, K. and Ashton, A.R., Biochim. Biophys. Acta, 870 (1986) 191.

5 Buchanan, B.B., Annu. Rev. Plant Physiol., 31 (1981) 341.

6 Metzler, M.C., Beverly, A., Rothermel, A. and Nelson, T., Plant Mol. Biol., 12 (1989) 713.
7 Luchetta, P., Cretin, C. and Gadal, P., Gene, 89 (1990) 171.
8 Scheibe, R., Kampfenkel, K., Wessels, R. and Tripier, D., Biochim. Biophys. Acta, 1076 (1991) 1.
9 Fickenscher, K., Scheibe, R. and Marcus, F., Eur. J. Biochem., 168 (1987) 653.
10 Decottignies, P., Schmitter, J.M., Miginiac-Maslow, M., Le Marechal, P., Jacquot, J.-P. and Gladal, P.J., Biol. Chem., 263 (1988) 11780.
11 Ashton, A.R. and Hatch, M.D., Arch. Biochem. Biophys., 227 (1983) 416.
12 Rebeille, F. and Hatch, M.D., Arch. Biochem. Biophys., 249 (1986) 164.
13 Ashton, A.R. and Hatch, M.D., Arch. Biochem. Biophys., 227 (1983) 406.
14 Birkoft, J.J., Rhodes, G. and Banaszak, L.J., Biochemistry, 28 (1989) 6065.
15 Wigley, D.B., Gamblin, S.J., Turkenburg, S.P., Dodson, E.J., Piontek, K., Muirhead, H. and Holbrook, J.J., J. Mol. Biol., 223 (1992) 317.
16 Adams, M.J., Hans, D.J., Jeffery, B.A., McPherson, A., Memall, H.L., Rossmann, M.G., Schevitz, R.W. and Wanacott, A.J., J. Mol. Biol., 41 (1969) 159.
17 Adams, M.J., Ford, G.C., Koekoek, R., Lentz, P.J., McPherson, A., Rossmann, M.G., Smiley, I.E., Schevitz, R.W. and Wonacott, A.J., Nature, 227 (1970) 1098.
18 Gribskov, M., Devereux, J. and Burgess, R.R., Nucleic Acids Res., 14 (1986) 227.
19 Thornton, J.M., J. Mol. Biol., 157 (1981) 209.
20 Hagler, A.T., J. Am. Chem. Soc., 101 (1979) 5122.
21 MacGregor, M.J., Islam, S.A. and Sternberg, J.E., J. Mol. Biol., 198 (1987) 295.
22 Connolly, M.L., Science, 221 (1983) 709.
23 Johnson, F.A., Lewis, S.D. and Shafter, J.A., Biochemistry, 20 (1981) 44.
24 Lodola, A., Parker, D.M. and Holbrook, J.J., Biochem. J., 173 (1978) 959.
25 Scrutton, N.S., Berry, A. and Perham, R.N., Nature, 343 (1990) 38.
26 Feeney, R., Clarke, A.R. and Holbrook, J.J., Biochem. Biophys. Res. Commun., 166 (1990) 667.
27 Sowdhamini, R., Srinivasan, N., Shoichet, B., Santi, D.V., Ramakrishnan, C. and Balaram, P., Prot. Eng., 3 (1989) 95.
28 Richardson, J., Adv. Prot. Chem., 34 (1981) 167.
29 White, J.L., Hackert, M.L., Buehner, M., Adams, M.J., Ford, G.C., Lentz, P.J., Smiley, I.E., Steindel, S.J. and Rossmann, M.G., J. Mol. Biol., 102 (1976) 759.
30 Clarke, A.R., Wigley, D.B., Chia, N.W., Barstow, D., Atkinson, A. and Holbrook, J.J., Nature, 328 (1987) 699.
31 Wilks, H.M., Hart, K.W., Feeney, R., Dunn, C.R., Muirhead, H., Chia, W.N., Barstow, D., Atkinson, A., Clarke, A.R. and Holbrook, J.J., Science, 242 (1988) 1541.
32 Scheibe, R., Rudolph, R., Reng, W. and Jaenicke, R., Eur. J. Biochem., 189 (1990) 581.
33 Girg, R., Rudolph, R. and Jaenicke, R., Eur. J. Biochem., 119 (1981) 301.
34 Opitz, U., Rudolph, R., Jaenicke, R., Erickson, L. and Neurath, H., Biochemistry, 26 (1981) 1399.
35 Clarke, A.R., Atkinson, T., Campbell, J.W. and Holbrook, J.J., Biochim. Biophys. Acta, 829 (1985) 387.
36 Jenkins, C.L.D., Anderson, L.E. and Hatch, M.D., Plant Sci., 45 (1986) 1.
37 Kagawa, K. and Bruno, P.L., Arch. Biochem. Biophys., 260 (1988) 674.
38 Fickenscher, K. and Scheibe, R., Arch. Biochem. Biophys., 260 (1988) 771.