Biochimica et Biophysica Acta, 1057 (1991) 157–185

ADONIS 000527289100093B

BBABIO 43358

Review

Enzymes depending on the pterin molybdenum cofactor: sequence families, spectroscopic properties of molybdenum and possible cofactor-binding domains

John C. Wootton [1,*], Rachel E. Nicolson [1], J. Mark Cock [1], Dawn E. Walters [1], Julian F. Burke [2], Wendy A. Doyle [2] and Robert C. Bray [2]

[1] *Department of Genetics, University of Leeds, Leeds (U.K.) and* [2] *Biochemistry Laboratory, University of Sussex, Brighton (U.K)*

(Received 17 July 1990)

Key words: Pterin molybdenum cofactor; Molybdenum enzyme; Sequence family

Contents

* Present address: National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD U.S.A.

Abbreviations: DMSO, dimethylsulphoxide; EXAFS, X-ray absorption extended fine structure.

Correspondence: R.C. Bray, School of Chemistry and Molecular Sciences, University of Sussex, Brighton, BN1 9QJ, U.K.

I. Introduction: Classification of the enzymes

I-A. Preface

Molybdenum enzymes play varied and important roles in organisms from primitive bacteria to higher plants and animals including man. Information on and understanding of these enzymes is increasing steadily. Individual enzymes catalyse steps in the biological nitrogen and sulphur cycles, making them of agricultural and environmental importance. Also, genetic deficiency related to human molybdenum enzymes is associated [1] with severe clinical abnormalities, making the enzymes of biomedical importance. Aspects of molybdenum enzymes have been reviewed in books that appeared, respectively, 10 [3], 5 [2] and 1 year ago [4], as well as in a recent article [5] by one of the authors. Therefore, a rather brief introduction is given here.

Molybdenum in enzymes is bound to the proteins via one or other of two small molecules: that are referred to as the iron-molybdenum cofactor and the pterin molybdenum cofactor. One enzyme only, nitrogenase, which converts dinitrogen to ammonia. contains the iron-molybdenum cofactor and thus falls outside the scope of this article. All other molybdenum enzymes that have been investigated contain the pterin molybdenum cofactor. The structure of this cofactor and the question of chemical differences between cofactor molecules from different molybdenum enzymes is discussed below.

A considerable number of molybdenum enzymes are known that depend on the pterin molybdenum cofactor. These may be classified according to source, biological role and enzymic properties, or according to the nature of redox-active constituents that are present in the enzymes in addition to molybdenum. Alternatively, they may be classified from spectroscopic and chemical data relating to molybdenum. or from amino acid sequence comparisons. So far, no X-ray crystallographic data are available for any of these enzymes.

In this article, amino acid sequence data, which have accumulated rapidly in the last few years, are emphasised. Discussion of this is prefaced by brief surveys of information from spectroscopic work, bearing on the nature of the molybdenum site in different enzymes, and of work on the structure of the pterin molybdenum cofactor, as well as by an introduction to sequence comparison methods. The comparisons presented permit new conclusions to be drawn concerning some of the molybdenum enzymes.

I-B. Brief survey of the enzymes

We consider, first, nitrate reductases. These may have either assimilatory or respiratory roles. Enzymes with the former function have been reviewed in a recent book [4] *. They occur in plants and in lower organisms such as *Chlorella vulgaris*, as soluble enzymes that catalyse the first step of nitrogen uptake, by reducing nitrate to nitrite. For the dissimilatory enzyme, which is generally membrane-bound and occurs in various microorganisms including *E. coli* and bacteria involved in denitrification, reduction of nitrate to nitrite is, anaerobically, the terminal step of the respiratory chain. Various formate dehydrogenases occur in microorganisms and these, like the nitrate reductases, may

* Two further reviews [4a, 4b] on assimilatory nitrate reductase appeared shortly after this article was submitted.

have either an assimilatory role, in this case in carbon uptake, or a dissimilatory one in energy metabolism. A particular distinction of *E. coli* formate dehydrogenase, is that it contains a selenocysteine residue whose co-translational incorporation into the enzyme has been the subject of considerable research (e.g., Ref. 6). Sulphoxides and amine oxides. in particular trimethylamine *N*-oxide and dimethylsulphoxide (DMSO), are widespread in the environment and reduction of these compounds is catalysed by molybdenum enzymes. Thus, DMSO can act as terminal electron acceptor for anaerobic growth of certain bacteria, including *E. coli* [7]. Among the enzymes involved in these processes, the DMSO reductase of *Rhodobacter capsulatus* and related organisms, a soluble periplasmic enzyme, is currently attracting attention [8–11] in view of its unusual simplicity, it being the only molybdenum enzyme so far described that contains molybdenum accompanied by no other redox-active components. For other molybdenum enzymes, iron-sulphur, flavin or haem centres are always present and are involved in intramolecular electron transfer to or from the molybdenum centre, where reaction of the primary substrate takes place. Strong light absorption by these centres masks the weaker absorption by the bound molybdenum cofactor, complicating its study in these enzymes by optical methods, whereas such a limitation does not apply to this DMSO reductase.

In animals, molybdenum enzymes comprise liver sulphite oxidase and members of the group known as molybdenum-containing hydroxylases *. The former catalyses the terminal step of catabolism of sulphur containing compounds, prior to excretion as sulphate (humans are said to excrete 1 g of SO_4^{2-} per day). The molybdenum-containing hydroxylases include xanthine oxidases and dehydrogenases as well as aldehyde oxidases. These enzymes are widely distributed in organisms that range from the primitive sulphate-reducing bacterium, *Desulfovibrio gigas* [12], through, for example, the fruit fly, *Drosophila*, to higher animals, including man [13]. The most extensively studied of them is xanthine oxidase from bovine milk. The amino acid sequence of *Drosophila melanogaster* xanthine dehydrogenase has been determined [14,15] by sequencing of the *rosy* gene. Surprisingly, the role of these enzymes is not always clear. Molybdenum-containing hydroxylases have exceptionally low specificities [16] making it difficult to determine, particularly for higher organisms, on which of their many possible substrates the enzymes actually act under normal in vivo conditions. Some lower organisms, including certain fungi [17,18] and bacteria [19–21], can utilize hypoxanthine or xanthine as sole nitrogen, or nitrogen and carbon source, and for these, oxidation of these compounds to uric acid represents the first step of nitrogen assimilation. In *Drosophila*, eye colour is influenced by the action of xanthine dehydrogenase on pteridine eye pigments, a fact that has long been utilized by geneticists working on this organism. On the other hand in man, though the precise significance of some of the work is uncertain, xanthine oxidase attracts attention in relation to heart disease. It has been suggested to be involved in ischaemica/reperfusion injury, supposedly because of the enzyme's ability to reduce oxygen to the superoxide radical [22]. Very recently evidence has been presented [23] for antibodies in normal humans to xanthine oxidase, with these significantly elevated in myocardial infarction patients.

Constituents of the various enzymes in addition to the pterin molybdenum cofactor are as follows. Molybdenum-containing hydroxylases contain FAD and two different types of iron-sulphur centre (an exception is the enzyme from *D. gigas*, which lacks the flavin). Sulphite oxidase contains a *b*-type cytochrome, as do the assimilatory nitrate reductases, with these additionally containing FAD. Bacterial nitrate reductases contain iron-sulphur centres and sometimes a *b*-type cytochrome. Formate dehydrogenases have various compositions depending on the source.

I-C. Classification from spectroscopic and chemical properties of the molybdenum

As is discussed elsewhere [5,24], and is taken up again below, most of what is known about the local environment of molybdenum in enzymes has come from two spectroscopic methods, EPR and EXAFS. Both provide information on the ligand environment of molybdenum in the enzymes. Where available, results of such work permit simple categorization of the enzymes. Such a categorization, based primarily on molybdenum(V) EPR spectra, is considered below.

Detailed discussion of EPR signals [5,25,26] is outside the scope of this article. However, we note that EPR spectra of Mo(V) in enzymes are unusually revealing in comparison of those from other transition metals in metalloproteins. Each enzyme gives rise under appropriate conditions, not to one but to a number of different Mo(V) EPR signals, each one of which corresponds to a different structure round the molybdenum. The different signals are usually referred to by names *.

* The term molybdenum-containing hydroxylase, as used here, has recently been approved by the IUB Nomenclature Committee. However, there is some confusion in the literature, since some workers have used it to include sulphite oxidase and nitrate reductase also.

* The main signals from molybdenum-containing hydroxylases are called Very Rapid, Rapid, Slow and Inhibited, respectively, from sulphite oxidase they are Low-pH and High-pH, from respiratory nitrate reductases the same names are applied and for assimilatory nitrate reductases the names Signal A and Signal B are used.

TABLE I

Assignment of molybdenum centres of different enzymes into families, from Mo(V) EPR, EXAFS and chemical data

Assignment has been made primarily on the basis of the EPR data. Where comparison of the Mo(V) EPR spectra in the references cited clearly indicates that an enzyme belongs to a particular family, the word 'yes' is entered, a '?' indicates that some uncertainty remains. EXAFS data on sulphido (= S) and oxo (= O) ligands, only, of molybdenum in the oxidized enzymes are presented, with ligand numbers in parentheses. The last two columns refer to interconversion between functional and desulpho forms of molybdenum-containing hydroxylases, by replacement of a sulphido ligand by an oxo, and vice versa.

Family	Enzyme	Source	EPR	EXAFS	SCN^- liberation by CN^-	Reactivation by S^{2-}
Molybdenum-containing hydroxylases	Xanthine oxidase	Bovine milk	yes [5]	= S(1), = O(1) [27,28] [a]	yes [29]	yes [29,30]
	Xanthine dehydrogenase	Chicken or turkey liver	yes [31]	= S(1), = O(1) [32] [a]	yes [33]	yes [33]
	Xanthine dehydrogenese	*Veillonela alcalescens*	yes [34]			
	Xanthine dehydrogenase	*Clostridium acidiurici*	yes [35]			
	Xanthine dehydrogenase	*Drosophila melanogaster*				yes [36,37]
	Aldehyde oxidase	Rabbit liver	yes [38,39,40]		yes [40,41]	yes [36]
	Mo-Fe protein (aldehyde oxidase)	*Desulfovibrio gigas*	yes [12]	[b]		
	Purine hydroxylase	*Aspergillus nidulans*	? [43]			
	Carbon monoxide oxidase	*Pseudomonas carboxydovorans*	? [44]		? [45]	
Animal sulphite oxidase	Sulphite oxidase	Chicken or bovine liver	yes [46,47]	= O(2) [48]		
Respiratory nitrate reductases (and related enzymes)	Nitrate reductase	*E. coli*	yes [49]	= O(1?) [50]		
	Nitrate reductase	*Pseudomonas aeruginosa*	yes [51]			
	Nitrate reductase	*Paracoccus denitrificans*	yes [52]			
	DMSO reductase	*E. coli*	? [52a]			
Assimilatory nitrate reductases	Nitrate reductase	*Chlorella vulgaris*	yes [53]	= O(2) [54]		
	Nitrate reductase	Spinach	yes [55,56]			
Formate dehydrogenase Type 1	Formate dehydrogenase	*Methanobacterium formicicum*	yes [57]		yes [57]	
Formate dehydrogenase Type 2	Formate dehydrogenase	*Pseudomonas aeruginosa*	yes [58]			
Formate dehydrogenase Type 3	Formate dehydrogenase	*Clostridium pasteurianum*	none [59]	= O(3?) [60]		

[a] EXAFS of the desulpho enzyme, with two oxo ligands, is also reported in the references cited.
[b] EXAFS of the desulpho enzyme, with two oxo ligands, is reported in Ref. 42.

As is indicated in Table I, the enzymes fall into one or other of four clearly defined families, or into a fifth and less well-defined family. Superficially, the EPR signals from all the enzymes are rather alike in that they have g_{av} values of about 1.96–1.98 * and in that most of them show splittings due to relatively strong hyperfine coupling of molybdenum to one or more exchangeable protons. These are believed to be present as -OH (or sometimes -SH) ligands of the molybdenum. However, closer inspection of the data (some of which have only recently become available) reveals the remarkable homogeneity of the EPR spectral forms and parameters ** when, and only when, any one signal is compared within a single family, as defined here. Thus, three respiratory nitrate reductases [49,51,52] show two types of molybdenum(V) signal that are virtually indistinguishable from one another. A similar conclusion applies to the two assimilatory nitrate reductases studied by EPR. Spectra of the spinach [55,56] and *Chlorella* [53] enzymes are essentially indistinguishable from one another, while being clearly distinct from those of the respiratory enzymes, as indeed they are from those of all the other families listed in Table I. The family for which the largest number of members has been studied in the greatest detail is that of the molybdenum-containing hydroxylases. Recent work, for example on the *Desulfovibrio gigas* enzyme [12] and on rabbit liver aldehyde oxidase (Ref. 40 and Turner, N.A. and Bray, R.C., unpublished work) re-emphasises the uniformity of the molybdenum(V) EPR signals within this family that was earlier established in relation to xanthine oxidases and dehydrogenases from bacterial as well as from mammalian and avian sources [5,31,34,35].

Uniformity of EPR signals within a family undoubtedly means uniformity of structure. Thus, it can confidently be stated that, within any one family (and with all the enzymes of the family in the same state), the atoms ligating molybdenum will be identical and the ligand distances and bond angles extremely similar to one another. The only qualification to be added is that molybdenum(V) EPR spectra tend to be relatively insensitive to the nature of a ligand atom occupying an axial coordination site, so that identity in this site may perhaps not always be fully guaranteed.

* g values measure the position in the EPR spectrum. The values quoted are in the normal range for Mo(V) compounds.

** Extraction of EPR parameters (g values, hyperfine couplings and sometimes angles of non-coincidence) from spectra can be a relatively complex process. If more than one signal-giving species is present in the same sample, their signals will tend to overlap, so that some form of spectral deconvolution will be needed for detailed interpretation. Parameters are usually obtained with the help of computer simulations. In Table I, firm assignment of an enzyme to a family has generally been made on the basis of very close similarity of at least two signals from the enzyme to the corresponding signals from other members of the family.

For molybdenum centres in enzymes, the catalytic and other properties (including the EPR spectra) depend [5] in no small measure on the presence of oxo groups (terminal = O ligands) and in some cases also of sulphido groups (terminal = S ligands) as ligands of the metal. The presence of these groupings, well known in simple complexes of the metal, has been established most directly in several of the enzymes by EXAFS spectroscopy and available data are summarized in Table I. The molybdenum-containing hydroxylases contain both an oxo ligand and a sulphido ligand. Unique investigations summarized in earlier reviews (Ref. 5; cf. Ref. 27), of transient species during enzyme turnover observed by EPR spectroscopy, using substitution with stable isotopes, have gone some way towards elucidating the roles of these groups in the catalytic reactions of these enzymes. Oxo groups are involved in oxygen transfer and sulphido ones apparently as proton acceptors. The sulphido ligand is relatively labile and may be replaced by a second oxo group by treating the enzymes with cyanide ions, so converting them to the inactive, desulpho forms, the cyanide being converted in the process to thiocyanate. It is sometimes possible to re-insert the sulphido ligand into the desulpho enzymes, simply by treatment with sulphide ions. Data on these points are summarized in Table I. Note that *Drosophila melanogaster* xanthine dehydrogenase is included in this family in Table I, exceptionally in the absence of any spectroscopic data, on the basis of reactivation by sulphide ions, this having been established [36,37] not only for the enzyme from wild-type flies but also for that from the *ma-l* mutant, which is genetically lacking in the ability to incorporate the sulphido ligand into the molybdenum centre of its xanthine dehydrogenase, which is thus inactive.

EXAFS data on the other families in Table I provide evidence for oxo ligands of molybdenum. Two such ligands are clearly established [48] in the sulphite oxidase molecule and probably also, from less detailed work [54], in the assimilatory nitrate reductases. On the other hand, there may be one oxo ligand only [50] in respiratory nitrate reductases with, apparently [60] and surprisingly, three in the one type of formate dehydrogenase examined by EXAFS. EPR data on formate dehydrogenases distinguish these enzymes from all the other enzyme families in Table I, but on presently available data they do not themselves form a single family. Note particularly that there is evidence [57], but not from EXAFS, suggesting a sulphido ligand in *Methanobacterium formicicum* formate dehydrogenase. If confirmed, this would be the only example outside the molybdenum-containing hydroxylases of an enzyme containing such a ligand *.

* Please see footnote * on next page.

Another important chemical property of molybdenum in the different enzymes is the redox potential of the metal. It is well established from model compound work [61] that this is strongly dependent on ligation of the metal. Available data on Mo(VI)/Mo(V) and Mo(V)/Mo(IV) potentials, from potentiometric titration measurements of the enzymes at pH 7 to 8, fit well with the categorisation into families as given in Table I. For many of the enzymes the two potentials are fairly close to one another, which has the effect that 100% conversion to the EPR-active Mo(V) state is never achieved **. Experimental differences mean that it is not always easy to compare redox potentials in the literature. In particular, changes of apparent pH on freezing are important [62,63] and furthermore, allowance has not always been made for the binding [46] of different anions to the enzymes. Within these limitations, the lowest reported potentials *** of around −400 mV [64] are for a formate dehydrogenase. Next come the molybdenum-containing hydroxylases, a number of which have been studied and all of which have potentials in the region of −350 mV [39,65–70]. The two assimilatory nitrate reductases investigated both have potentials around 0 mV [71–73]. For sulphite oxidase, though the two reported potentials of the metal [74] are unusually far apart (perhaps because of chloride effects) they are centred on a value comparable to that for the assimilatory nitrate reductases. By far the highest potential is shown by the one respiratory nitrate reductase that has been studied, that from *E. coli*, with a potential of about +200 mV [75].

II. Variability of the environment of molybdenum in different enzymes: roles of the cofactor and of the protein

II-A. The nature of the pterin molybdenum cofactor or cofactors

At present it is not clear whether the diverse properties of molybdenum in the different families of enzymes, as discussed above, depend solely on differences in the environment provided by the different proteins, or whether chemical differences in the cofactor molecules themselves are also involved. The possibility that such cofactor variants might exist has long been recognised [76–78] and strong indications [79] for them have recently been supplemented by certainty in one particular case [11]. Before considering these matters, which are central to understanding the enzymes at the chemical level, we take up the question of the nature of the molybdenum cofactor or cofactors. Earlier work has been reviewed [76–78]. A common molybdenum-containing cofactor for molybdenum enzymes was first postulated in 1964, but because of its extremely unstable nature when liberated from the enzymes and exposed to air, elucidation of its structure progressed slowly. Assay of the cofactor is almost universally done in terms of its ability to complement the apo-nitrate reductase in extracts of the *nit*-1 mutant of *Neurospora crassa*, to give nitrate reductase activity. In its original form, and as still often used, this assay is only semi-quantitative, though conditions for making it quantitative have been described [80].

Evidence that the cofactor contains a reduced pteridine derivative was first presented in 1980 [81]. Extensive study of the structure of molybdenum-free fluorescent products from its oxidative degradation, ultimately led Rajagopalan and co-workers [82] to conclude that one form of the cofactor, that from bovine xanthine oxidase and chicken sulphite oxidase, has the structure illustrated in Scheme Ia *. An additional form was recently identified [11] (Scheme Ib) as the cofactor in the enzyme dimethylsulphoxide reductase from *Rhodobacter sphaeroides*. This has a structure identical to that of Scheme Ia but with, as an additional component, a guanosine 5′-monophosphate residue attached to the phosphate group through a pyrophosphate linkage. Further confirmation of the basic structures of Scheme 1 and elucidation, for example, of details of the coordination of molybdenum might be desirable and could in principle be achieved by chemical synthesis (cf. Ref. 83, 84) or by spectroscopic studies of the free undegraded cofactor. No more than limited EPR work [85,86], which is considered further below, has been reported to date. However, the isolation has been described [87] under rigorously controlled anaerobic conditions and on a moderately large scale, starting from

* (p. 161) Formate dehydrogenase is unique amongst pterin molybdenum cofactor enzymes in that its reaction:

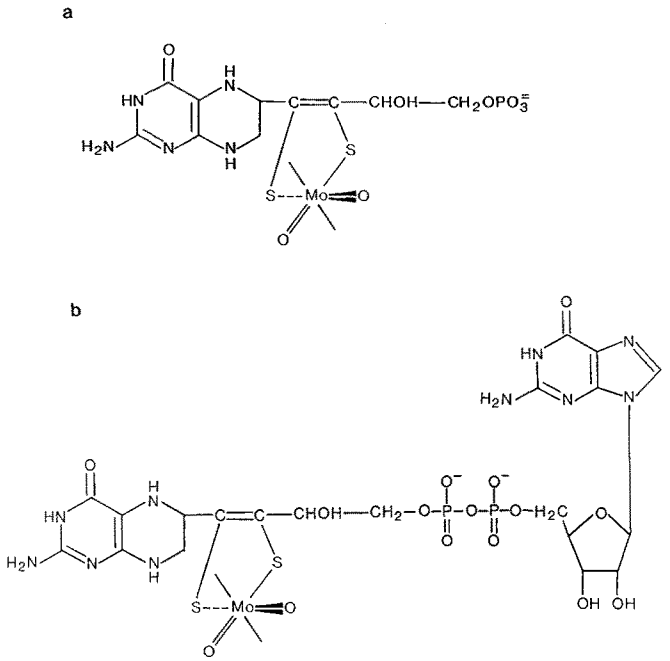
$$HCOO^- \rightarrow CO_2 + H^+ + 2e^-$$

formally, at least, involves only proton-, and not oxygen-transfer. The presence in its molecule of a sulphido ligand would thus be consistent with the experimentally demonstrated role [5] of this group as proton acceptor in the action of molybdenum-containing hydroxylases.

** The theoretical maximum conversion to Mo(V) for equal Mo(VI)/Mo(V) and Mo(V)/Mo(IV) potentials is 33%.

*** The catalytically inactive desulpho forms of molybdenum-containing hydroxylases, in which molybdenum bears two oxo ligands, also have low potentials.

* The name 'molybdopterin' is used by Rajagopalan and co-workers to denote the molybdenum-free form of the cofactor from sulphite oxidase and xanthine oxidase. Thus, according to these workers, the species denoted in Scheme 1a would be termed the molybdenum complex of molybdopterin. We find this nomenclature confusing and do not employ it here. Similarly, these workers term the cofactor from Rhodobacter sphaeroides dimethylsulphoxide reductase (cf. Scheme Ib), 'molybdopterin guanine dinucleotide'.



Scheme I. Proposed structures of the molybdenum cofactors from (a) chicken liver sulphite oxidase [82] and (b) *Rhodobacter sphaeroides* dimethylsulphoside reductase [11]. Structures have been modified so that the coordination of the metal [Mo(VI)] represents that, not in the free cofactor, but the oxidized enzyme. The nature of the unspecified ligands and the coordination geometry are discussed in the text. One = O ligand of molybdenum in structure (a) would be replaced by = S in xanthine oxidase and related enzymes.

purified milk xanthine oxidase, of the cofactor in an essentially undegraded, molybdenum-containing state. This could pave the way for further spectroscopic work on the undegraded molecule, particularly by NMR.

Relatively little is known of biosynthesis of the pterin cofactor, controlled in *E. coli* by several of the *chl* genes [88], though inherited deficiencies of the cofactor in humans are associated [1,89] with severe abnormalities. The search (e.g. Refs. 90, 91) for biological precursors of the cofactor led recently to the isolation and structural characterisation [92] of a product from oxidative degradation of such a molecule.

In principle, evidence for cofactor variants might come from quantitative enzymic assays of cofactor activity by the *nit*-1 assay, from spectroscopic or chromatographic studies, or from chemical analyses or degradation studies. As indicated above, special precautions are needed to make the *nit*-1 assay quantitative and not all molybdenum enzymes have had their cofactor tested carefully by this assay. Cofactor from chicken liver sulphite oxidase and xanthine oxidase are indistinguishable [80] in the *nit*-1 assay, as they are also by careful and detailed comparison [82] of their degradation products. Furthermore, important evidence has recently been presented [89], from studies of the nature of the defect in humans lacking the ability to synthesise the cofactor that late steps in the biosynthesis of the molecule are identical in man and in certain microorganisms. On the other hand, the cofactor from formate dehydrogenase of *Methanobacterium formicicum* has been reported [93] to have no activity in the *nit*-1 assay, and activity of that from two other proteins (the molybdenum-containing hydroxylase of *Desulfovibrio gigas* [12] and the demolybdo form of xanthine oxidase [94]) seems significantly low.

Another point of possible structural variation, in principle fairly easy to detect, concerns the phosphorus content of the cofactor. That from xanthine oxidase and sulphite oxidase bears a single phosphate group (cf. Scheme Ia [82,95]), while that from demolybdo-xanthine oxidase [94] appears [96] to be phosphate-free. In contrast, data [93] on the cofactor from *Methanobacterium formicicum* formate dehydrogenase, suggest that this molecule contains more than one phosphate group. Similarly, evidence has been presented by Meyer and co-workers [79], from ^{32}P incorporation studies, for the presence of two phosphate groups in the cofactor from carbon monoxide dehydrogenase from *Pseudomonas carboxydoflava*. This cofactor molecule was also reported [79] to be larger than that from xanthine oxidase, as indicated by gel filtration (M_r 730 and 330, respectively). These gel filtrations (see also Refs. 97, 98) were carried out on partly degraded cofactor molecules, on Sephadex G-15, in 4 M guanidium chloride at pH 1.2, and evidence was further presented that the cofactor from carbon monoxide dehydrogenase contains an extra non-pterin component, perhaps an aromatic grouping, associated with the second phosphate. On the evidence of M_r values of 690–750, determined [79,98] as above, for cofactor from eight other bacterial molybdenum enzymes (including nitrate reductase from *E. coli* and xanthine dehydrogenase from *Veillonella alcalescens*), it was concluded by Meyer and co-workers that the cofactor from bacteria * is different from that in mammals. Similarly high M_r values have recently been found [98a] in another laboratory for the cofactor from four further bacterial molybdenum enzymes.

Detailed degradation work on the cofactor from *Rhodobacter sphaeroides* has recently been reported [11]. The investigations described included conversion to the dicarboxamido derivative, in a procedure analogous to that used [82] in work on the cofactor from sulphite oxidase and xanthine oxidase. The product, containing two phosphate groups per molecule, was cleaved by treatment with a nucleotide pyrophosphatase to yield guanosine 5′-monophosphate and a carboxamido derivative of the cofactor, identical to that from the enzymes studied earlier. The analogy of the two pterin cofactor forms with FMN and FAD was stressed [11].

* The term 'bactopterin' was proposed [79] for the molybdenum-free form of this cofactor molecule.

A further way in which cofactor molecules could undoubtedly vary from enzyme to enzyme concerns the level of reduction of the pteridine derivative. The original form was isolated [82] with the pteridine in the tetrahydro form (cf. Scheme I), from sulphite oxidase and xanthine oxidase, but a higher oxidation state in the enzymes was not excluded. A problem, here, is that it is not necessarily valid to equate the oxidation state of the cofactor as isolated with that in the intact enzyme, since its reduction by -SH groups in the denatured protein, or by reducing agents added during the isolation procedures, cannot be excluded.

In summary, two forms of the pterin molybdenum cofactor are now well established but much work remains to be done on the distribution of these forms and the possibility that further forms may exist remains entirely open.

II-B. Possible ligation of molybdenum of the cofactor to amino acid side-chains of the protein

The pterin molybdenum cofactor is in all cases tightly bound to the various proteins by non-covalent interactions and is liberated from them only under denaturing conditions. In addition, it seems highly probable that amino acid side-chains of the protein provide one or more ligands of molybdenum in the cofactor. Most dioxomolybdenum(VI) complexes are six-coordinate, though five-coordinate complexes of the metal in this state are also known [99]. In keeping with the former, the metal in the structures of Scheme I is shown as six-coordinate *, with two thiolate ligands from the pterin cofactor, plus two oxo groups or one oxo and one sulphido group, and with two further unspecified ligands. In principle, spectroscopic data should be capable of yielding information on the nature and number of ligand atoms provided by the protein or, conceivably, by the cofactor molecule itself (cf. Ref. 100) **. Extensive spectroscopic data available on the various enzymes have so far failed to provide fully definitive information on this question. Positive evidence for sulphur or nitrogen ligands would be more readily obtainable than for oxygen. The first suggestions of sulphur ligation of molybdenum in xanthine oxidase came from early EPR work [101]. However, at that time the sulphido ligand and the pterin cofactor were unknown. As summarized below, there are indications of a cysteine ligand in respiratory nitrate reductase and sulphite oxidase, but no such indications for xanthine oxidase or dehydrogenase, and no positive indications of a nitrogen ligand in any of the enzymes.

As has already been discussed (cf. Table I), EXAFS has provided clear evidence for oxo and sulphido ligands of molybdenum in different enzymes. The technique also gives evidence for a number of thiolate ligands in all the enzymes; these might include, in addition to the two thiolate ligands of the structure of Scheme I, cysteine ligands from the protein. However, quantification of any such ligands is difficult. For all the enzymes, under all the conditions that have been tried, the number of thiolate ligands indicated by EXAFS [27,28,32, 48,50,54,102] ranges from about 2 to 4. Unfortunately, though, evaluation of ligand numbers by EXAFS is never very precise, a typical error being $\pm 25\%$ [103], and furthermore the technique cannot distinguish between ligands of similar atomic number. This latter point leads to particular problems in relation to sulphite oxidase and respiratory nitrate reductases. These enzymes undergo [46,49] a pH-dependent structural transition that involves the binding of an anion ligand such as chloride to molybdenum in the low-pH form. Thus, chloride ligands in the low-pH forms of the enzymes are liable to be counted as thiolates by EXAFS, as is the sulphido group when reduced to -SH, in reduced functional xanthine oxidase or dehydrogenase. When all such samples are excluded, the data suggest two thiolate ligands for xanthine oxidase [27,28], but three in sulphite oxidase [48] and respiratory nitrate reductase [50]. Thus, subtracting the two thiolate ligands assumed to come from the cofactor, we are left with the rather strong likelihood, according to the EXAFS data, of one thiolate ligand of molybdenum from a cysteine of the protein in sulphite oxidase and nitrate reductase, but with no support for (though probably not rigorous exclusion) of such a ligand in xanthine oxidase and dehydrogenase *.

EPR has so far failed to provide fully definitive evidence regarding ligation of amino acid residues to the metal in the enzymes. Among oxomolybdenum(V) complexes, it is well established that equatorially coordinated thiolate ligands, taking the place of oxygen or

* The precise coordination geometry in the enzymes is uncertain. That in Scheme I is intended to depict C_s symmetry. This symmetry accords with the EPR parameters of certain Mo(V) species derived from the cofactor [86] or xanthine oxidase [99a]. In particular, the coordination illustrated is that proposed [86] for a complex of the cofactor bearing two thiophenolate ligands.

** Phosphorus ENDOR has recently been used to show that in xanthine oxidase and sulphite oxidase the cofactor side-chain is in an extended conformation, thus making it clear that groups in it are not coordinated to molybdenum, (Howes, B.D., Bennett, B., Koppenhöfer, A., Lower, D.J. and Bray, R.C., unpublished work).

* The possibility of ligation of the sulphur of a methionine residue to the metal may also be considered. In the already complicated coordination sphere of molybdenum in xanthine oxidase and dehydrogenase an additional methionine would be quite difficult to detect by EXAFS. Though such a ligand was proposed from early work [104,105], it was not included in fits to the more extensive data of later workers [27,28,32,102] and there is thus no positive evidence favouring methionine.

nitrogen ligands, tend to increase the value of g_{av}. Thus, the free cofactor, when liberated from the enzyme in the presence of added thiols, has g_{av} values [85,86] higher than those from xanthine oxidase, a fact that argues against an all-thiolate coordination sphere for molybdenum in this enzyme, i.e., against two cysteine ligands of the metal (though not necessarily against one such ligand). More specifically in model compound work [61] involving a series of molybdenum complexes of a pyrazolylborate ligand, having an oxo group, three nitrogen ligands and two other variable ligands, two sulphur ligands are found to be sufficient to bring g_{av} up to a value matching closely those of most of the signal-giving species from xanthine oxidase [106]. Thus, evidence for a cysteine ligand of molybdenum in the molybdenum-containing hydroxylases is not forthcoming from EPR. In the case of sulphite oxidase and respiratory nitrate reductase, which lack the sulphido ligand and would therefore tend to have lower g values, the position is further complicated by the pH-dependent structural transition. However, the relatively high values of g_{av} for both these enzymes in the low-pH state tend to support the EXAFS data in suggesting, relative to xanthine oxidase, extra ligation to sulphur, presumably from cysteine.

Nitrogen ligands might be detectable from ^{14}N hyperfine coupling, seen in EPR, or if coupling is weak, in ENDOR. No such coupling is seen for any of the enzymes by EPR [106] or for xanthine oxidase by ENDOR [107]. This argues, though perhaps not definitively, against ligation of a nitrogen of an amino acid side-chain to molybdenum in the molybdenum-containing hydroxylases. ENDOR data are not available on other molybdenum enzymes and a nitrogen ligand cannot be excluded.

III. Limited evidence for locations of functional groups in proteins: timeliness of sequence analysis

Sequence information started to become available on pterin molybdenum cofactor enzymes, from protein sequencing as far back as 1977 [108], and from DNA sequencing from 1984 [109]. Sequencing is now providing an important input to research in this area.

Locating the binding sites of cofactors or prosthetic groups in proteins and enzymes such as those with which we are concerned, in the absence of X-ray crystallographic information, is a difficult problem. The precise nature of the environment provided for a cofactor by the protein, can of course have a very important influence on its reactivity in different enzymes. If the cofactor is reversibly dissociable and can be replaced by closely related analogues, then affinity labelling may in principle be used to identify amino acid residues in the sequence that are in the vicinity of the cofactor binding site. FAD is reversibly dissociable from xanthine oxidase and though analogues suitable for affinity work are available [110], no work on these lines seems to be have been carried out on this enzyme. In contrast to the FAD, the pterin molybdenum cofactor has so far been shown to be liberated from molybdenum enzymes only under conditions that lead to irreversible denaturation, so that this approach is not available for locating this cofactor molecule in the proteins.

Substrate binding sites in molybdenum-containing hydroxylases should be more readily identifiable than the pterin cofactor binding site. In particular, the exceptionally low substrate specificity of the enzymes should facilitate the design of affinity labels. However, to our knowledge no such work has so far been carried out on xanthine oxidase, or indeed on any other pterin molybdenum cofactor enzyme.

Many molybdenum-containing hydroxylases exist in oxidase and NAD^+-dependent dehydrogenase forms. The exact relation between these is not fully understood; however, re-oxidation of the reduced enzymes is, in either case, known to occur via the flavin. An important contribution to work in this area was the use [111] of 5-*p*-fluorosulphonylbenzoyladenosine to label the putative NAD^+-binding site of chicken liver xanthine dehydrogenase and so to identify the sequence FFTGYR in a relevant region of the protein.

The more traditional methods of selective modification of amino acid residues in active site regions seem to have been little used (or perhaps have been found of little use) in studies of molybdenum enzymes. Of particular note, however, is work [112] on modification of a lysine residue in the substrate-binding site of xanthine oxidase. This site is presumably adjacent to the pterin cofactor site. Modification of this lysine decreased reactivity of the enzyme substantially. So far, however, it has not been located in the sequence.

Work has also been carried out [113] on an essential cysteine residue in eukaryotic nitrate reductase from *Chlorella vulgaris*. This appears to be involved in NADH-binding and not, despite statements (e.g., Ref. 114) to the contrary in the literature, in the acceptance of reducing equivalents. Though this cysteine reacts with the affinity label 5-*p*-fluorosulphonylbenzoyladenosine, no attempts analagous to those just described regarding the NAD^+ site of xanthine dehydrogenase appear to have been made to isolate a labelled peptide from this enzyme.

As has already been indicated, X-ray crystallographic information on molybdenum enzymes is not yet available. That this is the case is not for want of effort, however. Crystals of subunits of chicken liver sulphite oxidase were described in 1977 [115]. Xanthine oxidase was first crystallised [116] many years ago, even before the structure of myoglobin was determined. Attempts (e.g., Ref. 117) to grow crystals of this enzyme suitable for X-ray diffraction studies are complicated by proteo-

lysis of the enzyme but have been made at various times and in different laboratories, including that of one of the authors, and are continuing. In the meantime, it is obviously important that sequence information should be available, both for future use in conjunction with crystallographic work and as an aid to probing the active centres of the enzymes by site-directed mutagenesis.

IV. Methods of sequence alignment and database exploration

IV-A. Scope of the methods

For this review we have applied some well-established computer methods (reviewed in Refs. 118, 119) and also some recently developed algorithms [120,121] to make a thorough analysis of the sequences of the various proteins that bind the pterin molybdenum cofactor. The purposes of these analyses are as follows:

1. To classify the polypeptides and their component domains into related families within which regions of sequence homology * can be defined.
2. To identify candidates for residues that might be involved in interaction with the pterin molybdenum cofactor, or with other cofactors, prosthetic groups or metal centres, or with substrates.
3. To explore databases of protein sequences for any other proteins that might resemble the enzymes depending on the pterin molybdenum cofactor in any interesting properties, for example in the overall structure of the polypeptides or their component domains or in patterns of critical functional residues. Even if there are no candidates for similar proteins in current databases, the comparisons with all other sequenced proteins are valuable for evaluation of the possible significance of the sequence patterns found within the families of enzymes covered in this review.

IV-B. A brief review of relevant general aspects of protein structure, function and evolution

It is well established [123–125] that homologous polypeptides, which are related by divergence from a common ancestral sequence, are typically similar in three-dimensional structure and folding topology, and usually have similar, but not necessarily identical, intermolecular interactions and biological functions. Three-dimensional folds are generally more conserved than sequences. Distantly related proteins of similar structure and function do not necessarily show significant sequence similarity, although a few amino acids critical for structure or molecular interactions might be conserved. Local sequence similarity is commonly found in different enzymes that catalyse different but related reactions or bind common cofactors. These have undergone functional divergence in evolution, following gene duplication, in addition to the changes that accompany species divergence. There are also cases [126–129] in which convergent evolution appears to have occurred, giving proteins of similar function in which a few crucial residues are brought by the tertiary structure into a particular spatial arrangement, despite these residues appearing in different orders in the different proteins and in different parts of the primary structure.

Many of the more complex proteins contain more than one structural domain linked into a single polypeptide, as is the case with some of the enzymes depending on the pterin molybdenum cofactor. Sequences conserved in evolution may correspond to substructures, such as domains, or smaller regions sometimes called 'motifs' [121,130,131]. These features may be very differently spaced or arranged in different members of a family of homologous proteins. The sizes of external loops and other structures interspersed in the sequences are very variable. The conserved sequences usually correspond to parts of the structural framework of the three-dimensional fold and to sets of residues involved in binding and catalysis. If the homologous proteins are very distantly related, these crucial regions may contain very few conserved regions and recognition of them by methods of sequence comparison is difficult. There are many cases where genetic rearrangements and/or duplications in evolution appear to have generated new combinations of structural modules and thus new coupled functionalities in proteins. In eukaryotic protein families some of the modules correspond approximately to exons, which supports the idea that the primitive evolutionary units of proteins were small substructures which have become fused and shuffled by genetic rearrangments in evolution [132–134].

IV-C. The range of computer methods applied

The problem of recognition of distant homology is highly pertinent to the diverse set of enzymes depending on the pterin molybdenum cofactor, so that an appropriately wide range of computer methods has been applied to the question of whether or not regions of these sequences are related to each other. Conventional methods of sequence comparison, alignment and database searching commonly encounter the problem of the 'Grey Area' [118] or 'Twilight Zone' [121], in which there is an overlap in similarity scores between, on the one hand, pairs of sequences which are structurally

* 'Homology' is used here in the strict sense [122] of relatedness by evolutionary descent from a common ancestral sequence. 'Sequence similarity' is used as a general term to indicate any observed resemblance in amino acid sequence, and does not imply whether or not such evolutionary relatedness by divergence exists.

unrelated but show statistically significant sequence similarity and, conversely, pairs of sequences that are structurally similar but whose sequence similarity is not statistically significant. This problem is a consequence partly of the large amount and diversity of sequence information available in databases and also of the way proteins evolve so that many alternative sequences are compatible with each characteristic three-dimensional fold. It is therefore an interesting problem in computer-aided pattern recognition to find and evaluate the sparse, irregular sets of small numbers of amino acids that are diagnostic of substructures common to members of a distantly related family of polypeptides.

In alignments of two or more polypeptides, domains or motifs, sequence similarity arises from patterns of residues. The residues in the two or more sequences that are compared may be either (a) identical, (b) chemically analogous, e.g., the hydrophobic amino acids, valine, isoleucine and leucine, or (c) members of sets of amino acids that are commonly observed to be substituted as alternatives in aligned sequences of evolutionarily related homologous proteins. The latter residue sets might be obviously chemically related or might have similar roles in determining the three-dimensional conformation of the protein, for example glycine and proline as helix breakers. A number of scoring schemes have been developed that use chemical or evolutionary data, of which Dayhoff's statistically weighted mutational matrices [135] are the most commonly used in methods of sequence alignment and database searching [118, 136–138]. In such schemes each pair of amino acids, including self-self pairs, receives a characteristic score, irrespective of the context in which the compared amino acids occur. This scheme is appropriate, for example, for the search of an amino acid sequence database with a single probe sequence, or for sequence alignment by dynamic programming methods, in the absence of prior knowledge of the location of crucial residues. The classical dynamic programming algorithm was first applied to molecular biology by Needleman and Wunsch [139] and Sellars [140] and modified for local similarity detection by Smith and Waterman [141]. More recent improvements include the use of concave or stepped gap penalty functions that more accurately reflect the observed occurence of insertions and deletions in protein evolution (Refs. 142, 143; Dix, N.I.M. and Wootton, J.C. unpublished work), increases in computational efficiency [144–146] and the extension, with appropriate novel weighting schemes, to the construction of multiple alignments [147–151]. However the problem of the Grey Area still exists in the evaluation of statistical significance of marginal levels of amino acid sequence similarity deduced from these methods.

Recent advances in the detection and evaluation of local patterns of crucial residues have made some inroads into the Grey Area. These methods use scoring schemes that are specific for the residue constraints of each of the features of interest ('discriminators'), together with techniques of refinement and evaluation of automated diagnosis, that were developed originally for other areas of science and engineering, including signal detection and information retrieval. Weight matrices [152–157] and other methods of encoding consensus sequences into 'signatures' or 'profiles' (reviewed in Ref. 121) have been used to produce and refine discriminators for many interesting features of proteins. In practice most of these features are described and defined by sequence alignments, structural superpositions or both. In some cases data such as secondary structure assignments, hydrophobicity, amphiphilicity and amino acid compositional constraints are part of the description of a feature. Discriminators are any sets of scores, weights or rules designed to distinguish a particular feature of protein sequence or structure from all other features. A good discriminator is a representation of a feature that is of both descriptive and diagnostic (or 'predictive') value. The development of a good discriminator for an interesting region of a diverse protein family is an extensive research process, which has not yet been fully automated in a satisfactory way, involving repeated refinement and evaluation of the feature-specific scoring scheme.

The discrimination power of a discriminator is a measure of its ability to distinguish the feature of interest from all other features, as shown by the results of searches of a comprehensive protein sequence database. This can be evaluated by several methods, including a modification of ROC analysis used for other diagnostic systems [120,158]. A test set of proteins, containing some of the known examples of the sequences of interest, is defined prior to the database searches, and other 'candidate' sequences might emerge from inspection of the database search results during the research process. Classification of the sets of proteins according to various uncertainty criteria enables the researcher to avoid circularity in the logic by maintaining a clear distinction between the sequences that are, for example 'canonical', 'candidates' and 'unknown'. This allows the scoring scheme to be refined in an unbiassed manner.

It is valuable to use such methods in 'exploratory' mode, in which the sequence database is browsed for possible new patterns. Preliminary similarity sets derived from multiple sequence alignments, or in some cases hypothetical patterns, can be encoded as one or more discriminators and used to rank the whole database of protein sequences in order of similarity to the pattern. Many such experiments lead to the rejection of the pattern as non-diagnostic of the protein functionality of interest, but novel features also emerge. This approach is also appropriate for testing whether any sequence patterns in one family, for example those in the prokaryotic pterin molybdenum cofactor proteins,

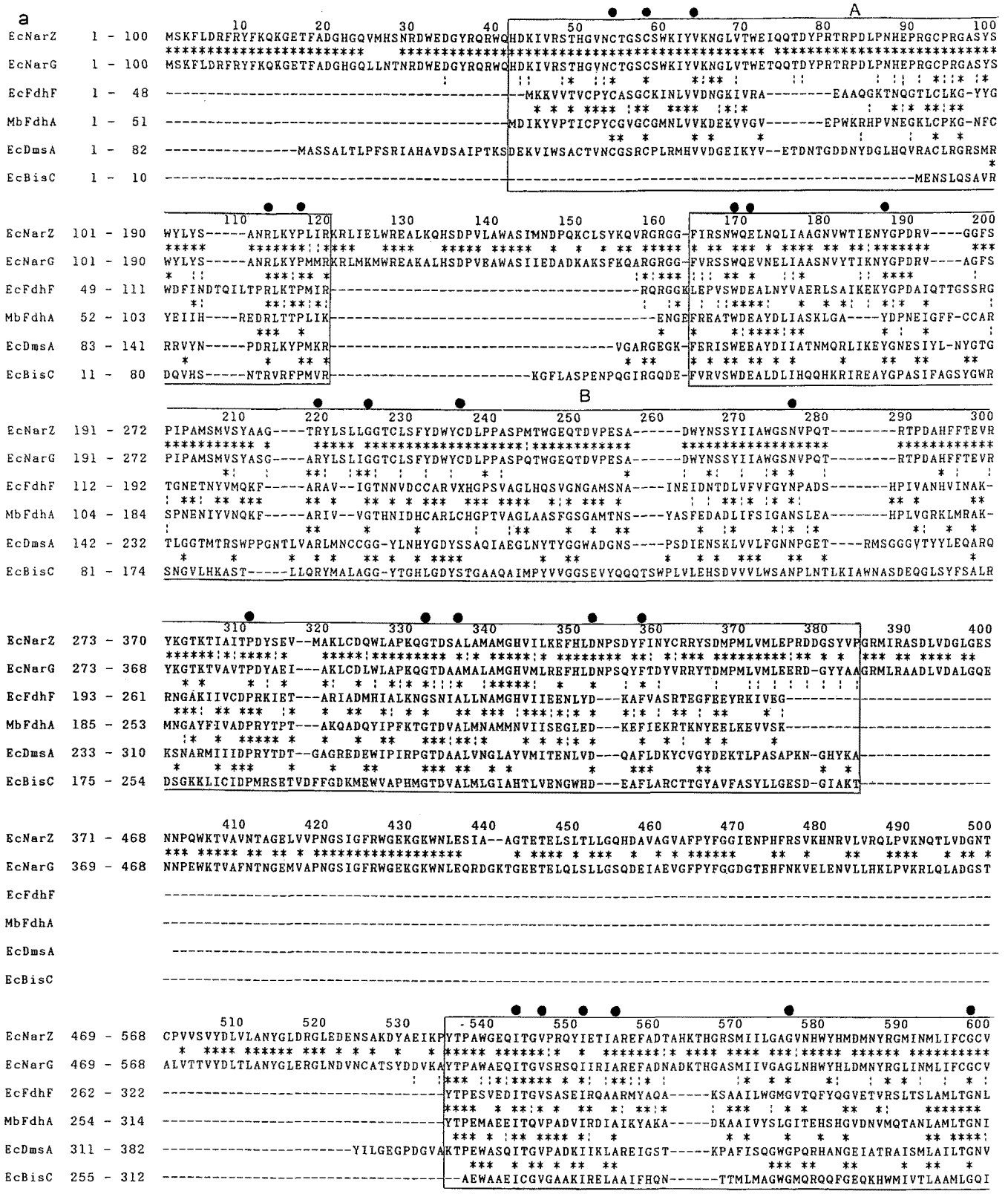
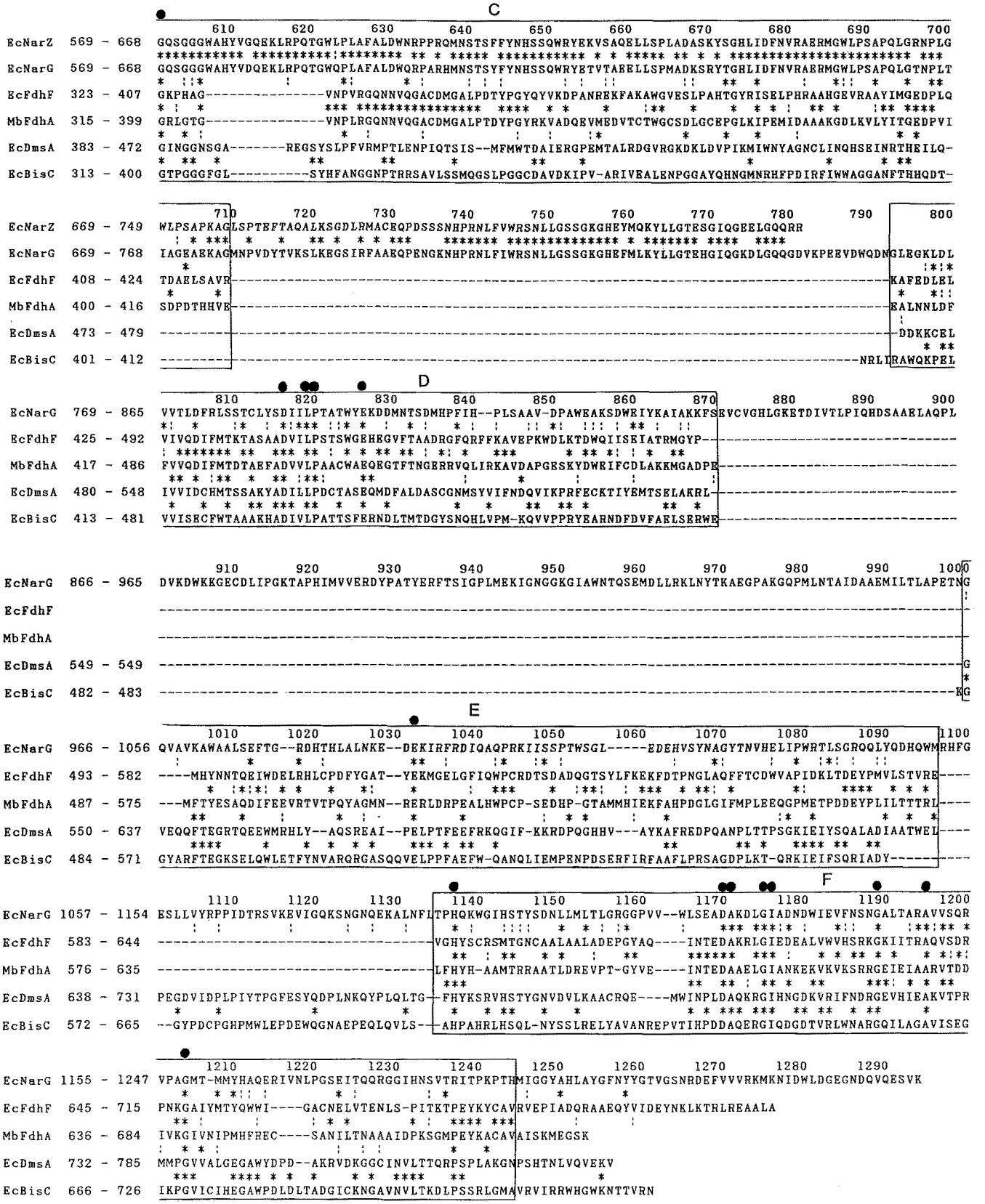


Fig. 1. (a) Alignment of amino acid sequences of some of the large polypeptides, thought to bind the pterin molybdenum cofactor, of prokaryotic pterin molybdenum cofactor enzymes. Proteins, from top to bottom, are: EcNarZ: *E. coli* second nitrate reductase (Walters, D.E., Cock, J.M. and Wootton, J.C., unpublished work), EcNarG: *E. coli* respiratory nitrate reductase α-subunit (Ref. 109; Cock, J.M., Baron, A.J., McPherson, M.J. and Wootton, J.C., unpublished work); EcFdhF: *E. coli* formate dehydrogenase, α-subunit [6], MbFdhA: *Methanobacterium formicicum* formate dehydrogenase, α-subunit [170], EcDmsA: *E. coli* dimethylsulphoxide reductase, α-subunit [168], EcBisC: *E. coli* biotin sulphoxide reductase [169].



The numbered scale above the aligned sequences (refered to as 'general numbers' in the text) defines the positions in the whole alignment, the residue numbers in the individual sequences are given at the left of each row. Symbols between the sequences: *, identical amino acids in adjacent rows, ¦, identical amino acids in non-adjacent rows. Closed circles above the sequences denote residues which are either identical in all the sequences or are of interest for other reasons, as discussed in the text, X denotes the selenocysteine residue of EcFdhF. (b) (Overleaf) The locations of regions of similar sequence, enclosed in boxes and lettered A to F, of the alignment given in (a).

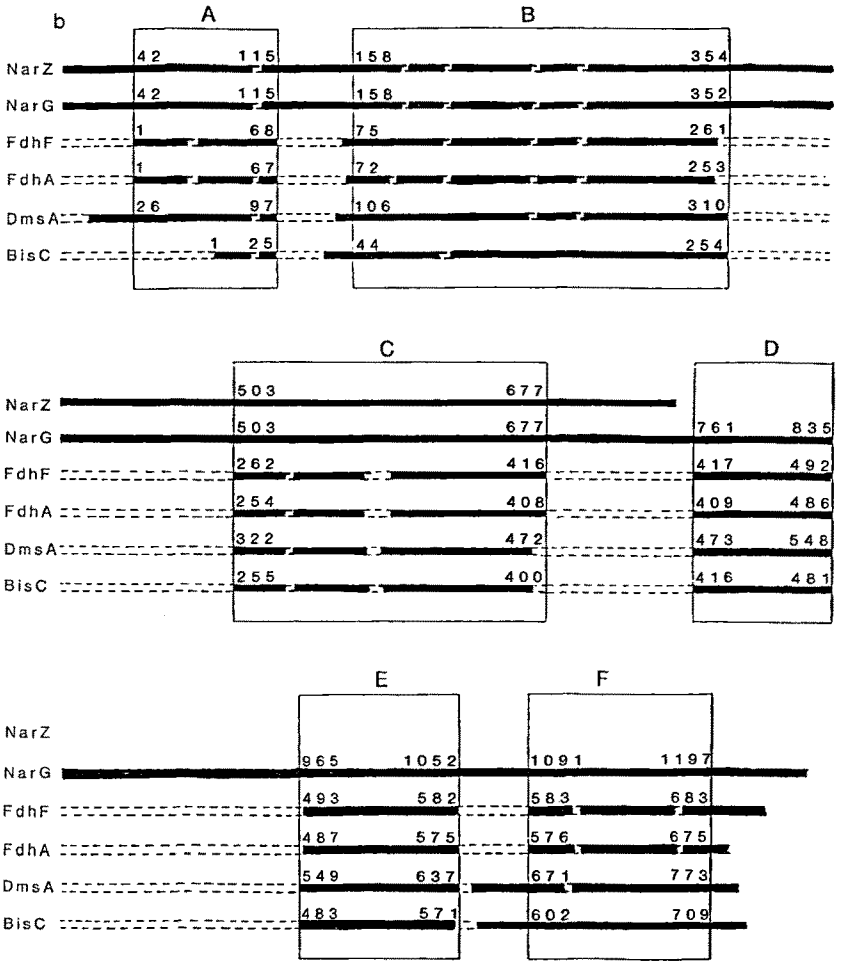


Fig. 1b.

can also be detected in another family, for example in the eukaryotic pterin molybdenum cofactor proteins. The criterion of a positive result is then if both families are adequately distinguished from the numerous other types of protein in the database. Caution is needed in both the application of this method and in the interpretation of the results, to ensure that the discriminators are made in an unbiassed way from only the members of the first family, and to take account of any proteins that unexpectedly have functionally significant similarity to the first family. These methods are more objective and sensitive than visual inspection of sequences, which can commonly detect for example the more striking patterns of cysteine and histidine residues, but not the more subtle patterns.

IV-D. Implementations used for this review

These methods * have been applied using the ISIS protein sequence/structure databases and software [120], together with ISIS-compatible SOMAP software for construction of multiple sequence alignments (Parry-Smith, D. and Attwood, T.K., University of Leeds, unpublished work), as follows.

Database searches and exploration used the OWL composite protein sequence database [159], which is an amalgamation of the non-redundant entries from seven original databases including a translation of the GenBank nucleotide sequence database. OWL is updated at 3-monthly intervals, and all searches for this review were repeated with the release 7.0 of January, 1990, which contained 20 894 entries and 5 979 599 residues. Initial global searches for sequence similarity used the SWEEP program which is a development (Bleasby, A.J. and Wootton, J.C., unpublished work) of the Wilbur-Lipman-Pearson method [136–138] modified to permit (1) flexible specification of scoring matrix and search parameters, including non-linear gap penalty functions, (2) user-specified automatic segmentation of the probe sequence and (3) formatting of the hit lists resulting from the searches into either pairwise or multiple alignment forms. These facilities are useful in allowing larger gaps, as commonly found in distantly related proteins, and in increasing the likelihood of

* All software used is available through the SERC Daresbury Laboratory, Daresbury. Warrington, WA4 4AD, U.K.

finding local similarities and domain boundaries in the long sequences of multi-domain proteins.

The multiply aligned output format from SWEEP was a useful preliminary guide to gap positions for the more refined alignment methods of SOMAP. The latter system includes a local segment-matching algorithm which was used to find the best alignments in the regions between the blocks of strong homology and to adjust the gap positions accordingly. In the alignment figures in this review (Figs. 1 and 2), the boxes enclose regions which are strongly homologous by any of the statistical criteria which are commonly applied to sequence comparison [118,160,161]. The regions between the boxes, although aligned to give the highest pairwise local segmental matches, do not have sufficient sequence similarity to give a confident indication of structural similarity, but are included because their positions in the sequences correspond. This might prove useful in future if some critical residues are located in these less conserved regions.

Pattern discriminator methods were then used for two purposes: first, to evaluate the extent to which local regions of the polypeptide domain families defined by sequence alignment were distinct from other sequences in the OWL database, and second, to test whether there are any features that show cross-similarities in the different families, for example in the prokaryotic and eukaryotic domains presumed to bind the pterin molybdenum cofactor. Components of the LUPES suite of ISIS [120] were used as follows. Program CAM was used to construct weight matrix discriminators from ungapped segments of the aligned families of lengths varying from 7 to 20 residues. Several matrices were made from each segment by using information selected judiciously from the residue frequencies at each of the aligned positions, differently adjusted according to different hypotheses of permitted residue substitution sets. CAM incorporates a database of such residue sets and algorithms that permit alternative methods of discriminator refinement by controlled adjustment of both positive and negative weights. These discriminators were used to search the whole OWL database by using program MEGASCAN and to scan other individual sequences considered in this review for the best matching segments by using program PATSCAN. Most matrices were rejected for further consideration because they were insufficiently specific in the database searches for the family from which they were constructed, but several were judged as 'interesting' for further use in the cross-family comparisons. Some of these used heavy weighting of possible hypothetical crucial residues, such as cysteine, histidine, tyrosine, aspartate and arginine, if these were conserved or largely conserved in the multiple alignments. Similar discriminator methods were also used to improve the more difficult parts of the eukaryotic multiple alignments involving xanthine dehydrogenase; in parts of the presumed pterin molybdenum cofactor binding domain and in the NAD^+/FAD domain, this sequence was very dissimilar to the other members of the eukaryotic family, and the more sensitive matrix methods were required to locate reasonable regions of best fit.

In addition, all the different classes of sequences considered in this review were subjected to several different methods of secondary structure prediction, hydrophobicity and amphiphilicity analysis. These results are not given in detail. The structure prediction methods, which are generally of no more than 60% accuracy on a per residue basis [162,163], gave results for the large subunits and iron-sulphur subunits of the prokaryotic family (see below) that are generally typical of globular proteins rich in predicted beta structure. Similar results were obtained from the presumed pterin molybdenum cofactor binding regions of the eukaryotic proteins. Because of their low reliability these results were not used as an aid to sequence alignment. Strong alpha-helix predictions, of a highly amphiphilic nature, were obtained for some segments within the large NarG and NarZ polypeptides. These are within the long inserts (Fig. 1) that are present in these nitrate reductase polypeptides and have no equivalents within the other, shorter members of this prokaryotic family of homologues. The functional significance of these inserts is unknown.

V. The prokaryotic sequence family

V-A. Sequences available

The sequences have been obtained for the genes coding for six pterin molybdenum cofactor enzymes that occur in prokaryotic organisms. These are as follows: the *narG, H, J* and *I* genes of *E. coli* code for the respiratory nitrate reductase complex. For this enzyme the whole operon has been sequenced in sections in different laboratories (Refs. 109, 164, 165; Cock, J.M., Baron, A.J., McPherson, M.H. and Wootton, J.C., unpublished work). The N-terminal segment of the *narZ* gene, which codes for the large subunit of a second nitrate reductase [166] in this organism has also been sequenced (Ref. 167; Walters, D.E., Cock, J.M. and Wootton, J.C., unpublished work). The *E. coli dmsABC* genes code for the enzyme dimethylsulphoxide reductase and for this the whole operon has been sequenced [168], as has the *bisC* gene from *E. coli*, which encodes biotin sulphoxide reductase [169]. Again in the same organism, the *fdhF* gene encoding the molybdenum-selenium polypeptide of the benzylviologen-linked formate dehydrogenase has been sequenced [6]. Finally, the *fdhAB* genes of *Methanobacterium formicicum* code for the formate dehydrogenase of this organism and have also been sequenced [170].

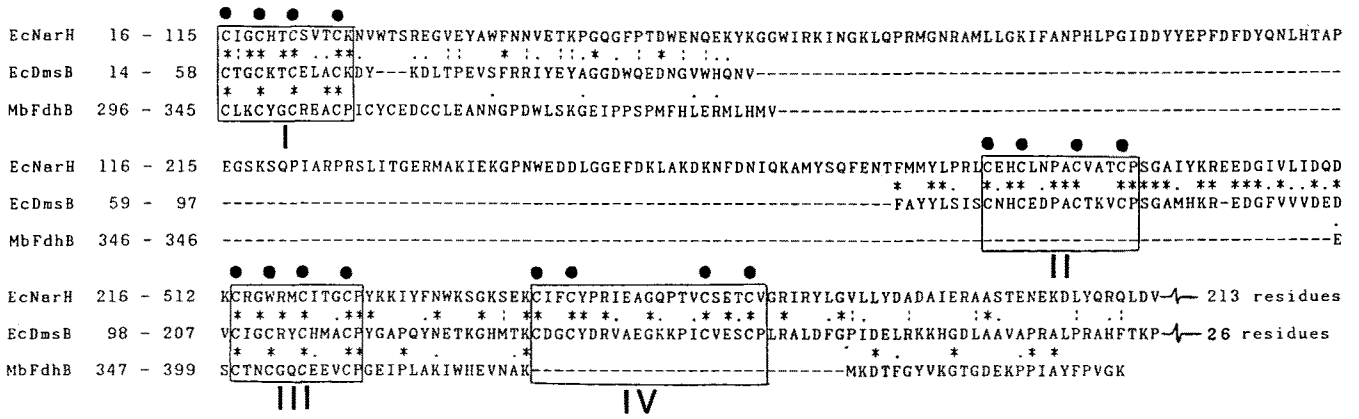


Fig. 2. Alignment of parts of the amino acid sequences of the iron-sulphur polypeptides of some prokaryotic molybdenum enzymes. The proteins are: EcNarH: *E. coli* respiratory nitrate reductase, β-subunit (Cock, J.M., Baron, A.J., McPherson, M.J. and Wootton, J.C., unpublished work), EcDmsB: *E. coli* dimethylsulphoxide reductase β-subunit [168], MbFdhB: *Methanobacterium formicicum* formate dehydrogenase, β-subunit [170]. Alignment of the sequences of the iron-sulphur clusters are shown in boxes I–IV, these being the only regions of similar sequence in these polypeptides. Symbols between sequences: *, identical amino acids in adjacent rows, ., similar amino acids, that are commonly found as alternatives in homologous proteins, in adjacent rows, ¦, identical or similar amino acids in non-adjacent rows. Closed circles above the sequences denote cysteine residues and the position in which tryptophan occurs instead of cysteine in the NarH polypeptide.

In those cases in which the coding sequences are co-transcribed as a single operon, the large polypeptide that is presumed to bind the pterin molybdenum cofactor (see below) is the promoter-proximal reading frame. This applies to NarG, DmsA and FdhA. The sequences of these polypeptides are aligned in Fig. 1. The second reading frames in these operons encode polypeptides containing cysteine clusters resembling those in ferredoxins. The genes for these are *narH*, *dmsB* and *fdhB* (cf. Fig. 2). The third reading frame (*narJ*, *dmsC*) encodes a very hydrophobic polypeptide which is presumed to be involved in membrane binding *. Evidence supporting such a role for *dmsC* has recently been obtained (Weiner, J., personal communication). In the absence of *dmsC*, *dmsA* and *dmsB* give rise to a soluble, rather than a membrane-bound complex of the A and B subunits. The *nar*I polypeptide is a *b*-type cytochrome that co-purifies with the respiratory nitrate reductase complex following release from the membranes (Refs. 172, 164; Wootton, J.C. and Baron, A.J., unpublished work).

V-B. Criteria of reliability of the sequences

With the proliferation of sequence information in this and other fields it is important to consider the reliability of the sequences. There are basically three criteria which may be used in judging sequence information. Where sequences are published in refereed journals, it is almost always a requirement of the journal that the complete sequence should have been determined on both strands of the DNA and that there should be no remaining ambiguities in the gel readings. Where this is the case the sequence should be reliable, although the onus is on the authors, not the referees, to ensure that these requirements are met in practice. A sequence should be reliable if it has been determined independently in different laboratories with identical results. Conversely, there may be lines of evidence which, while not definitive, may nevertheless be strongly suggestive of sequence errors. These include cases in which the bias of codon usage shifts to a different reading frame for parts of a polypeptide sequence [173]. This is commonly diagnostic of frame-shift errors that consist of compensating spaced pairs of nucleotide insertions or deletions. A further suggestive line of evidence for the correctness or otherwise of a sequence may be used where there is sequence homology with one or more polypeptides with similar biological function. In cases where there is conflict between proposed sequences, the one having the greater homology is usually, but not necessarily, more likely to be correct.

As already noted, various parts of the *narGHJI* operon have been sequenced in different laboratories (Refs. 109, 164, 165; Cock, J.M., Baron, M.J., McPherson, M.J. and Wootton, J.C., unpublished work). Conflicts have occurred in the and *G* and *H* reading frames, including frame-shifts. Positions of conflict have been checked retrospectively and results have confirmed that the sequence of Cock, Baron, McPherson and Wootton (unpublished work, see translated reading frames illustrated in Figs. 1 and 2) accurately reflects the DNA of the active, functionally expressed *nar* operon from cosmid clone pNAR10 [109]. Also the translated frame of NarG (Fig. 1) is more similar, in the regions of conflict, to the other proteins with which it has been

* It has been suggested [171] that *DmsC*, like *NarI*, may bear a haem group.

aligned than is the alternative sequence [165]. No conflicts of sequence have been reported for the genes encoding the other prokaryotic enzymes discussed in this section.

V-C. Which polypeptides bind the pterin molybdenum cofactor?

It is generally presumed but not rigorously proven that the large polypeptides whose sequences are aligned in Fig. 1 are the subunits that bind the pterin molybdenum cofactor. This is a difficult point to establish by biochemical analyses because of the lability of the cofactor and the usually irreversible inactivation of the proteins that follows cofactor dissociation. Experimental evidence addressing this point has been presented [174] for the *E. coli* respiratory nitrate reductase. The polypeptides of the purified, membrane-released enzyme were separated after treatment with low levels of SDS, then attempts were made to reconstitute activity from the separated products by dialysis against a source of pterin molybdenum cofactor. The 'subunit A' fraction, containing predominantly the large NarG subunit (called the A or alpha subunit in different laboratories) regained tightly bound molybdenum in parallel with recovery of some enzymic activity whereas 'subunit B' (containing predominantly the NarH, B or beta subunit) and other fractions, did not. However a problem with this work has emerged in retrospect, in the light of the amino acid sequences of these subunits. The active enzyme reconstituted from 'subunit A' was reported to contain 7.4 Fe per mol compared with 12.0 Fe for the native purified complex before SDS treatment [174], whereas 'subunit B' apparently contained no iron. As pointed out by others [165], this is difficult to reconcile with the sequence data. It is the NarH polypeptide whose sequence (Fig. 2) contains recognisable cysteine clusters characteristic of iron-sulphur centres such as [4Fe-4S] ferredoxins (although some of the NarH clusters are unusual, see below), whereas the NarG subunit does not contain any of the sequence features around its cysteine residues that are characteristic of known iron-sulphur centres. The significance of the work [174], as it relates to the molybdenum cofactor, is thus uncertain.

No other work has claimed to assign the role of molybdenum binding to any one subunit. However, independent evidence for the essential catalytic role of the large subunit has appeared in a preliminary communication [175]. Differential iodination of the purified complex was described, and it was reported that modification of the large subunit, not the other subunits, correlated with the loss of activity.

In the absence of definitive experimental evidence, what do the sequences suggest about the question of which subunit(s) bind the molybdenum pterin cofactor? First, following the precedents of other families of known sequence and three-dimensional structure in which critical binding sites are generally relatively strongly conserved, the cofactor binding sites may be presumed to be in the segments of the sequences that exhibit sequence similarity in all the homologous proteins. This excludes, for example, the large insertions in the NarG and NarZ polypeptides and the unconserved regions that are not enclosed in boxes (Fig. 1) as candidates for cofactor binding regions. Second, the sequence of the iron-sulphur subunit (NarH, DmsB, FdhB) is conserved only in the cysteine clusters themselves (Fig. 2). This implies that this subunit does not provide the ligands for the cofactor. Such a conclusion is further strengthened by the small size of this subunit, particularly DmsB, which seems too small to accommodate the cofactor in addition to the iron-sulphur clusters. Third, the large subunit of the *E. coli* formate dehydrogenase (FdhF) contains the selenocysteine residue (Fig. 1a, position 236 (general numbers)). Although there is no evidence to indicate whether or not this residue is involved in the catalytic reaction or in substrate binding, it is conceivable that it might have a role requiring it to be close to molybdenum in the three-dimensional structure. However, such an interaction might occur across an intersubunit interface. All these considerations are suggestive rather than definitive, but are consistent with the conclusion that the large subunits (Fig. 1) provide the ligands for the molybdenum cofactor.

V-D. Conserved features and possible critical residues in the prokaryotic family

The alignments of the prokaryotic sequences (Figs. 1 and 2) are, for reasons indicated above, in only partial agreement with those of Blasco et al. [165] and are more complete than those of these workers or of Bilous et al. [168]. The proteins show clear regions of sequence homology within which possible candidates for functionally critical residues might occur. Any residues which are identical or chemically similar in all the aligned sequences are likely to be crucial either for the structure of the polypeptide framework or for function. Such similarities occur only rarely by chance when three or more sequences are aligned. A caution in this type of interpretation is necessary because of a few unusual cases in which critical residues are in non-conserved regions or in different regions of the sequence in different members of the family. Nevertheless, suggestions from sequence alignments are valuable for judging priorities for protein engineering experiments in the absence of biochemical evidence for the roles of specific side-chains.

V-D.1. Interspersed conserved features

The large polypeptides (Fig. 1), that are presumed to bind the pterin molybdenum cofactor, show an interesting interspersed arrangement of the regions of sequence homology. It is a reasonable assumption, in the absence of any three-dimensional information, that the conserved blocks enclosed by boxes in Fig. 1 interact with each other spatially and form a characteristic structure that functions in binding the pterin molybdenum cofactor and in catalysis. This minimal 'molybdenum cofactor-binding fold', consisting of over 620 amino acids is evidently closely conserved across the diverse members of the prokaryotic family (the *narZ* sequence is incomplete). The strength of the evolutionary constraints on this structure is illustrated by the close similarity of the formate dehydrogenase sequences from *Methanobacterium formicicum*, an archaebacterium, and *E. coli*, a eubacterium, which are likely to be separated by more than 10^9 years of divergence. This structure is likely to consist of more than one globular domain, since it is very much larger than the 100 to 150 residues typical of the known crystal structures of coenzyme binding and catalytic domains of multidomain enzymes such as NAD^+- or FAD-dependent dehydrogenases, or ATP-binding proteins.

The homologous sequences forming this core molybdenum cofactor-binding fold are separated by sequences which are very different in length and in sequence in different members of the family. These interspersed non-conserved sequences might be external loops or inserted domains that are folded into compact globular structures outside the main conserved core structure. These 'external' regions are very long in the prokaryote nitrate reductase sequences. Such regions might function in protein–protein interactions within the multi-polypeptide complexes. Also, it cannot be ruled out that they might have roles in the specificity of substrate selection, although specificity in the catalytic reaction is likely to be determined by the local environment of the molybdenum itself.

The complex interspersed organisation of these large polypeptides underlines the difficulties likely to arise in analysing the structure–function relationships of these proteins by genetic engineering methods. Many multidomain proteins have been analysed by domain dissections by using DNA deletions and rearrangements, followed by functional analysis of the expressed products. A more complex strategy will be required in the case of the polypeptides presumed to bind the pterin molybdenum cofactor.

V-D.2. Possible ligands of molybdenum and the binding-site of the pterin molybdenum cofactor

Spectroscopic evidence on the prokaryotic enzymes (see above: Section II-B) indicates that thiolate ligands of molybdenum occur in all enzymes tested. There is no direct evidence as to whether any of these are provided by cysteine side-chains of the protein or by the pterin cofactor. Since, as discussed earlier, it is presumed that two are from the cofactor (Scheme I), the apparent presence of three thiolate ligands of molybdenum in respiratory nitrate reductase suggests that one is likely to be from the protein. Also, as was discussed earlier, the possibility of nitrogen or oxygen ligands from the protein cannot be excluded. Candidates for interesting side-chains other than cysteine are therefore methionine, histidine, lysine, arginine, glutamine, asparagine, serine, threonine, glutamate, aspartate, tyrosine, and possibly tryptophan because of its indole nitrogen. Also, the involvement of main chain carbonyl oxygens and amide nitrogens cannot be ruled out. In the interspersed conserved regions of the large subunits, there are therefore many possibilities for totally conserved nitrogen or oxygen ligands, amongst the side-chains highlighted in Fig. 1. These include all the residue types listed above except for lysine and serine. Any of these might be worth further investigation.

Possible thiolate ligands are more narrowly defined: there are three conserved cysteines in the N-terminal block (Fig. 1a), although it cannot be ruled out that these might be an unusual form of iron-sulphur cluster. However, this region is completely absent from the recently published BisC sequence and therefore might not be a general feature of the family, although there is still some doubt about the location of the initiation codon of the *bisC* reading frame, in the absence of a confirmatory protein sequence [169].

There is also a very interesting set of residues at position 236 (general numbers) in the alignment. This position is cysteine in the nitrate reductase sequences, and also in the *Methanobacterium formicicum* formate dehydrogenase. However, as already noted, the corresponding residue in the *E. coli* FdhF formate dehydrogenase is selenocysteine [6], which is not present in any of the other members of the family. The best-fitting local alignments of the DmsA and BisC sequences have serine in this position, although the sequence similarity is relatively weak. Thus, assuming that the polypeptide folds are relatively similar in these regions, a corresponding site in these proteins might be occupied by an oxygen, sulphur or selenium atom in different members of the family. There is no evidence * for or against the involvement of the selenium atom of the FdhF enzyme in functions such as substrate selection or interaction with the molybdenum atom. Clearly, the absence of selenium in the *Methanobacterium* homologue demon-

* Selenium is reported to be present in another pterin molybdenum cofactor enzyme, the xanthine dehydrogenase of *Clostridium acidiurici* [35]. However the amount of selenium in the enzyme was low (0.07 mol/mol of molybdenum) and no evidence for a functional role for it was presented.

strates that selenium and sulphur are equally compatible with the function of homologous proteins as formate dehydrogenases, but it is not known whether amino acid sequence differences elsewhere in the polypeptides have evolved to accommodate these alternative atoms.

The pterin cofactor in these enzymes contains one or perhaps two phosphate groups. The negative charges of these might be located in the protein by positively charged amino acid side-chains (cf. Ref. 176). Alternatively, a hydrogen-bonded network not involving the participation of any such groupings is at least as likely (cf. Ref. 177). If the former is the case, then either of the conserved arginine residues, in blocks A and B respectively, could fulfil this role. The possibility of a nucleotide unit forming a part of the cofactor in one or more of the enzymes made it important to search in the sequences for motifs, analagous to those detected [178] in other proteins, that might be associated with their binding sites. We were unable to find any such motifs in the enzymes under study.

V-D.3. The iron-sulphur polypeptides and the binding-sites for other cofactors

As already noted, the NarH; DmsB and FdhB alignment (Fig. 2) shows sequence homology only in the cysteine motifs. As pointed out by other workers (e.g., Ref. 165; Cammack, R. and Weiner, J., personal communication), these resemble the amino acid sequence patterns involved in binding the [4Fe-4S] clusters found in many bacterial ferredoxins [179]. However, there are some interesting differences from the consensus for this family. These ferredoxins contain pairs of cysteine motifs. The sequence pattern of each motif is -C-x-x-C-x-x-C-x-x-x-C-P- in most cases for ferredoxins bearing two [4Fe-4S] clusters, but with two or more cysteines out of the eight not conserved in ferredoxins having only one such cluster. In some cases the motif occurs with four residues between the second and third cysteines, which is possibly characteristic of a [3Fe-4S] cluster, as in the *Azotobacter vinelandii* ferredoxin I [180]. Almost all clusters from ferredoxins also have a glutamate or aspartate residue two to four positions N-terminal to the first cysteine, although this side-chain does not participate in contacts with the [4Fe-4S] cluster itself. Crystal structures are available for the *Peptococcus aerogenes* ferredoxin [181] and the *A. vinelandii* protein, for which the data have been redetermined and reinterpreted [180,182,183]. These show that the four protein ligands of each typical [4Fe-4S] cluster are provided by the sulphurs of the first three cysteine side-chains of one motif and the fourth cysteine of the other member of the pair of motifs. An interesting structural rearrangement of the protein has recently been observed in a variant of the *A. vinelandii* protein. When one of the cysteines ligating the [4Fe-4S] cluster was replaced by site-directed mutagenesis [184], the cluster structure was nevertheless maintained, by another cysteine residue, originally unligated, taking its place as a ligand.

The two cysteine motifs from *Methanobacterium formicicum* FdhB (Fig. 2) match the consensus properties of those of [4Fe-4S] proteins, but the four motifs of each of the NarH and DmsB sequences differ in the following aspects. Reading from the N-terminal end of the polypeptides, in motif (I) the residue following the fourth cysteine is lysine instead of proline. Motif (II) lacks the characteristic glutamate or aspartate in the few residues N-terminal to the first cysteine. Motif (III) is normal in the case of DmsB, but in the NarH homologue the second cysteine is replaced by tryptophan. Motif (IV) contains eleven residues between the second and third cysteines instead of the usual two or four, and in NarH the final residue is valine instead of proline. Further work is required to analyse the structural and functional meaning of these unusual sequences. The third motif of NarH, containing tryptophan instead of the second cysteine is interesting. Some ferredoxins, which are reported to contain three-iron clusters [185], have valine instead of the second cysteine [179] of the typical motif where this NarH motif contains tryptophan. Tryptophan and valine are commonly found as alternative hydrophobic residues in homologous proteins. It might therefore be speculated that the motif in NarH is associated with the three-iron cluster reported [186] to be present in *E. coli* nitrate reductase. X-ray crystallographic work has confirmed that some [3Fe-4S] clusters (typified by that of beef heart aconitase [187]), but not others [183], are readily convertible into [4Fe-4S] clusters. Whether the three-iron cluster reported in *E. coli* nitrate reductase is structurally analogous to one or other of these is not clear. However, X-ray work [183] on the [3Fe-4S] cluster of *A. vinelandii* ferredoxin, which is convertible only under denaturing conditions [188], shows that the sequence motif that provides ligands has four cysteines. Significantly, it is the second cysteine of the motif that fails to provide a ligand for the cluster, thus focusing further attention on the tryptophan in this position in NarH. Site-directed mutagenesis studies at present in progress on DmsB (Weiner, J., personal communication) involving mutations in the second cysteine residue in the iron-sulphur motifs of the protein may provide further insights into the roles of these residues in the iron-sulphur clusters.

Uniquely amongst the enzymes, formate dehydrogenase from *Methanobacterium formicicum* [189] contains FAD and has a binding site for coenzyme F_{420}. We have found no indications in the sequence comparisons of the present work that the binding sites for these might be located in FdhA or FdhB. Although it is conceivable that they might be present in the N-terminal part of FdhB (residues 1–290), the possibility that the enzyme contains additional subunits needs to be considered.

VI. The eukaryotic sequence family

VI-A. Sequences available

Since 1986, the amount of sequence data available for known eukaryotic molybdenum-containing enzymes has increased rapidly, so that at the time of writing, ten full and five partial sequences were known. These represent, however, only three different types of enzyme. The data have been obtained mostly by the sequencing of genomic or cDNA clones, although protein sequencing was used first to identify a haem-binding 'cytochrome b_5-like' domain in chicken liver sulphite oxidase [108] and *Neurospora crassa* nitrate reductase [190].

More sequences are available for plant and fungal nitrate reductases than for any other subgroup of the enzymes under consideration. For the plant enzymes, three monocotyledon nitrate reductase genes are sequenced either fully or partially, those of barley [191], corn [192] and rice (*nia*1) [193]. Dicotyledon gene sequences published include both *nia*1 and *nia*2 from *Arabidopsis thaliana* [194,195], *nia* from tomato [196] and *nia*2 from tobacco [197]; *nia*1 from tobacco is also sequenced (EMBL Acc. No. X14058 (NTNIA1)). The fungal group of nitrate reductases so far is smaller, with the full sequence of *niaD* from *Aspergillus nidulans* [114] being the only one available, except for the small length of peptide sequence [190] already mentioned for the enzyme from *Neurospora crassa*. From relatively elementary sequence comparisons it can be seen that all plant and fungal nitrate reductase proteins show a high degree of homology, with the proportion of identical residues falling slightly for the more distantly related species. Such good agreement of sequences, together with the confirmation of parts of the cDNA sequence of corn nitrate reductase by protein sequencing [192], suggests that the reliability of all the sequences is extremely good.

A similar situation applies to the second group of eukaryotic molybdenum enzymes sequenced, the molybdenum-containing hydroxylases, in particular the xanthine dehydrogenases. Three full sequences are available for this enzyme, those of the well-studied *rosy* gene in *Drosophila melanogaster* [14,15] and the equivalent genes in *Drosophila pseudoobscura* [198] and *Calliphora vicina* [199]. These genes show a high degree of sequence homology, though sequences have not been further checked by protein sequencing. It is perhaps surprising that sequencing work has not been reported on the most extensively studied eukaryotic molybdenum-containing hydroxylase, xanthine oxidase from bovine milk. However, a few years ago, cloning of the gene for the closely related rat xanthine dehydrogenase was reported [200] and sequencing work is expected to follow *. The last sequence to be noted is that of the enzyme chicken liver sulphite oxidase. This, unlike all the other full sequences, was determined entirely by protein sequence analysis [108,201].

In view of the relatively large amounts of extremely similar data for plant and fungal nitrate reductases and also for fruit fly xanthine dehydrogenases, a selection has been made in the protein sequence alignment diagrams (Figs. 3–5) shown in the following sections. These contain the chicken liver sulphite oxidase and *D. melanogaster* xanthine dehydrogenase sequences, with, at most, five of the nitrate reductase sequences, namely those of tobacco *nia*2, *A. thaliana nia*2, tomato *nia*, rice *nia*1 and *A. nidulans niaD*. These are all sequences that are complete.

VI-B. The overall structures of genes, in different eukaryotic species, coding for pterin molybdenum cofactor enzymes

In Fig. 3, the overall structures of *D. melanogaster* xanthine dehydrogenase, chicken liver sulphite oxidase, tomato nitrate reductase and *A. nidulans* nitrate reductase are compared. Each of these enzymes bears two or more cofactors or prosthetic groups on a single polypeptide chain and, as is indicated in Fig. 3 and discussed in detail below, in each case, specific regions of the gene can be identified that code for regions of the chain associated with the individual cofactors. The most striking feature of Fig. 3 is the different order in which these groups occur in the different enzymes. Thus, in xanthine dehdrogenase, the iron-sulphur centres are towards the N-terminal end, the FAD binding domain in the middle and the pterin molybdenum cofactor domain towards the C-terminus **. Relative to this, the order of the molybdenum cofactor domain and the flavin domain is reversed in the plant nitrate reductases, with the haem region inserted between them. Again, in sulphite oxidase the order of the haem domain and the molybdenum cofactor domain is reversed. Sequence homologies or similarities between corresponding regions of the different polypeptides are considered in detail

* Shortly after this article was submitted, the full sequence of rat liver xanthine dehydrogenase determined by sequencing the cDNA and by protein sequencing, was published [200a]. This enzyme consists of 1319 amino acid residues and is homologous to the *Drosophila melanogaster* enzyme, with 52% identity.

** In work published shortly after this article was submitted. Nishino and co-workers [200a] found that tryptic cleavage of rat liver xanthine dehydrogenase gave rise to three peptide fragments, 1–184, 185–539 and 540–1319. The cleavage sites thus correspond well with the domain boundaries proposed in Fig. 3 for the *Drosophila* enzyme. Furthermore. these workers came independently to conclusions identical to those presented here, regarding the respective association of the cofactors with the three domains.

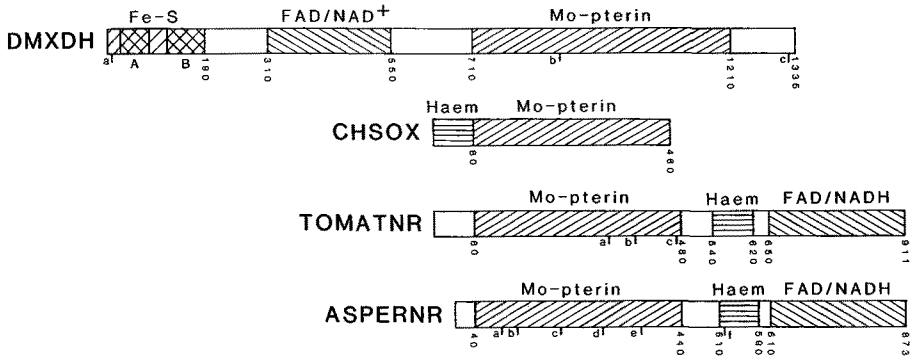


Fig. 3. Probable locations of functional domains in the amino acid sequences of some eukaryotic proteins binding the pterin molybdenum cofactor. The boundaries of the different domains are approximate. The proteins are, from top to bottom: DMXDH, *Drosophila melanogaster* xanthine dehydrogenase, [14,15], CHSOX, chicken liver sulphite oxidase [201], TOMATNR, tomato nitrate reductase (*nia* gene product) [196], ASPERNR, the fungal nitrate reductase of *Aspergillus nidulans* (*niaD* gene product) [114]. The small letters, a–f signify intron positions in the sequences, if these are known. Letters A and B denote the probable binding sites of the two [2Fe-2S] centres in *D. melangoaster* xanthine dehydrogenase. Different types of shading correspond to regions where there are sequence homologies or similarities, as discussed in the text.

```
                                10        20        30        40        50        60        70        80        90       100
DMXDH    715 - 813  DYPRFVTKGNVEEALSQADHTFEGTCRMGGQEHFYLETH-AALAVPRDSDELELFCSTQHPSEVQKLVAHVTALPAHRVVCRAKRLGGGFGGKESRGISV
                    *  .  :  .. : *  :  :   *  *. .*: * :*:  *:*   *     :*** .:.* .  :   .:::::*  .            *:**.  :
CHSOX     85 - 169  DEAPAAPDAQDPFAGDP--PRHPGL-RVNSQKPFNAEPP-AELLAERFLTPNELFFTRNHLPVPAVEPSSYRRLRVD-----------GPGGRTLSLSLA
                    .  *.  *..* ...*.  :*.*.. *.  .. ***.*.* ::*. . *.** :*:. *** :** :  .:   *.           *  * .:...
ARABNR    87 - 174  ELEPSVLDPRDEYTADSWIERNPSMVRLTGKHPFNSEAPLNRLMHHGFITPVPLHYVRNHGHVPK---AQWAEWTVEVT---------GFVKRPMKFTMD
RICENR    88 - 175  EVEPPVRDARNEGTADAWTERNPSLIRLTGKHPLNCEPPLAELMHHGFITPAALHYVANHGAVPR---GDWSTWTVDVT---------GLVKRPMRLTMD
ASPERNR   43 - 133  SKEPTEVLSLDKTTPDSHVPRDPRLIRLTGVHPFNVEPPLTALFQQGFLTPPELFYVRNHGPVPHVRDEDIPNWELRIE---------GLVEKPITLSFK
TOBACNR   79 - 166  ELEPSVHDTRDEGTADNWIERNFSMIRLTGKHPFNSEPPLNRLMHHGFITPVPLHYVRNHGPVPK---GTWDDWTVEVT---------GLVKRPMKFTMD
TOMATNR   84 - 171  ELEPSIHDTRDEGTADNWIERNFSLIRLTGKHPFNSEPPLSRLMHHGFITPVPLHYVRNHGPVPK---ASWSDWTVEVT---------GLVKRPMKFTMD

                               110       120       130       140       150       160       170       180       190       200
DMXDH    814 - 913  ALPVALAAYRMGRPVRCMLDRDEDMLITGTRHPFLFKYKVGFTKEGLITACDIECYNNAGWSMDLSFSVLERAMFHFENCYRIPNVRVGGWVCKTNLPSN
                     *  :..:  .:  ..*  .*  .*                  .  .:::  . **   . * *.  :  :** .* *: *:          *  **   * ..
CHSOX     170 - 247  ELRSRFPKHEVTATLQCAGNRRSEM---------------SRVRPVKGLPWDIGAISTARWGGASLRDVLLHAGFPEELQ-------GGEHVCFEGLDAD
                    .* * *.:.*:.*** ****** *                  . *.  **.:*. ...**. * *:.*:*** .:*.     :      *. .***** .
ARABNR    175 - 252  QLVSEFAYREFAATLVCAGNRRKEQ---------------NMVKKSKGFNWGSAGVSTSVWRGVPLCDVLRRCGIFSRKG-------GALNVCFEGSEDL
RICENR    176 - 253  ELVNGFPAVEIPVTLVCAGNRRKEQ---------------NMVQQTVGFNWGAAGVSTSVWRGARLRDVLRRCGIMPSKG-------GALNVCFEGADDL
ASPERNR   134 - 208  QILQNYDQITAPITLVCAGNRRKEQ---------------NTVRKSKGFSWGSAALSTALFTGPMMADIIKSAKPLRR----------AKYVCMEGADNL
TOBACNR   167 - 244  QLVNEFPCRELPVTLVCAGNRRKEQ---------------NMVKQTIGFNWGAAAVSTTIWRGVPLRALLKRCGVFSKNK-------GALNVCFEGADVL
TOMATNR   172 - 249  QLVNEFPSREFPVTLVCAGNRRKEQ---------------NMVKQTIGFNWGAAAVSTTVWRGVPLRALLKRCGVQSKKK-------GALNVCFEGSDVL

                               210       220       230       240       250       260       270       280       290       300
DMXDH    914 -1013  TAFRGFGGPQGMYAGEHIIRDVARIVGRDVVDVMRLNFYKTGDYTHYHQQLEHFPIERCLEDCLKQSRYDEKRQDIARFNRENRWRKRGMAVVPTKYGIA
                     .  :.*..   :.  : : ..   .  . ::*  :: : ..   .  .  :   * .   . .  :.  ::: .:* :.    :  : ::::
CHSOX     248 - 339  PGGAPYGASIPYGRA------LSPAADVLLAYEMNGTELPRDHRFPVRVVVPGVVGARSVKWLRRVAVSPDESPSRWQQNDYKGFSPCVDWDTVDYRT--
                    ****  ..*  ** .     . *. *..*** :** :*: ** ****...** .*.*:****.*. *. :** .: . .* . ... ** .  *
ARABNR    253 - 350  PGGAGTAGS-KYGTSIKKEYAMDPSRDIILAYMQNGEYLTPDHGFPVRIIIPGFIGGRMVKWLKRIIVTTKESDNFYHFKDNRVLPSLVDAELADEE-GW
RICENR    254 - 348  PGGGGS----KYGTSITRQWALDPSRDIMLAYMQNGEPPLLPDHGFPVRAIIPGCIGGRMVKWVKRIIVTTAESDNYYHYKDNRVLPSHVDAELANAD-AW
ASPERNR   209 - 301  PNG-------NYGTSIKLNWAMDPNRGIMLAHKMNGEDLRPDHGRPLRAVVPGQIGGRSVKWLKKLIITDAPSDNWYHIYDNRVLPTMVTPDMSSQNPSW
TOBACNR   245 - 339  PGGGGS----KYGTSIKKEFAMDPARDIIVAYMQNGEKLAPDHGFPVRMIIPGFIGGRMVKWIKRIIVTTQESDSYYHFKDNRVLPPHVDAELANTE-AW
TOMATNR   250 - 344  PGGGGS----KYGTSIKKEFAMDPSRDIIVAYMQNGEMLSPDHGFPVRMIIPGFIGGRMVKWLKRIVVTTQESESYYHYKDNRVLPPHVDAELANAE-AW

                               310       320       330       340       350       360       370       380       390       400
DMXDH    1014 -1113  FGVMHLNQAGSLINIYGDGSVLLSHGGVEIGQGLNTKMIQCAARALGIPSELIHISETATDKVPNTSPTAASVGSDLNGMAVLDACEKLNKRLAPIKEAL
                       :   :  ..  : .                          : ** . *: ::    .  .           :   *  :* . : *   :.*: :
CHSOX     340 - 404  --APA--IQELPVQSAVTQP-------------------RPGAAVPPGEL------TVKGYAWSGGGRRVVRVDVSLDGGRTWKVARLMGDKAP--------
                       *   *.** ..*:.* *                   .  :   :.        *.**** ****. * **.*..***:**.*::*   .  *
ARABNR    351 - 427  WYKPEYIINELNINSVITTP-------------------CHEEILPINAFTTQRP-YTLKGYAYSGGGKKVTRVEVTVDGGETWNVCALDHQEKPNK------
RICENR    349 - 425  WYKPEYIINELNVNSVITTP-------------------GYDEILPINGITTQRG-YTMKGYAYSGGGKRITRVEVTLDGGETWLVCVLDLPEKPTK------
ASPERNR   302 - 380  WRDERYAIYDLNVNSAAVYP-------------------QHKETLDLAAA---RPFYTAKGYAYAGGGRRITRVEISLDKGKSWRLARIEYAEDKYR-DFEG
TOBACNR   340 - 416  WYKPEYIINELNINSVITTP-------------------CHEEILPINAWTTQRP-YTLRGYSYSGGGKKVTRVEVTLDGGETWQVSTLDHPEKPTK------
TOMATNR   345 - 421  WYKPEYIINELNINSVITTP-------------------CHEEILPINAWTTQRP-YTLRGYAYSGGGKKVTRVEVTLDGGETWSVCTLDHPEKPTK------

                               410       420       430       440       450       460       470       480
DMXDH    1114 -1181  -----------PGGTWKEWINKAYFDR-VSLSATGFYAMPGIGYHPETNPNARTYSYYTNGVGVTVVEIDCLTGDHQVLS
                              **:.*  :.    ::   .:.:                 *  :.*:  ** :.  .*. .          :: :
CHSOX     405 - 460  -----------PGRAWAWALWELTVPV-EAGTEL------------EIVCKAVDSSYNVQPDSVAPIWNLRGVLSTAWHR
                               *. * *:.* * *:* . :.                 **: .*:* . *:**. .  **** *... :* *
ARABNR    428 - 482  -----------YGKFWCWCFWSLEVEVLDLLSAK------------EIAVRAWDETLNTQPEKM--IWNLMGMMNNCWFR
RICENR    426 - 480  -----------YGKHWCWRFWPVEVEVFDLLGAK------------EIAARAWDQSHNTQPEKL--IWNLMGMMNNCWFK
ASPERNR   381 - 446  TLYGGRVDMAWREACFCWSFWSLDIPVSELASSD------------ALLVRAMDEALSLQPKDM--YWSVLGMMNNPWFR
TOBACNR   417 - 471  -----------YGKYWCWCFWSLEVEVLDLLSAK------------EIAVRAWDETLNTQPEKL--IWNVMGMMNNCWFR
TOMATNR   422 - 476  -----------YGKYWCWCFWSLEVEVLDLLSAK------------EIAVRATDETLNTQPEKL--IWNVMGMMNNCWFR
```

Fig. 4. Alignment of the amino acid sequences of the domain, presumed to bind the pterin molybdenum cofactor, in some eukaryotic molybdoproteins. Proteins are: DMXDH, *D. melanogaster* xanthine dehydrogenase (*rosy* gene product) [14,15], CHSOX, chicken liver sulphite oxidase [201], ARABNR, RICENR, ASPERNR, TOBACNR and TOMATNR, nitrate reductases of, respectively, *Arabidopsis thaliana* (*nia*2 gene product) [195], rice (*nia*1 gene product) [193], *Aspergillus nidulans* (*niaD* gene product) [114], tobacco (*nia*2 gene product) [197] and tomato (*nia* gene product) [196]. The numbering is as in Fig 1. Symbols between sequences emphasise both identities and similarities, as in Fig 2. Closed circles above sequences denote conserved or interesting residues discussed in the text.

below. These are such as to imply (cf. Ref. 195) that for the different enzymes, the same set of ancestral genes is employed, each coding for a domain specific for a particular cofactor, but with these gene segments combined together in different orders.

The positions of known introns are also marked in Fig. 3. Not all of these occur in the regions of domain boundaries and the roles of some of them are not clear, e.g., those near the N- and C-termini of xanthine dehydrogenase. However, the data do in general conform to the hypothesis [202] that introns are deleted during evolution. In particular, there are more introns in the gene for *A. nidulans* nitrate reductase than in that for the enzyme from *A. thaliana*.

VI-C. Critical residues in the pterin molybdenum cofactor binding domain: lack of similarity to the large subunit of the prokaryotic enzymes

Work on the sequence of nitrate reductase from *A. thaliana* [195] led to the conclusion (see below) that about half of this protein is accounted for by haem and flavin domains [203], leaving the rest of the molecule as the putative pterin molybdenum cofactor binding domain. This conclusion is greatly strengthened by the finding [195,201] of sequence similarities between this nitrate reductase domain and the non-haem part of sulphite oxidase.

In Fig. 4, the sequences of the molybdenum cofactor binding domain of five plant nitrate reductases are compared with those of chicken liver sulphite oxidase and *D. melanogaster* xanthine dehydrogenase. Our alignments, except for those of the last named enzyme, are similar or identical to those of earlier workers [195,196,201]. No detailed analysis of the xanthine dehydrogenase sequence [14,15] has previously been published. However, Kinghorn and Campbell [114] attempted to align some parts of its sequence with those of some of the enzymes with which it is compared in Fig. 4. It is apparent, though, that these workers aligned it in a manner quite different from that presented here.

Fig. 4 confirms sequence homology, over the segments of the polypeptide chains that are aligned, between all the plant nitrate reductases and between these and sulphite oxidase. There is also similarity between these sequences and that of xanthine dehydrogenase. However, significance is attached to the similarity of this enzyme only in the ranges of residues 1–197 and 447–464 (general numbers).

We have been able to find no significant similarities between the eukaryotic sequences in Fig. 4 and the corresponding prokaryotic ones illustrated in Fig. 1. This is despite our having used extensively the computer methods described earlier. Matrices were either determined from the prokaryotic alignment and used to search for similarities in the eukaryotic sequences, or conversely areas of maximum homology between the eukaryotic enzymes were used to build matrices to look for similar motifs in the prokaryotic sequences. Once made, these matrices were refined rigorously. However, though such matrices showed excellent discrimination within the group (i.e., prokaryotic or eukaryotic) from which they were derived and supported the alignment of the xanthine dehydrogenase sequence shown in Fig. 4, no sequences were found within the other group that had scores over and above those of random sequences from the OWL database. Thus, it was not possible to locate the molybdenum cofactor binding domain further, by comparison between prokaryotic and eukaryotic enzymes.

As discussed in the section on the prokaryotic enzymes, in examining the sequences in Fig. 4 to locate regions possibly associated with interactions with the pterin molybdenum cofactor, it is necessary to look in regions where the sequences show the greatest similarities to one another. The match in the 33-amino acid stretch between residues 27 and 60 (general numbers) is good, with 15% of the xanthine dehydrogenase residues indentical to those in all the other enzymes, a further 18% being identical to those in sulphite oxidase only and a further 30% identical or of a type similar to that in one or more of the other enzymes. If it is the case that positively charged groups are involved in interactions with the phosphate group of the cofactor, then arginines 27 and 121 (general numbers) could be involved. A cysteine ligand of molybdenum could be provided by either of the conserved residues, 117 or 193*. The latter is next to lysine 194, which could conceivably turn out to be the residue, mentioned earlier, identified in xanthine oxidase in selective chemical modification studies.

VI-D. Critical residues in the haem, iron-sulphur and flavin domains

The haem domain of chicken liver sulphite oxidase shows extensive sequence homology with that of the corresponding plant and fungal nitrate reductases, as well as with cytochrome b_5 [203b]. This homology has been extensively commented on by earlier workers (e.g.,

* In work brought to our attention after this article was submitted. Barber and Neame [203a] reported on the amino acid sequence of rat liver sulphite oxidase. Cysteine 193 (general numbers of Fig. 4) is replaced by glutamine in this enzyme. On the other hand, cysteine 117 is conserved and is therefore still a candidate for molybdenum ligation. Indeed, these workers came independently, from sequence comparisons of sulphite oxidase and assimilatory nitrate reductases only, to the conclusion that this latter cysteine residue may provide a ligand of molybdenum in these enzymes. They did not, however, make comparisons with the xanthine dehydrogenases.

Refs. 114, 195, 196, 201, 203) and limited data are presented in Fig. 5a. The conserved histidines presumed to be ligands of the haem are in positions 37 and 63 (general numbers).

Of the enzymes being considered, only xanthine dehydrogenase contains iron-sulphur centres. These are of the [2Fe-2S] type. As for the [4Fe-4S] centres that predominate in the prokaryotic enzymes, [2Fe-2S] centres are associated with groups of cysteine residues in the sequences. However, the spacings of the cysteines differ somewhat for the two classes of iron-sulphur protein. The large number of cysteine residues at the N-terminal end of the xanthine dehydrogenase polypeptide provides strong evidence that its two [2Fe-2S] centres are located in this part of the molecule. In Fig. 5b, a short section of the xanthine dehydrogenase polypeptide is compared with part of the sequence [204] of a typical and extensively studied [2Fe-2S] protein, spinach ferredoxin. Considerable sequence similarity is apparent in the region of the five conserved cysteine residues. In spinach ferredoxin, cysteines 21, 26 29 and 61 (general numbers) are those involved in the iron-sulphur centre [204a,204b] and it is likely that these residues bind one of these centres in xanthine dehydrogenase, also. Note

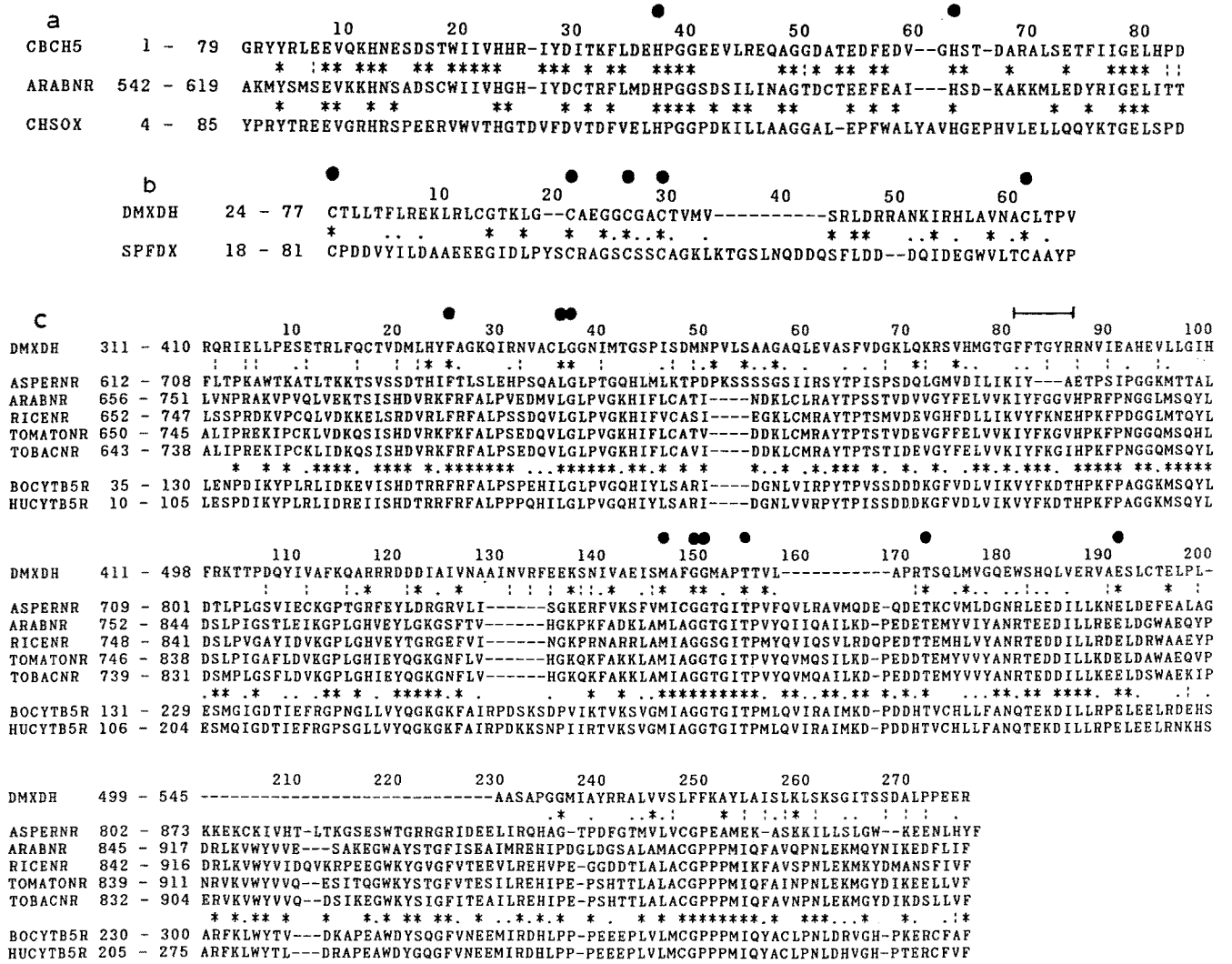


Fig. 5. Alignment of the amino acid sequences of various functional domains of some eukaryotic molybdenum enzymes depending on the pterin molybdenum cofactor. Proteins are coded as in Fig. 4. The numbering method is as in Fig. 1. Symbols between sequences indicate identities in (a), as in Fig. 1, or both identities and similarities in (b) and (c), as in Fig. 2. (a) Shows the haem-binding domain that resembes cytochrome b_5, CBCH5 [203b], in both chicken liver sulphite oxidase and in all plant and fungal nitrate reductases. Data on *Arabidopsis thaliana* (*nia*2 gene product) are illustrated. (b) Shows the ferredoxin-like N-terminal iron-sulphur domain of *D. melanogaster* xanthine dehydrogenase, aligned with part of the spinach ferredoxin sequence, SPFDX [204], used as a representative member of the known [2Fe-2S] ferredoxins. (c) Shows alignment of the C-terminal FAD/NAD^+-binding domains of five plant or fungal nitrate reductases, ASPERNR, ARABNR, RICENR, TOMATNR and TOBACNR, with the strongly conserved homologous domains of bovine and human cytochrome b_5 reductase, BOCYTB5R, HUCYTB5R [205,206] The presumed FAD/NAD^+-binding domain of *D. melanogaster* xanthine dehydrogenase is also shown, in possible alignment with these domains, though it is not strongly similar in sequence. The bar above this sequence denotes the six-amino-acid sequence identified in the chicken liver xanthine dehydrogenase by affinity labelling [111].

that these residues are present not only in the *Drosophila melanogaster* enzyme but also in those from *Drosophila pseudoobscura* and *Calliphora vicina*. Ligands of the second iron-sulphur centre in this molecule are less readily identified, but a strong candidate would be the four cysteines comprising residues 150, 169, 175 and 183 (xanthine dehydrogenase numbers).

The flavin domain of the enzymes is illustrated in Fig. 5c, in which appropriate parts of the sequences of *D. melanogaster* xanthine dehydrogenase and five plant or fungal nitrate reductases are aligned with those of bovine [205] and human [206] cytochrome b_5 reductases. Homology of the nitrate reductases with the cytochrome b_5 reductases has been pointed out by previous workers [114,195,196,203] and our alignments are similar to those of these workers. However, the sequence similarity of xanthine dehydrogenase to these proteins has not been pointed out previously, though our alignment of this protein should be regarded as only tentative *. Comparison of the various sequences with those of other flavin-containing or NAD^+-dependent enzymes reveals no sequence similarities. Those tested include various NAD^+-dehydrogenases, glutathione reductase, lipoamide dehydrogenase and mercuric reductase. Thus, the flavin domains and the associated NAD^+-binding sites in molybdenum-containing enzymes and in cytochrome-b_5 reductases, represent a family distinct from that for other enzymes of known sequence.

Residues conserved in all the enzymes as aligned, and perhaps having a functional role in FAD or NAD^+/NADH binding, include (using general numbers), phenylalanine-25, methionine-146, threonines-154 and -172 and glutamate-191. The sequence of the peptide from chicken xanthine dehydrogenase, identified [111] by affinity-labelling as a putative NAD^+ binding site and referred to earlier, is conserved in the *Drosophila* enzyme and indicated by a bar in Fig. 5c and corresponds to residues 81–86 (general numbers) **. As aligned, there is very little sequence similarity in this region between xanthine dehydrogenase and the other enzymes. This suggests that the role of this sequence in relation to NAD^+ binding may not have been unambiguously established by the affinity work. It raises particularly the possibility of the label having reacted at a site perhaps slightly removed from the actual NAD^+ binding site. Particular importance was attached in the affinity labelling work to the tyrosine in position 85 (general numbers) of xanthine dehydrogenase. The other enzymes in Fig. 5c show a conserved tyrosine at position 82. Altering the alignments so as to bring the tyrosine residues into register would reduce the overall quality of the alignments slightly in this region of the polypeptides, but not to a significant extent. A functional role of this residue, e.g., in flavin binding, is therefore not excluded.

VII. Mutational analysis of molybdenum-containing enzymes and site-directed mutagenesis studies

Ultimately the various hypotheses, put forward in the preceding sections concerning the functions in the different molybdenum-containing enzymes of the various domains or subunits and the binding-sites for the different cofactors, have to be put to the test. This requires the ability to mutagenize specific sequences in order to change single amino acids and to express at a high level the resultant protein in a system where the enzyme can be correctly assembled by the cell, and produced in sufficient quantities for biochemical and, where possible, also biophysical studies.

At first sight the prokaryotic enzymes are advantageous in this respect, the endogenous enzyme can be inactivated by mutagenesis and the modified enzyme introduced on a selectable plasmid. However, difficulties arise if the locus encoding the enzyme is duplicated [166], then mutants can be selected only at very low frequency. If the gene is vital to cell function, then again mutants can be obtained only by the selection of rare temperature-sensitive alleles. Finally, in some cases overproduction of the recombinant protein, particularly if it is normally membrane-bound, can in itself be a lethal event. These technical difficulties, which with some effort could be overcome, have so far limited mutational studies of the most obvious molybdenum containing enzyme for such work, nitrate reductase from *E. coli*. However, progress has recently been made in over-expressing another *E. coli* enzyme, dimethylsulphoxide reductase encoded by the *dmsABC* operon. Over-expression of this enzyme, which is a heterotrimer of subunits A (alpha), B (beta), C (gamma), to a content of up to 30% of the total membrane protein has been achieved (Ref. 52a; Weiner, J., personal communication). Even more interestingly, as noted earlier, a high level of over-expression of the A and B subunits to form an active enzyme, in soluble form, was achieved by genetic deletion of the membrane-attaching C subunit. Furthermore, as also mentioned above, site-directed mutagenesis work on the B subunit (DmsB) is reported by these workers to be in progress.

An alternative approach is to make mutants and to study the activity of eukaryotic enzymes in eukaryotic systems. The molecular cloning and isolation of several plant nitrate reductase genes is now well advanced, as is

* See Note added in proof on p. 185.

** In work [200a] published shortly after this article was submitted, the putative six-amino-acid NAD^+ binding site in chicken liver and *Drosophila* xanthine dehydrogenases was shown to be also well conserved in the rat enzyme, with only one amino acid substitution, threonine to proline.

the selection of nitrate reductase mutants (i.e., mutants lacking nitrate reductase activity because of modification of the structural gene for this enzyme) in *Nicotiana tabacum* [207], *Nicotiana plumbaginifolia* [208] and tomato (Koornneef, M., cited in Ref. 196). By using these, mutant enzymes can be assayed in the absence of interfering wild-type activity. Homologous plant systems, when used for the production and incorporation of the various co-factors should not be limiting in the expression of the engineered proteins. On the other hand, it is difficult to imagine that the successful expression of enzymes as complex as the molybdenum enzymes can be easily achieved in heterologous systems such as the yeast *Saccharomyces cerevisiae*, even though it has well-defined genetics and plasmid-based expression systems [209]. Alternatively, other microbial systems such as *Pseudomonas putida*, which has been shown [19,209a] to contain xanthine dehydrogenase could, be used. However, such work would be handicapped by lack of well-defined genetic markers, transposons and basic mutagenesis systems.

A thorough study of molybdenum-containing enzymes through mutagenesis requires the use of genetics, biochemistry and biophysics. In this respect xanthine dehydrogenase from *Drosophila melanogaster* encoded by the *rosy* locus is a good candidate. Chovnick's group has isolated many mutants and a number have been mapped to the DNA sequence encoding xanthine dehydrogenase (Chovnick, A., personal communication; Ref. 209b). Although the gene is large and contains non-coding introns, the use of the polymerase chain reaction [210], transfection procedures for *Drosophila* tissue culture cells [211], and P element [212] or *hobo* [213] transformation systems, should allow efficient mutagenesis and the establishment of cell lines or flies containing the mutant gene or genes. In this respect, the similarity between xanthine dehydrogenase and other eukaryotic molybdenum-containing enzymes, as well as the availability of DNA sequences from various *rosy* mutants, should permit the rapid confirmation of the various co-factor domains from xanthine dehydrogenase *, as discussed above and make possible more detailed studies by site-directed mutagenesis. An alternative insect expression system is Baculovirus, widely used for the expression of recombinant proteins [215]. However, as far as we are aware, no protein with the complex cofactor requirements of molybdenum-containing enzymes has so far been expressed in an active form in this system.

* Recent work [214] on such *rosy* mutant xanthine dehydrogenases, predicted from the present work to have mutations, respectively, in the molybdenum, flavin and iron-sulphur domains, has provided new insights into intramolecular electron transfer pathways in the enzyme and the roles of these different centres in interactions with oxidizing substrates.

VIII. Conclusions

A number of conclusions may be drawn from analyses of the data, from widely differing types of investigation, that have been presented in the preceding sections of this review.

The pterin molybdenum cofactor has now been shown to exist in two different forms, termed by Rajagopalan and co-workers 'molybdopterin' and 'molybdopterin guanine dinucleotide'. The structures of these forms seem essentially to be established, though experimental demonstration for the co-ordination of molybdenum is still lacking. The possibility that further structural variants of the cofactor may exist remains entirely open, as does the question of the distribution of different cofactor forms in different enzymes and in different species. In particular, more work is needed to establish the correctness or otherwise of a hypothesis by Meyer and co-workers [79] that one form of the cofactor is present in eukaryotes and another in prokaryotes.

With one exception (the DMS0 reductase from photosynthetic microorganisms), molybdenum cofactor enzymes contain redox-active centres (haem, iron-sulphur or flavin) in addition to molybdenum. In those prokaryotic enzymes that have been studied, these additional centres are present in subunits of the enzymes that are separate from the subunit carrying the pterin molybdenum cofactor. In contrast, in the eukayotic enzymes, the other centres are all present within different domains of a single polypeptide chain. For either type of enzyme, sequence analysis has provided important information not only regarding the molybdenum domains, as summarized below, but also (as indicated above) in relation to the binding of the other centres.

Regarding the molybdenum subunits or domains, both spectroscopic and sequence data permit classification of the enzymes into families. Not all the enzymes studied by the former methods have been studied by the latter, and *vice versa*. However, sequence analysis has permitted division of the enzymes into two broad families, prokaryotic enzymes (including several respiratory nitrate reductases and two formate dehydrogenases) and eukaryotic enzymes (including plant and fungal assimilatory nitrate reductases, chicken sulphite oxidase and fruit fly xanthine dehydrogenase). Sequence similarities have not been found between the two families. Spectroscopic and chemical data confirm that the prokaryotic nitrate reductases form a distinct class and, further, permit the subdivision of the eukaryotic family into three distinct subfamilies, each of which has its molybdenum co-ordinated in a manner that is distinct, but has yet to be completely defined.

The subfamily of the molybdenum-containing hydroxylases (especially the xanthine dehydrogenases) calls for particular comment, since its members include both prokaryotic and eukaryotic enzymes, yet according to

the spectroscopic data, all the enzymes belong to a single grouping. So far, sequence data are available only on eukaryotic enzymes of this class. It will therefore be of particular interest, when a sequence of an analgous prokaryotic enzyme becomes available, to see whether this shows similarities to both of the basic sequence families.

Sequence data have provided important information regarding the binding site of the pterin molybdenum cofactor within the molybdenum domain of the different enzymes. In the eukaryotic enzymes, comparisons of the sequences, amongst the relatively different enzymes that have been studied, have permitted identification of a relatively small number of amino acid residues that are likely to be involved in the binding or functioning of the pterin molybdenum cofactor. For the prokaryotic enzymes the greater similarities of the sequences has made it more difficult to locate the binding site with precision. However, a number of critical residues have been identified.

Expression systems have been or are being developed for a number of the enzymes and the way is thus now clearly open for the further probing of the structures and functioning of pterin molybdenum cofactor enzymes by site-directed mutagenesis.

Acknowledgements

We are most grateful to the following for providing us, during the course of our work, with information prior to publication: J.R. Andreesen, M.J. Barber, R. Cammack, A. Chovnick, R. Harrison, J. Kinghorn and J. Weiner. Work at the University of Sussex is suported by a grant to R.C.B. and J.F.B. from the Science and Engineering Research Council under the Molecular Recognition Initiative. We thank Dr. G.N. George for commenting on the co-ordination of molybdenum in the pterin cofactor. Work at the University of Leeds was supported by SERC grants to J.C.W. from the Biological Sciences Committee for sequencing nitrate reductase genes and from the Protein Engineering Initiative of the Biotechnology Directorate for development of computer software and data bases. We are grateful to Mrs. N. Ford for typing the manuscript.

References

1 Johnson, J.L. and Wadman, S.K. (1989) in Metabolic basis of inherited diseases (Scriver, C.R., Bandet, A.L., Sly, W.S. and Valle, D., eds.), 6th Edn., Vol 1, pp. 1463–1475, McGraw-Hill, New York.
2 Spiro, T.G. (1985) Molybdenum Enzymes, Wiley-Interscience, New York.
3 Coughlan, M. (1980) Molybdenum and Molybdenum-Containing Enzymes, Pergamon, Oxford.
4 Wray, J.L. and Kinghorn, J.R. (1989) Molecular and Genetic Aspects of Nitrate Assimilation, Oxford Science Publications, Oxford.
4aSolomonson, L.P. and Barber, M.J. (1990) Annu. Rev. Plant Physiol. Plant Mol. Biol. 41. 225–253.
4bCampbell, W.H. and Kinghorn, J.R. (1990) Trends Biochem. Sci. 15, 315–319.
5 Bray, R.C. (1988) Q. Rev. Biophys. 21 299–329.
6 Zinoni, F., Birkmann, A., Stadtman, T.C. and Bock, A. (1986) Proc. Natl. Acad. Sci USA 83 4650–4654.
7 Weiner, J.H., Mac Isaac, D.P., Bishop, R.E. and Bilous, P.T. (1988) J. Bacteriol. 170 1505–1510.
8 McEwan, A., Wetzstein, H.G., Meyer, O., Jackson, J.B., and Ferguson, S.J. (1987) Arch. Microbiol. 147, 340–345.
9 Satoh, T. and Kurihara, F.N. (1987), J. Biochem. 102, 191–197.
10 Bastian, N.R. and Rajagopalan, K.V. (1988), J. Cell. Biol. 107, (6iii) 845a.
11 Johnson, J.L., Bastian, N.R. and Rajagopalan, K.V. (1990) Proc. Natl. Acad. Sci. USA 87, 3190–3194.
12 Turner, N.A., Barata, B., Bray, R.C., Deistung, J., Le Gall, J. and Moura, J.J.G. (1987) Biochem. J. 243, 755–761.
13 Krenitsky, T.A., Spector, T. and Hall, W.W. (1986) Arch. Biochem. Biophys. 247, 108–119.
14 Lee, C.S., Curtis, D., McCarron, M., Love, C., Gray, M., Bender, W. and Chovnick, A. (1987) Genetics 116, 55–66.
15 Keith, T.P., Riley, M.A., Kreitman, M., Lewontin, R.C., Curtis, D. and Chambers, G. (1987) Genetics 116, 67–73.
16 Krenitsky, T.A., Shannon, M.N., Elion, G.B. and Hitchings, G.H. (1972) Arch. Biochem. Biophys. 150, 585–599.
17 Darlington, A.J., Scazzocchio, C. and Pateman, J.A. (1965) Nature 206, 599–600.
18 Scazzocchio, C. and Gorton, D. (1977) In Genetics and Physiology of *Aspergillus*, (Smith, J.E. and Pateman, J.A., eds.), pp. 255–265, Academic Press, New York.
19 Woolfolk, C.A. and Downward, J.S. (1977) J. Bacteriol. 130 1175–1191.
20 Sin, I.L. (1975) Biochim. Biophys. Acta 410, 12–20.
21 Woolfolk, C.A. and Downard, J.S. (1978) J. Bacteriol. 135, 422–428.
22 McCord, J.M. (1985), N. Engl. J. Med. 312, 159–163.
23 Harrison, R., Benboubetra, M., Bryson, S., Thomas, R.T. and Elwood, P.C. (1990) Cardioscience 1, 183–189.
24 Cramer, S.P. (1983) in Advances in Inorganic and Bioinorganic Mechanisms (Sykes, A.G., ed.), Vol. 2, pp. 259–316, Academic Press, New York.
25 Bray, R.C. (1980) Adv. Enzymol. Rel. Areas Mol. Biol. 51, 107–165.
26 Bray, R.C. (1980) in Biological Magnetic Resonance (Berliner L.J. and Reuben, J., eds.), Vol. 2, pp. 45–84, Plenum, New York.
27 Turner, N.A., Bray, R.C. and Diakun, G.P. (1989) Biochem. J. 260, 563–571.
28 Cramer, S.P. and Hille, R. (1985) J. Am. Chem. Soc. 107, 8164–8169.
29 Massey, V. and Edmondson, D. (1970) J. Biol. Chem. 245 6595–6598.
30 Malthouse, J.P.G., George, G.N., Lowe, D.J. and Bray, R.C. (1981) Biochem. J. 199 629–637.
31 Barber, M.J., Bray, R.C., Lowe, D.J. and Coughlan, M.P. (1976) Biochem. J. 153 297–307.
32 Cramer, S.P., Wahl, R. and Rajagopalan, K.V. (1981) J. Am. Chem. Soc. 103, 7721–7727.
33 Wahl, R.C. and Rajagopalan, K.V. (1982) J. Biol. Chem. 257, 1354–1359.
34 Dalton, K., Lowe, D.J., Pawlik, R.T. and Bray R.C. (1976) Biochem. J. 153, 287–295.
35 Wagner, R., Cammack, R. and Andreesen, J.R. (1984) Biochim. Biophys. Acta 791, 63–74.
36 Ventom, A.M. (1987) DPhil. Thesis, University of Sussex, U.K.

37 Wahl, R.C., Warner, C.K., Finnerty, V. and Rajagopalan, K.V. (1982) J. Biol. Chem. 257, 3958–3962.
38 Bray, R.C., George, G.N., Gutteridge, S., Norlander, L., Stell, J.G.P and Stubley, C. (1982) Biochem. J. 203, 263–267.
39 Barber, M.J., Coughlan, M.P., Rajagopalan, K.V. and Siegel, L.M. (1982) Biochemistry 21, 3561–3568.
40 Turner, N.A. (1988) DPhil, Thesis, University of Sussex, U.K.
41 Branzoli, V. and Massey, V. (1974) J. Biol. Chem. 249, 4346–4349.
42 Cramer, S.P., Moura, J.J.G., Xavier, A.V. and Le Gall, J. (1984) J. Inorg. Biochem. 20, 275–280.
43 Coughlan, M.P., Mehra, R.K., Barber, M.J. and Siegel, L.M. (1984) Arch. Biochem. Biophys. 229, 596–603.
44 Bray, R.C., George, G.N., Lange, R. and Meyer, O. (1983) Biochem. J. 211, 687–694.
45 Meyer, O., Jacobity, S. and Kruger, B. (1986) FEMS Microbiol. Rev. 39, 161–179.
46 Bray, R.C., Gutteridge, S., Lamy, T.M. and Wilkinson, T. (1983) Biochem. J. 211 227–236.
47 Cohen, H.J., Fridovich, I and Rajagopalan, K.V. (1971) J. Biol. Chem. 246, 374–382.
48 George, G.N., Kipke, C.A., Prince, R.C., Sunde, R.A., Enemark, J.H. and Cramer, S.P. (1989) Biochemistry 28, 5075–5080.
49 George, G.N. and Bray, R.C., Morpeth, F.F. and Boxer, D.H. (1985) Biochem. J. 227, 925–931.
50 George, G.N., Turner, N.A., Bray, R.C., Morpeth, F.F., Boxer, D.H. and Cramer, S.P. (1989) Biochem. J. 259. 693–700.
51 Godfrey, C., Greenwood, C., Thomson, A.J., Bray, R.C. and George, G.N. (1984) Biochem. J. 224, 601–608.
52 Turner, N., Ballard, A.L., Bray, R.C. and Ferguson, S. (1988) Biochem. J. 252, 925–926.
a. Cammack, R. and Weiner, J.H. (1990) Biochemistry 29, 8410–8416.
53 Kay, C.J. and Barber, M.J. (1989) Biochemistry 28, 5750–5758.
54 Cramer, S.P., Solomonson, L.P., Adams, M.W.W. and Mortenson, L.E. (1984) J. Am. Chem. Soc. 106 1467–1471.
55 Gutteridge, S. and Bray, R.C., Notton, B.A., Fido, R.J. and Hewitt, E.J. (1983) Biochem. J. 213, 137–142.
56 Barber, M.J., Notton, B.A., Kay, C.J. and Solomonson, L.P. (1988) Plant Physiol. 90, 70–74.
57 Barber, M.J., May, H.D. and Ferry, J.G. (1986) Biochemistry 25, 8150–8155.
58 Gadsby, P.M.A., Greenwood, C., Coddington, A., Thomson, A.J. and Godfrey, C. (1987) Biochem. J. 243 235–239.
59 Prince, R.C., Liu, C.L., Morgan, T.V. and Mortenson, L.E. (1985) FEBS Lett. 189, 263–266.
60 Cramer, S.P., Liu, C.L., Mortenson, L.E., Spence, J.T., Liu, S.M., Yamamoto, I. and Ljungdahl, L.G. (1985) J. Inorg. Biochem. 23, 119–124.
61 Cleland, W.E., Barnhart, K.M., Yamanouchi, K., Collison, D., Mabbs, F.E., Ortega, R.B. and Enemark, J.H. (1987) Inorg. Chem 26, 1017–1025.
62 Williams-Smith, D.C., Bray, R.C., Barber, M.J., Tsopanakis, A.D. and Vincent, S.P. (1977) Biochem. J. 167 593–600.
63 Orii, Y. and Morita, M. (1967) J. Biochem (Japan) 81, 163–168.
64 Barber, M.J., Siegel, L.M., Schauer, N.L., May, N.D. and Ferry, J.G. (1983) J. Biol. Chem. 258, 10839–10845.
65 Barber, M.J., Bray, R.C., Cammack, R. and Coughlan, M.P. (1977) Biochem. J. 163, 279–289.
66 Barber, M.J., Coughlan, M.P., Kanda, M. and Rajagopalan, K.V. (1980) Arch. Biochem. Biophys. 201, 468–475.
67 Kay, C.J. and Barber, N.J. (1990) Anal. Biochem. 184 11–15.
68 Barber, M.J. and Siegel, L.M. (1982) Biochemistry 21, 1638–1647.
69 Porras, A.G. and Palmer, G. (1982) J. Biol. Chem. 257, 11617–11626.
70 Cammack, R., Barber, M.J. and Bray, R.C. (1976) Biochem. J. 157, 469–478.
71 Barber, M.J., Notton, B.A. and Solomonson, L.P. (1987) FEBS Lett. 213, 372–374.
72 Kay, C.J., Barber, N.J. and Solomonson (1988) Biochemistry 27, 6142–6149.
73 Spence, J.T., Barber, N.J. and Solomonson, L.P. (1988) Biochem. J. 250, 921–923.
74 Cramer, S.P., Gray, H.B., Scott, N.S., Barber, M. and Rajagopalan, K.V. (1980) in Molybdenum Chemistry of Biological Significance (Newton W.E. and Otsuka, S., eds.), pp. 157–168, Plenum, New York.
75 Vincent, S.P. (1979) Biochem. J. 177, 757–759.
76 Johnson, J.L. (1980) in Molybdenum and Molybdenum-Containing Enzymes (Coughlan, M.P., ed.), pp. 345–383, Pergamon, Oxford.
77 Cramer, S.P. and Stiefel, E.I. (1985) in Molybdenum Enzymes (Spiro, T.G., ed.), pp. 411–441, Wiley-Interscience, New York.
78 Rajagopalan, K.V. (1989) in Molecular and Genetic Aspects of Nitrate Assimilation (Wray, J.L. and Kinghorn, J.R., eds.), pp. 212–226, Oxford Science Publications.
79 Kruger, B. and Meyer, O. (1986) Eur. J. Biochem. 157, 121–128.
80 Hawkes, T.R. and Bray, R.C. (1984) Biochem. J. 219, 481–493.
81 Johnson, J.L., Hainline, B.E. and Rajagopalan, K.V. (1980) J. Biol. Chem. 255, 1783–1786.
82 Kramer, S.P., Johnson, J.L., Ribiero, A., Millinton, D.S. and Rajagopalan, K.V. (1987) J. Biol. Chem. 262, 16357–16363.
83 Taylor, E.L., Ray, P.S., Darwish, I.S., Johnson, J.L. and Rajagopalan, K.V. (1989) J. Am. Chem. Soc. 111 7664–7665.
84 Larsen, L., Rowe, D.J., Garner, C.D. and Joule, J.A. (1989) J. Chem. Soc. Perkin Trans. I, 2317–2327.
85 Hawkes, T.R. and Bray, R.C. (1984) Biochem. J. 222, 587–600.
86 George, G.N., Hawkes, T.R., Jones, G.D. and Bray, R.C. (1986) Polyhedron 5, 587–589.
87 Deistung, J. and Bray, R.C. (1989) Biochem. J. 263, 477–483.
88 Boxer, D.H. (1989) in Molecular and Genetic Aspects of Nitrate Assimilation (Wray, J.L. and Kinghorn, J.R., eds.), pp. 27–39, Oxford Science Publications, Oxford.
89 Johnson, J.L., Wuebbens, M.M., Mandell, R. and Shih, V.E. (1989) J. Clin. Invest. 83, 897–903.
90 Johnson, M.E. and Rajagopalan, K.V. (1987) J. Bacteriol. 169, 110–116.
91 Boxer, D.H., Low, D.C., Pommier, J.A. and Giordano, G. (1987) J. Bacteriol. 169, 4678–4685.
92 Johnson, J.L., Wuebbens, M.M. and Rajagopalan, K.V. (1989) J. Biol. Chem. 264, 13440–13447.
93 May, H.D., Schauer, N.L. and Ferry, J.G. (1986) J. Bacteriol. 166, 500–504.
94 Ventom, A.M., Deistung, J. and Bray, R.C. (1988) Biochem. J. 255, 949–956.
95 Fish, K.M. and Massey, V. (1987) in Flavins and Flavoproteins (Edmondson, D. and McCormick, D. eds.), pp. 421–424, De Gruyter, Berlin.
96 Gardlik, S., Barber, M.J. and Rajagopalan, K.V. (1987) Arch. Bioch. Biophys. 259, 363–371.
97 Kruger, B. and Meyer, O. (1987) Biochim. Biophys. Acta 912, 357–364.
98 Kruger, B. and Meyer, O., Nagel, M., Andreesen, J.R., Meinike, M., Bock, E., Blumle, S. and Zumft, W.G. (1987) FEMS Microbiol. Lett. 48, 225–227.
a. Nagel, M., Koenig, K. and Andreesen, J.R. (1989) FEMS Microbiol. Lett. 60, 323–326.
99 Meyer, J.M. (1988) Inorg. Chem. 27, 3899–3903.
a. George, G.N. and Bray, R.C. (1988) Biochemistry 27, 3603–3609.
100 Kusthardt, U. and Enemark, J.H. (1987) J. Am. Chem. Soc. 109, 7926–7927.
101 Bray, R.C. and Meriwether, L.S. (1966) Nature 212, 467–469.

102 Hille, R., George, G.N., Eidsness, M.K. and Cramer, S.P. (1989) Inorg. Chem. 28, 4018–4022.
103 Flood, A. (1986) PhD Thesis, University of Manchester, U.K.
104 Tullius, T.D., Kurtz, D.M., Conradson, S.D. and Hodgson, K.O. (1979) J. Am. Chem. Soc. 101, 2776–2779.
105 Bordas, J., Bray, R.C., Garner, C.D., Gutteridge, S. and Hasnain, S.S. (1980) Biochem. J. 191, 499–508.
106 Bray, R.C. (1980) in Biological Magnetic Resonance (Berliner, L.J. and Reuben, J. eds.), Vol. 2, pp. 45–84, Plenum, New York.
107 Howes, B.D., Pinhal, N.M., Turner, N.A., Bray, R.C., Anger, G., Ehrenberg, A., Raynor, J.B. and Lowe, D.J. (1990) Biochemistry 29, 6120–6127.
108 Guiard, B. and Lederer, F. (1977) Eur. J. Biochem. 74, 181–190.
109 McPherson, M.J. Baron, A.J., Pappin, D.J.C. and Wootton, J.C. (1984) FEBS Lett. 177, 260–264.
110 Massey V., Ghisla, S and Yagi, K. (1986) Biochemistry 25, 8095–8102.
111 Nishino, T., and Nishino, T. (1989) J. Biol. Chem. 264, 5468–5473.
112 Nishino, T., Tushima, K., Hille, R. and Massey, V. (1982) J. Biol. Chem. 257, 7348–7353.
113 Barber, M.J. and Solomonson, L.P. (1986) J. Biol. Chem. 261, 4562–4567.
114 Kinghorn, J.R. and Campbell, E.I. (1989) in Molecular and Genetic Aspects of Nitrate Assimilation (Wray, J.L. and Kinghorn, J.R., eds), pp. 385–403, Oxford Science Publications, Oxford.
115 Johnson, J.L. and Rajagopalan, K.V. (1977) J. Biol. Chem. 252, 2017–2025.
116 Avis, P.C., Bergel, F., Bray, R.C. and Shooter, K.V. (1954) Nature 173, 1230–1231.
117 Rodrigues, C.G., Farchione, F., Hanson, G.R., Wilson, G.L., O'Conner, M.S. and Wedd, A.G. (1985) Climax 5th international conference on the chemistry and uses of molybdenum, Abstracts, pp. 55–56.
118 Doolittle, R.F. (1990) Methods Enzymol. 183, 99–110.
119 Feng, D. and Doolittle, R.F. (1987) J. Mol. Evol. 25, 351–360.
120 Akrigg, D., Bleasby, A.J., Dix, N.I.M., Findley, J.B.C., North, A.C.T., Parry-Smith, D., Wootton, J.C., Blundell, T.L., Gardner, S., Hayes, F., Islam, S., Sternberg, M.J.E., Thornton, J.M., Tickle, I.J. and Murray-Rust, P. (1988) Nature 335, 745–746.
121 Taylor, W.R. (1988) Protein Engin. 2, 77–86.
122 Reeck, G.R., de Haen, C., Teller, D.C., Doolittle, R.F., Fitch, W.M., Dickerson, R.E., Chambon, P., McLachlan, A.D., Margoliash, E., Jukes, T.H. et al. (1987) Cell 50, 667.
123 Rossmann, M.G., and Argos, P. (1982) Annu. Rev. Biochem. 50, 497–532.
124 Wilson, A.C., Carlson, S.S., White, T.J. (1977) Annu. Rev. Biochem. 46, 573–639.
125 Clothia, C. and Lesk, A.M. (1987) Cold Spring Harb. Symp. Quant. Biol. 52, 399–405.
126 Doolittle, R.F. (1981) Science 214, 149–159.
127 Slice, L.W., Freedman, J.H. and Rubin, C.S. (1990) J. Biol. Chem. 265, 256–263.
128 Liao, D.I. and Remington, S.J. (1990) J. Biol. Chem. 265, 6528–6531.
129 Carter, P. and Wells, J.A. (1988) Nature 332, 564–568.
130 Blundell, T.L., Sibanda, B.L., Sternberg, M.J. and Thornton, J.M. (1987) Nature 326, 347–53.
131 Gribskov, M., Homyak, M., Edenfield, J. and Eisenberg, D. (1988) Comput. Appl. Biosci. 4, 61–6.
132 Gilbert, W. (1987) Cold Spring Harb. Symp. Quant. Biol. 52, 901–5.
133 Go, M., Nosaka, M. (1987) Cold Spring Harb. Symp. Quant. Biol. 52, 915–924.
134 Holland, S.K. and Blake, C.C. (1987) Biosystems 20, 181–206.
135 Dayhoff, M.O., Schwartz, R.M., Orcutt, B.C. (1978) in Atlas of Protein Sequence and Structure, Vol. 5, Suppl. 3 (Dayhoff, M.O., ed.), pp. 345–352, Natl. Biomed. Res. Found., Washington.
136 Wilbur, W.J. and Lipman, D.J. (1983) Proc. Natl. Acad. Sci. USA 80, 726–730.
137 Lipman, D.J. and Pearson, W.R. (1985) Science 227, 1435–1441.
138 Pearson, W.R. and Lipman, D.J. (1988) Proc. Natl. Acad. Sci. USA 85, 2444–8.
139 Needleman, S.B. and Wunsch, C.D. (1970) J. Mol. Biol. 48, 443–453.
140 Sellers, P.H. (1974) SIAM, J. Appl. Math. 26, 787–793.
141 Smith, T.F., Waterman, M.S. (1981) J. Mol. Biol. 147, 195–197.
142 Altschul, S.F. and Erickson, B.W. (1986) Bull. Math. Biol. 48, 603–616.
143 Miller, W. and Myers, E.W. (1988) Bull. Math. Biol. 50, 97–120.
144 Fickett, J.W. (1984) Nucleic Acids Res. 12, 175–180.
145 Myers, E.W. and Miller, W. (1988) Comput. Appl. Biosci. 4, 11–17.
146 Spouge, J.L. (1989) SIAM, J. Appl. Math. 49, 1552–1566.
147 Johnson, M.S. and Doolittle, R.F. (1986) J. Mol. Evol. 23, 267–78.
148 Barton, G.J. and Sternberg, M.J. (1987) J. Mol. Biol. 198, 327–337.
149 Taylor, W.R. (1987) Comput. Appl. Biosci. 3, 81–87.
150 Altschul, S.F. and Lipman, D.J. (1989) SIAM J. Appl. Math. 49, 197–209.
151 Lipman, D.J., Altschul, S.F. and Kedecioglu, J.D. (1989) Proc. Natl. Acad. Sci. USA 86, 4412–4415.
152 Stormo, G.D., Schneider, T.S., Gold, L. and Ehrenfeucht, A. (1982) Nucleic Acids Res. 10, 2997–3011.
153 Staden, R. (1984) Nucleic Acids Res. 12, 505–519.
154 Gribskov, M., McLachlin, A.D. and Eisenberg, D. (1987) Proc. Natl. Acad. Sci. USA 84, 4355–4358.
155 Gribskov, M., Luthy, R. and Eisenberg, D. (1990) Methods Enzymol. 183, 146–159.
156 Drummond, M.H., Whitty, P. and Wootton, J.C. (1986) EMB0 J. 5, 441–447.
157 Wootton, J.C. and Drummond, M.H. (1989) Protein Engin. 2, 534–543.
158 Swets, J.A. (1988) Science 240, 1285–1293.
159 Bleasby, A.J. and Wootton, J.C. (1990) Protein Engin. 3, 153–159.
160 Collins, J.F. and Coulson, A.F. (1990) Methods Enzymol. 183 474–87.
161 Lipman, D.J., Wilbur, W.J., Smith, T.F. and Waterman, M.S. (1984) Nucleic Acids Res. 12, 215–226.
162 Palau, J. Argos, P. and Puigdomenech, P. (1982) Int. J. Pept. Protein Res. 19, 394–401.
163 Kabsch, W. and Sander, C. (1983) FEBS Lett. 155, 179–82.
164 Sodergren, E.J. and DeMoss, J.A. (1988) J. Bacteriol. 170, 1721–1729.
165 Blasco, F., Iobbi, C., Giordano, G., Chippaux, M. Bonnefoy, V. (1989) Mol. Gen. Genet. 218, 249–256.
166 Iobbi, C., Santini, C-L., Bonnefoy, V. and Giordano, G. (1987) Eur. J. Biochem. 168, 451–459.
167 Bonnefoy, B., Burini, J-F, Giardaro, G., Pascal, M.-C, and Chippaux, M. (1987) Mol. Microbiol. 1, 143–150.
168 Bilous, P.T., Cole, S.T., Anderson, W.F. and Weiner, J.H. (1988) Molecular Microbiol. 2, 785–795.
169 Pierson, D.E. and Campbell, A. (1990) J. Bacteriol. 172, 2194–2198.
170 Shuber, A.P., Orr, E.C. Recny, M.A., Schendel, P.F., May, H.D., Schauer, N.L. and Ferry, J.G. (1986) J. Biol. Chem. 261, 12942–12947.
171 Esposti, M.D. (1989) Biochim. Biophys. Acta 977, 249–265.
172 Morpeth, F.F. and Boxer, D.H. (1985) Biochemistry 24, 40–46.
173 Grosjean, H. and Fiers, W. (1982) Gene 18, 199–209.

174 Chaudhry, G.R. and MacGregor, C.H. (1983) J. Bacteriol. 154, 387–394.
175 Graham, A. and Boxer, D.H. (1980) Biochem. Soc. Trans. 8, 329–330.
176 Lundquist, T. and Schneider, G. (1989) J. Biol. Chem. 264, 7078–7083.
177 Pflugrath, J.W. and Quiocho, F.A. (1988) J. Mol. Biol. 200, 163–180.
178 Wievenga, R.K., Terpstra, P. and Hol, W.G. (1986) J. Mol. Biol. 187, 101–107.
179 Bruschi, M. and Guerlesquin, F. (1988) FEMS Microbiol. Rev. 54, 155–176.
180 Stout, G.H., Turley, S., Sieker, L.C. and Jensen, L.H. (1988) Proc. Nat. Acad. Sci. USA 85, 1020–1022.
181 Adman, E.T., Sieker, L.C. and Jensen, L.H. (1976) J. Biol. Chem. 251, 3801–3810.
182 Stout, C.D. (1988) J. Biol. Chem. 263, 9256–9260.
183 Stout, C.D. (1989) J. Mol. Biol. 205, 545–555.
184 Martin, A.E., Burgess, B.K., Stout, C.D., Cash, V.L., Dean, D.R., Jensen, G.M. and Stephens, P.J. (1990) Proc. Nat. Acad. Sci. USA 87, 598–602.
185 Beinert, H. and Thomson, A.J. (1983) Arch. Biochem. Biophys. 222, 333–361.
186 Johnson. M.K., Bennett, D.E., Morningstar, J.E., Adams, M.W.W. and Mortenson, L.E. (1985) J. Biol. Chem. 260, 5456–5463.
187 Robbins, A.H. and Stout, C.D. (1989) Proc. Natl. Acad. Sci. USA 86, 3639–3643.
188 Morgan, T.V., Stephens, P.J., Brugess, B.K. and Stout, C.D. (1984) FEBS Lett. 167, 137–141.
189 Schauer, N.L. and Ferry, J.G. (1986) J. Bacteriol. 165, 405–411.
190 Le, K.H.D. and Lederer, F. (1983) EMBO J. 2, 1909–1914.
191 Cheng, C.H., Dewdney, J., Kleinhofs, A. and Goodman, M.H. (1986) Proc. Natl. Acad. Sci. USA 83, 6825–6828.
192 Gowir, G. and Campbell, W.H. (1989) Plant Physiol. 90, 792–798.
193 Choi, H.K., Kleinhofs, A. and An, G. (1989) Plant Mol. Biol. 13, 731–733.
194 Cheng, C.L., Dewdney, J., Nam, H-G., Dem Boer, B.G.W. and Goodman, H.M. (1988) EMBO J. 7, 3309–3314.
195 Crawford, N.M., Smith, M., Bellissino, D. and Davis, R.W. (1988) Proc. Natl. Acad. Sci. USA 85, 5006–5010.
196 Daniel-Vedele, F., Dorbe, M.-F, Caboche, M. and Rouze, P. (1989) Gene 85. 371–380.
197 Vaucheret, H., Vincentz, M., Kronenberger, J., Caboche, M. and Rouze, P. (1989) Mol. Gen. Genet. 216, 10–15.
198 Riley, M.A. (1989) Mol. Biol. Evol. 6, 33–52.
199 Houde, M., Tiveron, M-C, and Bregregere, F. (1989) Gene 85, 391–402.
200 Yamazaki, K., Nishino, T. and Tsushima, K. (1987) in Flavins and Flavoproteins (Edmonson, D.E. and McCormick, D.B. eds.), pp. 425–428, De Gruyter, Berlin.
a. Amaya, Y., Yamazaki, K., Sato, M., Noda, K., Nishino, T. and Nishino, T. (1990) J. Biol. Chem. 265, 14170–14175.
201 Neame, P.J. and Barber, M.J. (1989) J. Biol. Chem. 264, 20894–20897.
202 Gilbert, W. (1985) Science 228, 823–824.
203 Calza, R., Huttner, E., Vincentz, M., Rouse, P., Galangau, F., Vaucheret, H., Cherel, I., Meyer, C., Kronenberger, J. and Caboche. M. (1987) Mol. Gen. Genet. 209, 552–562.
a. Barber, M.J. and Neame, P.J. (1990) J. Biol. Chem. 265, 20912–20915.
b. Nobrega, F.G. and Ozols, J. (1971) J. Biol. Chem. 246, 1706–1717.
204 Matsubara, H. and Sasaki, R.M. (1968) J. Biol. Chem. 243, 1732–1757.
a. Tsukihara, T., Fukuyama, K., Nakamura, M., Katsube, Y., Tanaka, N., Kakudo, M. Wada, K., Hase, T. and Matsubara, H. (1981) J. Biochem. 90, 1763–1773.
b. Tsutsui, T., Tsukihara, T., Fukuyama, K., Katsube, Y., Hase, T., Matsubara, H., Nishikawa, Y. and Tanaka, N. (1983) J. Biochem. 94, 299–302.
205 Ozols. J., Korza, G., Heinemann, F., Hediger, M. and Strittmatter, P. (1985) J. Biol. Chem. 260, 11953–11961.
206 Yubisui, T., Miyata, T., Iwanaga, S., Tamura, M., Yoshida, S., Takeshita, M. and Nakajima, H. (1984) J. Biochem. 96, 579–582.
207 Müller, A.J. and Grafe, R. (1978) Mol. Gen. Genet, 161, 67–76.
208 Gabard, J., Marion-Poll, A. Cherel, I., Meyer, C. Müller, A.J. and Caboche, M. (1987) Mol. Gen. Genet. 209, 596–606.
209 Derynk, R., Singh. A. and Goeddel, D.V. (1983) Nucleic Acids Res. 11, 1819.
a. Koenig, K. and Andreesen, J.R. (1990) J. Bacteriol. 172, 5999–6009.
b. Gray, M., Charpentier, A., Walsh, K., Wu, P. and Bender, W. (1991) Genetics 127, 139–149.
210 Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. and Pease, L.R. (1989) Gene 77, 51–59.
211 Saunders, S.E., Rawls, J.M., Wardle, C.J. and Burke, J.F. (1989) Nucleic Acids Res. 17, 6205–6216.
212 Engels, W.R. (1989) in Mobile DNA (Berg, D.E. and Howe, M.M., ed.), pp. 437–484, Am. Soc. Microbiol. Publ., Washington.
213 Blackman, R.K., Koehler, M.M.D., Grimaila, R. and Gelbart, W.M. (1989) EMBO J. 8, 211–217.
214 Bray, R.C., Hughes, R.K., Doyle, W.A., Whittle, J.R.S., Burke, J.F. and Chovnick, A. (1990) in Flavins and Flavoproteins (Curti, B., Zanetti, G. and Ronchi, S., eds.), De Gruyter, Berlin, in press.
215 Luckow, V.A. and Summers, M.D. (1988) Biotechnology 6, 47–55.

Note added in proof (Received 15 February 1991)

Work published shortly after this article was accepted has led us to believe that our alignment of the flavin domain of *D. melanogaster* xanthine dehydrogenase with nitrate reductases and cytochrome b_5 reductases (Fig. 5c) requires adjustment. P.A. Karplus, M.J. Daniels and J.R. Herriott (Science 251 (1991) 60–66) have located, within the sequence of the spinach ferredoxin : $NADP^+$ oxidoreductase, residues that are involved in the flavin co-factor interactions. These residues are well conserved in the plant/fungal nitrate reductases, but not in our aligment of *D. melanogaster* xanthine dehydrogenase. The overall conclusion (c.f. Refs. 200a, 214) that this region of xanthine dehydrogenase is the site of FAD/NAD^+ binding is in no way affected, and we are currently seeking to improve our alignment by taking into account the structural conclusions from the work of Karplus et al.