Bacterial morphine dehydrogenase further defines a distinct superfamily of oxidoreductases with diverse functional activities

Neil C. BRUCE,* David L. WILLEY,* Andrew F. W. COULSON† and Jonathan JEFFERY‡

*Institute of Biotechnology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QT, †Biocomputing Research Unit, Institute of Cell and Molecular Biology, University of Edinburgh, Kings Buildings, Mayfield Road, Edinburgh EH9 2JR, and ‡Department of Molecular and Cell Biology, University of Aberdeen, Marischal College, Aberdeen AB9 1AS, U.K.

Pseudomonas putida morphine dehydrogenase is shown to be closely homologous to 18 proteins, defining a superfamily within which morphine dehydrogenase particularly resembles two bacterial, 2,5-dioxo-D-gluconic acid reductases, and two eukaryotic proteins of unknown functions. Relationships within the superfamily are extensive and complex. Residue identities between protein pairs range from 29–90 %. Three subgroups are proposed. Nevertheless, on the basis of residue conservations/exchanges it is suggested that the nicotinamide coenzyme binding and substrate reduction occur in all the enzymes by broadly analogous mechanisms, among which some probable differences are identified.

INTRODUCTION

Among the several hundred known oxidoreductases dependent upon nicotinamide coenzymes, two extended superfamilies have long been recognized [1]. The enzymes of one contain catalytically important zinc, many utilizing primary alcohols as substrates, though use of secondary alcohols [2–4] and other activity [5] are also found. Enzymes of the other superfamily generally do not contain any catalytically active metal atom, have somewhat shorter subunits, and frequently but not always use large secondary alcohols as substrates. These superfamilies differ in structural type and catalytic mechanism [6–9].

Enzymes of a proposed third superfamily [10] are not homologous to those of the two superfamilies mentioned above, and their polypeptide chain lengths generally lie between those of the two superfamilies. They contain no metal atom, and their substrates include primary and secondary alcohols.

It is shown here that bacterial morphine dehydrogenase [11] is homologous to the proposed third superfamily. Structural and functional characteristics of this superfamily are demonstrated, and relationships between its members are described.

MATERIALS AND METHODS

The morphine dehydrogenase sequence [11] was compared with all sequences in the database SwissProt [12] using the computer program described [13]. Morphine dehydrogenase and all proteins showing extensive sequence similarity to it were then aligned with one another, making adjustments by eye to obtain an overall alignment. Preference was given to minimizing the numbers of gaps introduced, and to grouping the gaps together, rather than to obtaining more identities by the introduction of extra, dispersed, gaps. The alignment was made entirely without reference to any predicted secondary structure or any determined tertiary structure.

The program DSSP [14] was used together with the atomic coordinates of aldose reductase [15] to deduce aldose reductase secondary-structure elements.

RESULTS AND DISCUSSION

Proteins comprising the proposed superfamily

The databases were found to contain 18 protein sequences (of actual proteins or proteins predicted to exist from DNA data) that show considerable similarity to that of morphine dehydrogenase. Four of these proteins were mammalian aldose reductases [16–19], the others a mammalian aldehyde reductase [10], a mammalian 3α-hydroxysteroid dehydrogenase [20], a mammalian 3-oxosteroid-4-ene 5β-reductase [21], a putative reductase from *Leishmania* [22], a mammalian chlordecone reductase [23], a mammalian prostaglandin F synthase [24], a plant phytoalexin synthesis protein [25], two bacterial 2,5-dioxo-D-gluconic acid reductases [26,27], an amphibian lens protein [28], a yeast xylose reductase [29], a yeast protein of unknown function [30], a mammalian androgen-dependent protein of unknown function [31] and a plant protein of unknown function [32]. Figure 1 shows the sequences aligned with the morphine dehydrogenase sequence. Throughout the present paper, residue positional numbers refer to the continuous numbering of the morphine dehydrogenase sequence shown in Roman type (Figure 1). Where further positional identification is necessary (Table 1 and the subsection 'Active-site pocket' below), numbers are shown in *italics*, in parentheses, and refer to the human aldose reductase numbering shown in Figure 1.

General structural characteristics

All 19 sequences are identical at 22 positions (Figure 1), these conserved residues consisting of 5 Gly, 3 Arg, 2 Ala, 2 Asp, 2 Glu, 2 Pro, 1 Asn, 1 Gln, 1 Leu, 1 Lys, 1 Ser and 1 Trp.

The high proportion of Gly among these conserved residues indicates that space restrictions are limiting at similar positions in all of the proteins, consistent with conserved chain folding (cf. [33,34]).

The identities are distributed throughout most of the polypeptide chain, from residue 20 to residue 250 (Figure 1). However, the regions 38–59 and 158–164 account for 10 of the 22 identities.

‡ To whom correspondence should be addressed.

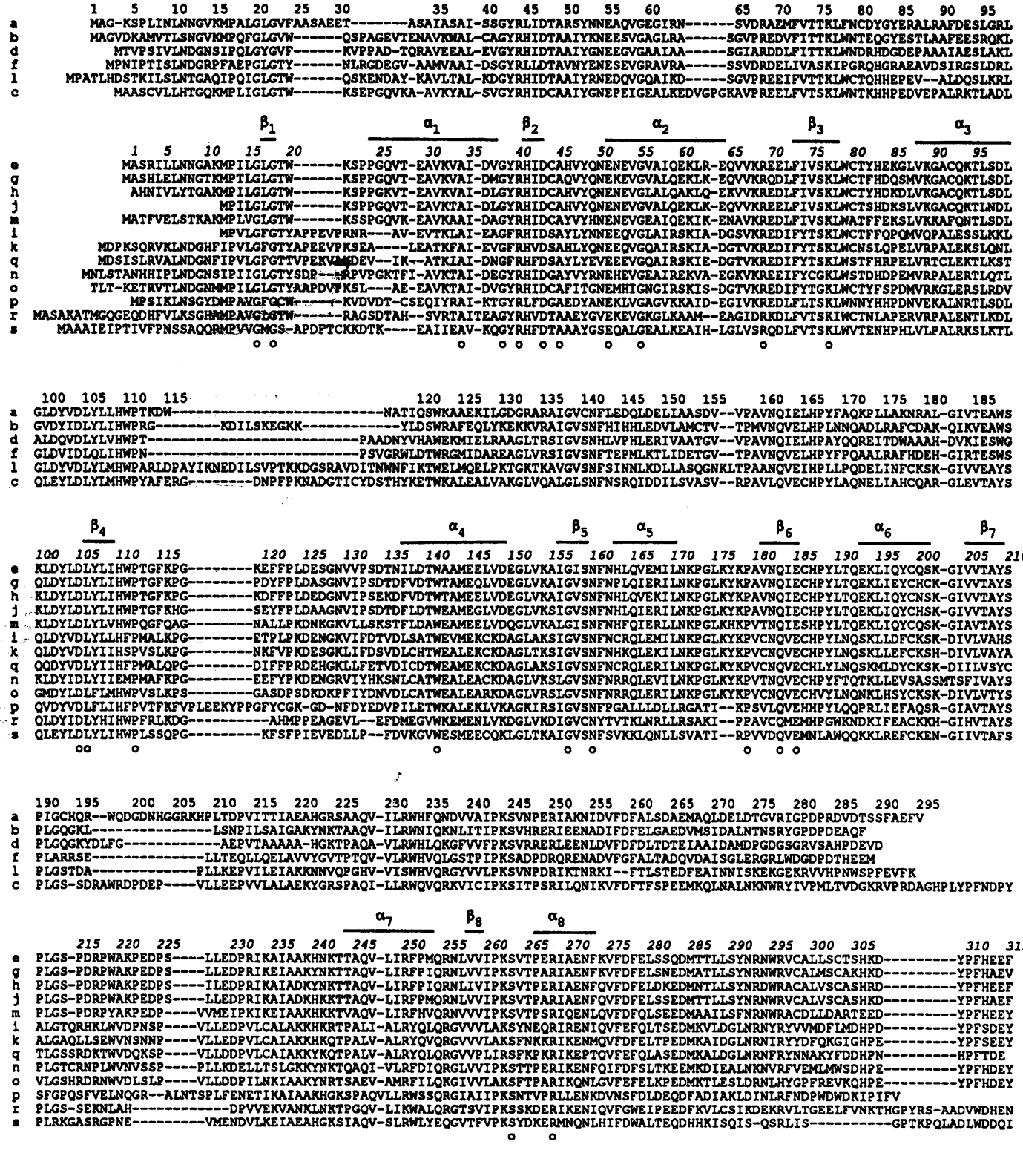

Figure 1 Alignment of primary structures of morphine dehydrogenase and homologous proteins

Residues identical in all structures are indicated (open circles below sequences). Residue positional numbers are shown above sequence a (in Roman) and sequence e (in *italic*), each numbering referring to the residue beneath its first digit. In the text, the numbering system of sequence a is generally used for all sequences, and numbers are in Roman. Where, in a few cases, this is not adequate, the sequence e number is given in parentheses in *italic*. The bars above sequence e show the α and β secondary-structure elements deduced from the crystallographic atomic co-ordinates of aldose reductase, and designated 'core barrel structures' (see the text). The proteins a–s are defined in the legend to Table 2.

Exchanges at many positions are seen to be limited and conservative (Figure 1). For example, position 101 has Asp in 17 structures and Glu in two; position 171 has Gln in 18 cases and Asn in one; and position 244 has Lys in 18 cases and Arg in one. Variations of this kind are consistent with considerable structural uniformity.

Also identified are positions at which exchanges with residues having very different properties are accepted. For example, position 37 is occupied by 11 different residues (acidic, basic, neutral and aromatic) in the 19 sequences, and positions 81 and 280 also show many non-conservative exchanges. Such variation is not restricted to regions of generally low sequence similarity nor to gap borders. At a few positions, exchange for a residue with very different properties has been accepted where otherwise one residue is preponderant, for example at position 132 (18 Gly and 1 Lys).

Gaps introduced in constructing the alignment occur principally in six regions of the polypeptide chain (between the following residues: 23–24, 64–65, 112–118, 156–157, 193–212, and after 281) (Figure 1), consistent with evolution of all these proteins from a common ancestor.

It can be seen from Figure 2 that most of the gaps correspond to segments of loops of the aldose reductase structure, compatible with conservation of a fundamental tertiary structure.

Structural and functional relationships

On the basis of crystallographic studies, functional roles have been ascribed to several residues in human aldose reductase [15]. Table 1 summarizes the functions, lists the residues involved, and relates them to the corresponding residues in the other structures. Figure 2 shows a computer-constructed representation of the α-carbon backbone of the human aldose reductase–NADPH complex, based on the crystallographic data published by Wilson et al. [15].

Using the aldose reductase atomic co-ordinates [15] and the program DSSP [14] we have identified secondary-structure elements which were not provided in the crystallographic report [15]. They are consistent with a $\beta 8/\alpha 8$ barrel motif, as proposed in [15]. The barrel is not perfectly symmetrical and, although it was mainly straightforward to distinguish between secondary structures that are part of the core barrel structure, and those that are peripheral to it, some discretionary choices were involved. The designations made for the core on this basis are shown in Figure 1 (above line e).

Further secondary structure elements (not shown in Figure 1) are a β-hairpin formed by residues (*3*)–(*13*), in which residues (*3*)–(*5*) and (*11*)–(*13*) comprise the sheet, a helix formed by residues (*231*)–(*240*), and another helix, formed from residues (*282*)–(*289*). Residues (*3*)–(*13*) are in an N-terminal region that differs considerably among the 19 sequences (Figure 1). The segment (*231*)–(*240*) is preceded by a gap-rich region, and followed by helix α_7. Residues (*282*)–(*289*) are located on the C-terminal side of helix α_8, but before a C-terminal region in which some marked differences can be seen among the various sequences (Figure 1).

Most of the 22 residues conserved in all sequences occur either within or very close to the secondary structures designated here as part of the core barrel structure (Figure 1). Of the few that do not, most are either residues to which a specific function has been ascribed, for example S-(*263*), or they are located close to such residues, for example A-(*45*), and P-(*112*) (Table 1). An exception is *R*- (*69*), which is located between α_2 and β_3, and has not yet been ascribed any specific function (Table 2).

Most of the 16 designated core structures are in regions of the alignment that do not contain gaps. However, helices α_1, α_2 and α_7 are in regions in which gaps do occur, albeit to different extents.

Hydrogen transfer

During the aldose reductase reaction, the 4-*pro*-(*R*) hydrogen of the coenzyme (NADPH) is transferred as hydride to the carbonyl carbon of the substrate, and a proton is acquired by the carbonyl oxygen. This proton is considered most likely to be provided by Tyr-52, with transfer facilitated by Lys-76 and Asp-47 [15]. Asp-47 and Lys-76 are conserved (Table 1) but Tyr-52 is exchanged for Thr in ρ-crystallin [28]. Thr could be expected to be a less ready proton donor, and ρ-crystallin has, in fact, relatively poor enzymic activity.

His-109, which has been suggested as a less likely but possible alternative to Tyr-52 for mediating the proton transfer [15], is conserved except in 3-oxosteroid-4-ene 5β-reductase, in which it is exchanged for Glu [21]. In appropriate microenvironments, His and Glu are each capable of participating in hydrogen-bonding and proton release/acceptance. Moreover, the enzyme that has Glu-109 catalyses the reduction of a carbon-to-carbon double bond, and the naturally selected catalytic steps that serve this function might differ from those evolved in the other enzymes for reduction of oxo or endoperoxide groups.

Consequently, substrate reduction is probably achieved in an exactly similar manner in 17 of the enzymes, and in analogous, slightly differing ways in the other two.

Coenzyme binding

In aldose reductase, Asn-141, Ser-140, and Gln-162 are thought to form hydrogen bonds to the coenzyme nicotinamide carboxamide group, orienting the nicotinamide ring [15]. Asn-141 and Gln-162 are conserved, but Ser-140 is replaced by Cys in two of the enzymes (morphine dehydrogenase and the barley protein of unknown function). The Cys thiol could not be expected to form a hydrogen bond exactly equivalent to that proposed for the Ser hydroxy group, but similar orientation of the nicotinamide ring should nevertheless be possible.

Residues of an 11-residue loop 244–254 in aldose reductase are thought to bind the coenzyme adenosine-2′-monophosphate moiety [15]. In all cases (Table 1), residue 244 (Lys/Arg) can form a salt bridge, Ser-245 is conserved, residue 246 is required to bond only via its main-chain NH group, the residue exchanges at 247 could all allow some form of side-chain hydrogen-bonding, and Arg-250 is conserved. Treatment of aldose reductase with phenylglyoxal resulted in selective inactivating modification of Arg-250, against which the presence of coenzyme afforded protection [35]. Engineered aldose reductase and aldehyde reductase mutants in which Met replaced Lys-244 showed altered kinetic behaviour, indicating for this residue a critical role possibly concerned with coenzyme binding and conformational changes induced by coenzyme binding [36].

Interactions proposed for Glu-253 and Asn-254 of aldose reductase [15] could not be closely mimicked by Lys/Thr/Gln or Pro/Leu respectively.

The coenzyme nicotinamide-ribose is thought to bond in aldose reductase to residues 23, 24 and 47 [15]. Lys-47 is conserved, and residues 23 and 24 bond through their main chain NH groups.

In aldose reductase, the pyrophosphate is described as forming a salt bridge with Lys-31, hydrogen-bonds with the main-chain NH groups of Ser-189, Ser-193, Leu-191 and Lys-244, and with the side chain hydroxy groups of Ser-189 and Ser-193 [15]. In other structures (Table 1), Lys-31 is exchanged for Thr/Asn/

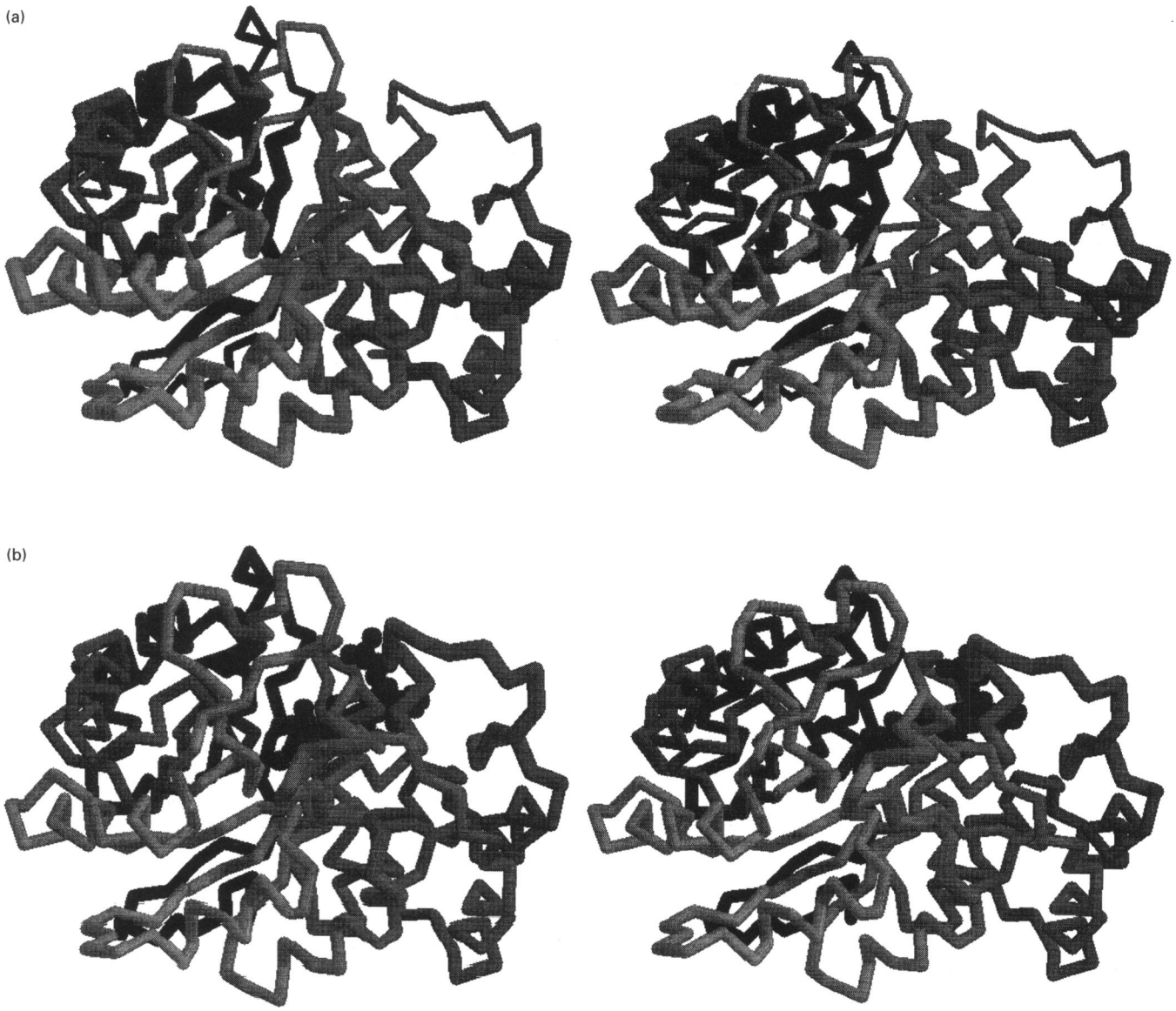

Figure 2 (a) Stereo view of the α-carbon backbone of human aldose reductase in the complex with NADPH (NADPH not shown) and (b) view as (a), but with a space-filling representation of the NADPH added

(**a**) is based on atomic co-ordinates of Wilson et al. ([15]; Brookhaven Database entry P1ads). The chain is shaded from the N-terminus (dark grey) to the C-terminus (light grey). The view is into the side of the eight-strand β/α-barrel, the axis of which runs from lower left to upper right. Regions of the sequence that correspond to gaps in any of the sequences aligned in Figure 1 are drawn with thinner rod. (**b**) the C-terminal 13 residues (centre front), and residues 193–212 (upper right) are drawn with thicker rod. Removal of the former segment lowers the catalytic activity of aldose reductase (see text), and movements of the loop formed by residues 193–212 are thought to occur with binding/release of the coenzyme (see text).

Gln/Arg, which might interact through their side chains with the pyrophosphate by hydrogen-bonding, or, in the case of Arg, by salt-bridge formation. Ser-189 and Ser-193 have numerous exchanges (Table 1) that could allow hydrogen-bonding with the main-chain NH group, but do not provide a side-chain hydroxy group, except in the case of Thr.

In the catalytic cycle of aldose reductase, a 15-residue loop consisting of residues Gly-192 to Leu-209 is thought to move in a manner resembling a hinged flap, facilitating the binding/release of the coenzyme [15]. This region has many gaps, and, where there are not gaps, non-conservative residue exchanges often occur (Figure 1).

In summary, there are probably similarities in coenzyme binding in all the enzymes, but with some variations in bonding interactions, and differences concerning binding of the pyrophosphate group and parts of the flanking ribose units.

Active-site pocket

The active-site pocket deduced for aldose reductase [15] is lined with 11 hydrophobic residues (Table 1). None of these is conserved, but four (Trp-24, Tyr-52, Trp-78, Trp-110 and Trp-197) are exchanged only for other hydrophobic residues (Table 1). Exchanges of Val-51, however, are often not conservative, and the other five residues [Table 1, Phe-(*121*), Phe-(*122*), Pro-(*218*), Leu-282, Leu-283] not only show a variety of exchanges,

Table 1 Human aldose reductase residues to which defined roles have been ascribed on the basis of crystallographic studies [15], and the corresponding residues of homologous enzymes

Residue numbers in Roman without parentheses refer to the morphine dehydrogenase numbering (Figure 1, shown in Roman), those in *italic* within parentheses refer to human aldose reductase (shown in Figure 1 in *italic*; numbered as in [15]).

Role ascribed [15]	Residues in human aldose reductase	Corresponding residues (Figure 1 alignment)
Hydrogen-transfer mechanism	Asp-47(*43*)	Asp
	Tyr-52(*48*)	Try, Thr
	or His-109(*110*)	or His, Glu
	Lys-76(*77*)	Lys
Hinged flap	Fifteen residues Gly-192(*213*) to Leu-209(*227*)	Region 192–209 has gaps and many residue exchanges
Hydrophobic side chains lining active-site pocket	Trp-24(*20*)	Trp, Phe, Tyr, Thr
	Val-51(*47*)	Ser, Ile, Asn, Asp Val, Glu, Ala, Leu
	Tyr-52(*48*)	Tyr, Thr
	Trp-78(*79*)	Trp, Phe, Pro
	Trp-110(*111*)	Trp, Phe, Ser, Met
	Phe-(*121*)	(*121*) is gap or Glu, Pro, Tyr, Ser, Leu, Phe
	Phe-(*122*)	(*122*) is gap or Gly, Thr, Cys, Asp, Leu Phe, His, Ser, Val, Tyr
	Pro-(*218*)	(*218*) is gap or Leu, Glu, Asn, Pro, Gly, Thr, Ala
	Trp-197(*219*)	Trp, Phe, Leu, Tyr, Pro
	Leu-282(*300*)	Asp, Val, Pro, Phe, Leu Gly, Met, Ala, or gap
	Leu-283(*301*)	Pro, Ser, Val, Trp, Arg Leu, Met, Glu, Gln, Asp Lys, or gap
Polar side chains within active-site pocket	Gln-53(*49*)	Asn, Gly, Glu, Lys, Arg Ala, His, Gln
	His-109(*110*)	His, Glu
	Cys-280(*298*)	Gly, Asp, Lys, Asn, Cys Leu, Ile, Tyr, Val
Residues interacting with the coenzyme	Thr-23(*19*)	Val, Thr, Cys, Ser
	Trp-24(*20*)	Phe, Tyr, Trp, Thr
	Lys-31-(*21*)	Thr, Lys, Asn, Gln, Arg
	Asp-47(*43*)	Asp
	Ser-140(*159*)	Ser, Cys
	Asn-141(*160*)	Asn
	Gln-162(*183*)	Gln
	Tyr-188(*209*)	Trp, Tyr, Phe, His
	Ser-189(*210*)	Ser, Gly, Ala, Cys
	Leu-191(*212*)	Ile, Leu, Phe,
	Ser-193(*214*)	Cys, Gln, Arg, Ser Pro, Lys, Ala, Thr
	Lys-244(*262*)	Lys, Arg
	Ser-245(*263*)	Ser
	Val-246(*264*)	Val, Ala, Asn, Phe Ser, Tyr, Thr, Ile
	Thr-247(*265*)	Asn, Arg, Asp, His Thr, Lys
	Arg-250(*268*)	Arg
	Glu-253(*271*)	Lys, Glu, Thr, Gln
	Asn-254(*272*)	Asn, Pro, Leu*

* The structure of a rat enzyme with broad substrate specificity, including 3α-hydroxysteroid dehydrogenase activity has also been reported [45–47], and differs from the 3α-hydroxysteroid dehydrogenase structures shown in Figure 1 at three positions, having Asn-247, Pro-248 and Leu-254.

but also fall in gap regions, where the polypeptide chain arrangement is likely not to be uniform in all the proteins.

Of the three residues with polar side chains noted within the active-site pocket (Gln-53, His-109, and Cys-280) [15], none is conserved, and only His-109 is exchanged solely for another residue with a polar side chain.

Crystallographic studies of an engineered aldose reductase mutant, in which Ser replaced Cys-280, indicated that when NADPH binds to aldose reductase, a conformational change occurs that moves residues 189–191 into van der Waals contact with the coenzyme, and folds down residues 192–196 over the pyrophosphate and ribose 2′-phosphate moieties of the NADPH, shutting off the top of the coenzyme-binding cleft, and enclosing the adenosine 2′-phosphate moiety of the NADPH [37]. This is broadly similar to steps in the catalytic cycle of aldose reductase proposed on the basis of crystallographic studies of the wild-type enzyme [15].

However, the mutant with the Cys-280-to-Ser mutation had altered kinetic properties, and it was thought plausible to attribute these to different interactions between the side chain of residue 280 and the coenzyme nicotinamide ring, mediated either directly or via different solvation properties of the active sites [38].

Deletion of the 13-residue C-terminal end of aldose reductase [Cys-(*303*) to Phe-(*315*)] greatly lowered the catalytic efficiency of the enzyme for uncharged substrates (as opposed to 2,5-dioxogluconic acid substrates) [39].

The segment 192–209 has many gaps, and numerous non-conservative residue exchanges (Figure 1). Also, residue 280 is Cys in only five sequences, namely the aldose reductases (Figure 1, sequences e, g, h, j and m). Moreover, the segments from residue 280 to the C-terminus exhibit considerable non-uniformity, only 10 or 11 of the 19 sequences showing similarities (Figure 1, sequences e, g, h, j, m, i, k, q, n, o, and, to a very much lesser extent, also c). Consequently, it is clear that, over the superfamily as a whole, there are significant differences in some aspects of active-site construction. This conclusion accords well with the considerable number of β/α-barrel proteins among which a variety of enzymic activities is found, employing different loop movements, but with active-site residues generally located in the loops that connect a β-strand to the following α-helix and forming the active sites in the C-terminal halves of the barrel domains [40–43].

The structures compared are of proteins from a wide range of organisms (bacteria to plants and mammals). They include dehydrogenases with widely differing substrate specificities and also proteins with presently unknown functions. The demonstrated conservation of key structural elements, together with residue divergence, in a background of general structural similarity, establishes that these proteins comprise a true superfamily, in which understanding of the structure and mechanism of one now begins to illuminate the natures of the others.

Subgroups within the superfamily

Residue identities between all pairs of the 19 structures (Figure 1) are listed in Table 2. Among these 171 pairwise comparisons, residue identities cover the wide range 29–90 %. Subdivisions can be made, as shown in the inset in Table 2.

The 15 pairwise comparisons between the proteins c, e, g, h, j and m show 49–90 % identities, and these proteins are designated Subgroup I (Table 2, inset). Four (e, g, h and j) are mammalian aldose reductases with identities (over 80 %) typical of species variants from evolutionarily close organisms [33,34]. Another is an androgen-dependent mouse protein (m), with identities with them of around 70 %, suggesting that this is not simply the mouse form of the same aldose reductase. The remaining protein assigned to Subgroup I is human aldehyde reductase (c), with

Table 2 Numbers of residue identities in pairwise comparisons of morphine dehydrogenase and 18 homologous proteins

Main table, all proteins; top half, percentage identities; bottom half, number of identities/number of residue pairs compared (excluding gaps). Below (selected proteins), subgroups, percentage identities. The 19 proteins were as follows: a, *Pseudomonas* morphine dehydrogenase [11]; b, *Leishmania* putative reductase [22]; c, human aldehyde reductase [10]; d, *Corynebacterium* 2,5-dioxo-D-gluconic acid reductase [26]; e, human aldose reductase [19]; f, *Corynebacterium* 2,5-dioxo-D-gluconic acid reductase [27]; g, rat aldose reductase [17]; h, bovine aldose reductase [18]; i, human chlordecone reductase [23]; j, rabbit aldose reductase [16]; k, bovine prostaglandin F synthase [24]; l, *Saccharomyces* GCY gene product (function unknown) [30]; m, mouse androgen-dependent protein of unknown function [31]; n, rat 3-oxosteroid-4-ene 5β-reductase [21]; o, frog ρ crystallin [28]; p, *Pichia* xylose reductase [29]; q, rat 3α-hydroxysteroid dehydrogenase [20]: r, barley protein of unknown function [32]: s, soybean reductase of phytoalexin synthesis [25].

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s
a	–	47	41	42	40	41	39	38	38	39	36	37	36	34	34	34	31	32	33
b	124/265	–	42	45	45	39	46	47	38	45	40	43	43	39	39	35	34	36	36
c	115/280	116/275	–	36	51	33	52	50	44	51	41	41	49	44	43	38	40	36	37
d	112/265	121/267	98/276	–	39	39	38	39	37	38	37	40	36	36	34	33	34	32	34
e	110/275	122/274	160/312	106/275	–	35	85	87	49	90	50	43	71	52	50	38	47	40	35
f	109/265	104/270	90/274	107/272	95/274	–	35	34	34	35	34	37	33	35	30	32	32	31	32
g	107/275	126/274	161/312	104/275	269/316	95/274	–	84	49	89	49	42	69	51	50	38	47	43	36
h	104/274	127/273	154/311	106/274	273/315	94/273	264/315	–	49	88	50	42	69	52	50	37	47	42	36
i	103/270	100/260	130/298	97/261	147/300	88/260	147/300	147/300	–	50	74	41	48	58	59	35	69	39	36
j	102/263	119/262	154/300	101/263	273/304	93/262	270/304	266/304	151/300	–	52	43	72	53	51	37	49	43	36
k	102/285	109/275	129/311	102/275	157/311	92/273	151/311	155/310	227/308	155/299	–	41	47	55	59	37	69	36	38
l	101/272	119/279	121/294	108/269	125/291	101/273	123/291	121/290	115/278	120/279	121/292	–	41	39	39	33	36	38	35
m	99/275	117/274	154/313	100/275	225/315	91/274	218/315	217/314	145/301	219/303	147/312	121/292	–	48	47	37	45	38	38
n	97/282	108/279	140/315	100/277	164/314	96/276	160/314	164/313	176/304	160/302	174/319	116/297	151/315	–	53	34	51	34	31
o	97/285	108/275	133/310	92/273	155/312	83/273	157/312	155/311	181/308	154/300	189/320	115/293	146/313	170/319	–	36	56	34	29
p	94/280	98/278	114/303	90/276	113/301	88/273	113/301	110/301	101/288	106/290	112/300	98/297	110/301	104/302	109/299	–	33	36	35
q	89/283	94/275	124/311	92/274	147/311	87/274	145/311	145/310	212/306	146/299	222/320	106/293	141/312	164/319	178/320	100/299	–	35	33
r	88/274	97/273	112/309	88/272	120/300	84/270	128/300	127/299	110/285	123/288	109/299	111/292	115/300	105/305	101/300	105/291	104/300	–	40
s	89/273	97/266	112/300	89/262	103/292	85/262	104/292	104/291	104/285	100/280	114/301	99/283	111/292	92/300	88/300	98/284	99/299	119/300	–

Subgroup I

	e	g	h	j	m
c	51	52	50	51	49
e	–	85	87	90	71
g	–	–	84	89	69
h	–	–	–	88	69
j	–	–	–	–	72

Subgroup II

	k	q	n	o
i	74	69	58	59
k	–	69	55	59
q	–	–	51	56
n	–	–	–	53

Subgroup III

	b	d	f	l
a	47	42	41	37
b	–	45	39	41
d	–	–	39	40
f	–	–	–	37

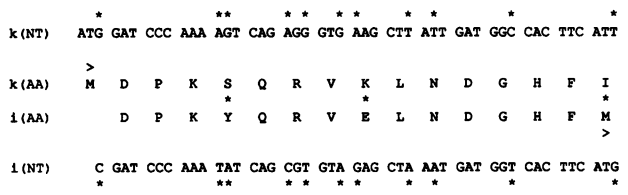

Figure 3 Comparison of the 5′-terminal segments of the nucleotide sequences determined for prostaglandin F synthase cDNA and chlordecone reductase cDNA, and the amino acid sequences they could encode

k(NT), nucleotide sequence of bovine prostaglandin F synthase cDNA 5′-terminal segment (from [24]). k(AA), N-terminal segment of the amino acid sequence of the protein (one-letter code) predicted in [24] with the initiator methionine residue indicated (> above the letter). i(NT), nucleotide sequence of human chlordecone reductase cDNA 5′-terminal segment (from [23]). i(AA), N-terminal segment of the amino acid sequence of the protein predicted in [23] begins with the initiator methionine indicated (> below the letter). The preceding 42 nucleotides, read in the same reading frame, would code for the 14 amino acid residues shown, though these nucleotides were supposed not to be translated, being 5′ to the putative initiator methionine codon [23]. The comparison shows high conservation between the predicted actual amino acid sequence of the bovine prostaglandin F synthase (k) and the hypothetical (supposedly untranslated) amino acid sequence of human chlordecone reductase (i). Non-identities are marked with an asterisk between the sequences. This amino acid level conservation (12 out of 14 residues) exceeds the conservation of the nucleotide sequences (33 out of 42 nucleotides; nucleotide non-identities are marked with an asterisk above or below the nucleotide).

around 50% identities with the other members of the subgroup. An engineered chimaeric protein in which the first 73 residues of human aldehyde reductase (that is, Ala-4 to Glu-70 of c in Figure 1), replaced the first 71 residues of human aldose reductase (that is, Ala-5 to Glu-70 of e in Figure 1) utilized various substrates with efficiency within 4-fold of that of the wild-type aldose reductase [39], indicating an interesting degree of functional equivalence between these segments.

Proteins i, k, q, n and o are assigned to a different subgroup (Subgroup II). The 10 pairwise comparisons show 51–74% identities within this subgroup (Table 2, inset). Three of the proteins (i, k and q) are mammalian reductases with different substrate specificities, one (n) is a mammalian reductase that catalyses reduction of a carbon–carbon double bond (rather than an oxo or endoperoxide group), and one is an amphibian lens protein (o) that may have a structural role, but is reported to possess low prostaglandin H_2 9,11-endoperoxide reductase activity (about 2% that of protein k).

Bovine prostaglandin F synthase (k) and human chlordecone reductase (i) are, in fact, even more closely related than can be seen from Table 2 and from Figure 1, in which different lengths at their N-termini make a noticeable dissimilarity. Comparison of the cDNA sequences [23,24] shows that 42 supposedly untranslated nucleotides 5′ to the ATG taken to correspond to the chlordecone reductase translation initiation codon would translate in the same reading frame to a 14 amino acid sequence closely similar to the predicted actual amino acid sequence of prostaglandin F synthase (Figure 3). The conservation is clearly higher at the amino acid level than at the nucleotide level (Figure 3). Selective pressures would seem unlikely to have conserved codons specifying a hypothetical amino acid sequence that was not expressed. Differential splicing of the primary transcript might provide a mechanism by which the 42 nucleotides in question could be expressed in some circumstances (accounting for the codon conservation), but not in others (accounting for the cDNA sequence reported [23]). Alternatively, the chlordecone reductase cDNA [23] may have been incomplete. Rabbit aldolase

reductase appears to be similarly short in the N-terminus (Figure 1, sequence j), and it is noteworthy that its N-terminus was not established [16].

Selections of proteins to form a Subgroup III are made more tentatively, because large differences in percentage identities are not found among the candidate proteins. The proteins assigned are a, b, d, f and l (Table 2, inset). Thus the *Pseudomonas* morphine dehydrogenase (a) is classified in this subgroup. It is 47% identical with protein b. Proteins d and f are both 2,5-dioxo-D-gluconic acid reductases from species of the same genus (*Corynebacterium*), but they are only 39% identical. Proteins b and l are of unknown functions, and are from dissimilar eukaryotes (*Leishmania* and *Saccharomyces*); they are 41% identical. The degree of identity within the subgroup (around 42%) is impressive for proteins that are evolutionarily diverse and of uncertain function *in vivo*.

Proteins clearly belonging to the superfamily, but not at present accommodated in the above subgroups, are the xylose reductase from a *Pichia* (p), the protein of unknown function from barley (r), and the soybean protein considered to be the reductase of phytoalexin synthesis (s). Additional structural and functional information would be necessary to trace their relationships further.

In conclusion, it is now established that bacterial morphine dehydrogenase is a member of a protein superfamily with distinctive structural characteristics which the morphine dehydrogenase structure helps to define. The structural archetype of the superfamily is a β/α-barrel arrangement [15]. Arrangements of this kind are well known in a variety of proteins [40], but in two other superfamilies [1,9,44] in which functional activities closely similar to those of the present superfamily occur, the structural types are quite different. Thus these three superfamilies have provided three different types of protein, each of which has proved to possess considerable structural stability and functional adaptability, permitting the evolution of structurally dissimilar enzymes that catalyse similar reactions by different mechanisms.

REFERENCES

1 Jörnvall, H., Persson, M. and Jeffery, J. (1981) Proc. Natl. Acad. Sci. U.S.A. **78**, 4226–4230
2 Horjales, E. and Brändén, C.-I. (1985) J. Biol. Chem. **260**, 15445–15451
3 Jeffery, J. and Jörnvall, H. (1988) Adv. Enzymol. Relat. Areas Mol. Biol. **61**, 47–106
4 Aronson, B. D., Somerville, R. L., Epperly, B. R. and Dekker, E. E. (1989) J. Biol. Chem. **264**, 5226–5232
5 Koivusalo, M., Baumann, M. and Uotila, L. (1989) FEBS Lett. **257**, 105–109
6 Eklund, H. (1989) Biochem. Soc. Trans. **17**, 293–296
7 Eklund, H., Horjales, E., Jörnvall, H., Brändén, C.-I. and Jeffery, J. (1985) Biochemistry **24**, 8005–8012
8 Persson, B., Krook, M. and Jörnvall, H. (1991) Eur. J. Biochem. **200**, 537–543
9 Ghosh, D., Weeks, C. M., Grochulski, P., Duax, W. L., Erman, M., Rimsay, R. L. and Orr, J. C. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 10064–10068
10 Bohren, K. M., Bullock, B., Wermuch, B. and Gabbay, K. H. (1989) J. Biol. Chem. **264**, 9547–9551
11 Willey, D. L., Caswell, D. A., Lowe, C. R. and Bruce, N. C. (1993) Biochem. J. **290**, 539–544
12 Bairoch, A. and Boeckmann, B. (1991) Nucleic Acids Res. **19**, 2247–2249
13 Coulson, A. F. W., Collins, J. F. and Lyall, A. (1987) Comput. J. **30**, 420–424
14 Kabsch, W. and Sander, C. (1983) Biopolymers **22**, 2577–2637
15 Wilson, D. K., Bohren, K. M., Gabbay, K. H. and Quiocho, F. A. (1992) Science **257**, 81–84
16 Garcia-Perez, A., Martin, B., Murphy, H. R., Uchida, S., Murer, H., Cowley, B. D., Handler, J. S. and Burg, M. B. (1989) J. Biol. Chem. **264**, 16815–16821
17 Nishimura, C., Wistow, G. and Carper, D. (1989) Enzymology and Molecular Biology of Carbonyl Metabolism 2 (Weiner, H. and Flynn, T. G., eds.), pp. 211–220, Alan R. Liss, New York
18 Schade, S. Z., Early, S. L., Williams, T. R., Kézdy, F. J., Heinrickson, R. L., Grimshaw, C. E. and Doughty, C. C. (1990) J. Biol. Chem. **265**, 3628–3635
19 Chung, S. and LaMendola, J. (1989) J. Biol. Chem. **264**, 14775–14777
20 Cheng, K. C., White, P. C. and Qin, K. N. (1991) Mol. Endocrinol. **5**, 823–828
21 Onishi, U., Noshiro, M., Shimosato, T. and Okuda, K. (1991) FEBS Lett. **283**, 215–218
22 Samaras, N. and Spithill, T. W. (1989) J. Biol. Chem. **264**, 4251–4254
23 Winters, C. J., Molowa, D. T. and Guzlian, P. S. (1990) Biochemistry **29**, 1080–1087
24 Watanabe, K., Fujii, Y., Nakayama, K., Ohkubo, H., Kuramitsu, S., Kagamiyama, H., Nakanishi, S. and Osamu, H. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 11–15
25 Welle, R., Schroeder, G., Schlitz, E., Grisebach, H. and Schroeder, J. (1991) Eur. J. Biochem. **196**, 423–430
26 Anderson, S., Marks, C. B., Lazarus, R., Miller, J., Stafford, K., Seymour, J., Light, D., Rastetter, W., Estell, D. (1985) Science **230**, 144–149
27 Grindley, J. G., Payton, M. A., De Pol, H. and Hardy, K. G. (1988) Appl. Environ. Microbiol. **54**, 1770–1775
28 Fujii, Y., Watanabe, K., Hayashi, H., Urade, Y., Kuramitsu, S., Kagamiyama, H. and Hayaishi, O. (1990) J. Biol. Chem. **265**, 9914–9923
29 Amore, R., Kötter, P., Küster, C., Ciracy, M. and Hollenberg, C. P. (1991) Gene **109**, 89–97
30 Oechsner, U., Magodolen, V. and Bandlow, W. (1988) FEBS Lett. **238**, 123–128
31 Pailhoux, E. A., Martinez, A., Veyssiere, E. M. and Jean, C. G. (1990) J. Biol. Chem. **265**, 19932–19936
32 Bartles, D., Engelhardt, K., Roncarate, R., Schneider, K., Rotter, M. and Salamini, F. (1991) EMBO J. **10**, 1037–1043
33 Smith, E. L. and Margoliash, E. (1964) Fed. Proc. Fed. Am. Soc. Exp. Biol. **23**, 1243–1247
34 Jörnvall, H., Persson, B. and Jeffery, J. (1987) Eur. J. Biochem. **167**, 195–201
35 Flynn, G. T. and Kubiseski, T. J. (1993). Adv. Enzyme Regul. **33**, 197–206
36 Bohren, K. M., Page, J. L., Shankar, R., Henry, S. P. and Gabbay, K. H. (1991) J. Biol. Chem. **266**, 24031–24037
37 Borhani, D. W., Harter, T. M. and Petrash, J. M. (1992) J. Biol. Chem. **267**, 24841–24847
38 Petrash, J. M., Harter, T. M., Devine, C. S., Olins, P. O., Bhatnagar, A., Liu, S.-Q. and Srivastava, S. K. (1992) **267**, 24833–24840
39 Bohren, K. M., Grinshaw, C. E. and Gabbay, K. H. (1992) J. Biol. Chem. **267**, 20965–20970
40 Farber, G. K. and Petsko, G. A. (1990) Trends Biochem. Sci. **15**, 228–234
41 Wilmanns, M., Hyde, C. C., Davies, D. R., Kirschner, K. and Jansonius, J. N. (1991) Biochemistry **30**, 9161–9169
42 Alber, T., Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D., Rivers, P. S. and Wilson, I. A. (1981) Philos. Trans. R. Soc. London B **293**, 159–171
43 Anderson, I., Knight, S., Schneider, G., Lindquist, Y., Lundquist, T., Brädnén, C.-I. and Lorimer, G. (1989) Nature (London) **337**, 229–234
44 Eklund, H., Nordström, B., Zeppezauer, E., Söderlund, G., Ohlsson, I., Boiwe, T., Söderberg, B.-O., Tapia, O., Brändén, C.-I. and Akeson A. (1976) J. Mol. Biol. **102**, 27–59
45 Pawlowski, J. E., Huizinga, M. and Penning, T. M. (1991) J. Biol. Chem. **266**, 8820–8825
46 Stolz, A., Rahimi-Kiani, M., Ameis, D., Chan, E., Ronk, M. and Shively, J. E. (1991) J. Biol. Chem. **266**, 15253–15257
47 Penning, T. M., Abrahams, W. R. and Pawlowski, J. E. (1991) J. Biol. Chem. **266**, 8826–8834

Received 5 August 1993/23 November 1993; accepted 1 December 1993