



Laphorn, A. J., Zhu, X., and Ellis, E. M. (2013) The diversity of microbial aldo/keto reductases from *Escherichia coli* K12. *Chemico-Biological Interactions*, 202(1-3), pp. 168-177. (doi:10.1016/j.cbi.2012.10.008)

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/124204/>

Deposited on: 07 October 2016

Enlighten – Research publications by members of the University of Glasgow
<http://eprints.gla.ac.uk>

The diversity of microbial aldo/keto reductases from *Escherichia coli* K12

Adrian J. Laphorn¹, Xiaofeng Zhu^{2,3} and Elizabeth M. Ellis³

1 School of Chemistry, Joseph Black Building, University of Glasgow, Glasgow G12 8QQ

2 College of Life Science and State Key Laboratory of Biotherapy and Cancer Centre, Sichuan University, Chengdu, China

3 Strathclyde Institute of Pharmacy and Biomedical Sciences, 161 Cathedral Street, Glasgow, G4 0RE

Corresponding author:

Adrian J. Laphorn
School of Chemistry,
Joseph Black Building,
University of Glasgow,
Glasgow G12 8QQ

Tel: +44 141-330 5940

Fax: +44 141-330 4888

E-mail: Adrian.Laphorn@glasgow.ac.uk

Key Words: Aldo-Keto reductases, methylglyoxal reductase, 2,5-diketo-D-gluconate reductase, tyrosine auxotrophy suppressor protein, L-glyceraldehyde 3-phosphate reductase, AKR quaternary structure.

Abstract

The genome of *Escherichia coli* K12 contains 9 open reading frames encoding aldo/keto reductases (AKR) that are differentially regulated and sequence diverse. A significant amount of data is available for the *E. coli* AKRs through the availability of gene knockouts and gene expression studies, which adds to the biochemical and kinetic data. This together with the availability of crystal structures for nearly half of the *E. coli* AKRs and homologues of several others provides an opportunity to look at the diversity of these representative bacterial AKRs. Based around the common AKR fold of $(\beta/\alpha)_8$ barrel with two additional α -helices, the *E. coli* AKRs have a loop structure that is more diverse than their mammalian counterparts, creating a variety of active site architectures. Nearly half of the AKRs are expected to be monomeric, but there are examples of dimeric, trimeric and octameric enzymes, as well as diversity in specificity for NAD as well as NADP as a cofactor. However in functional assignments and characterisation of enzyme activities there is a paucity of data when compared to the mammalian AKR enzymes.

1. Introduction

Aldo-keto reductases (AKRs) comprise a large superfamily of NADPH dependent oxidoreductases known to reduce a wide range of aromatic and aliphatic aldehydes and ketones in a stereospecific manner [1]. Although these reactions can also be catalysed by enzymes from the short-chain dehydrogenase/reductase (SDR) [2] and medium-chain dehydrogenase/reductase (MDR) [3] superfamilies, the AKR enzymes tend to favour aldehyde reduction leading to the formation of alcohols. There is no shortage of

potential substrates for the AKR enzymes, as aldehydes and ketones such as sugars and amino acids form key intermediates in cellular metabolic pathways. In addition oxidative stress results in the formation of peroxides and free radicals that damage components of the cell, including proteins, lipids, and DNA. In particular lipid peroxidation results in the consequent formation of the mutagenic aldehydes, malondialdehyde, β -hydroxyacrolein and 4-hydroxynonenal [4]. Microbial toxins, plant phytoalexins or manmade chemicals are yet another source of reactive aldehydes and ketones in the environment that need to be detoxified for cell survival.

AKRs are typically active against a broad range of potential substrates, which makes the task of unravelling the cellular functions of these enzymes all the more difficult. The basic folded structure of the protein, an $(\beta/\alpha)_8$ barrel with two additional α -helices H1 and H2, was first demonstrated for the archetypal mammalian AKR, human aldose reductase (AKR1B1) [5, 6] (Figure 1). The NADP(H) cofactor is found in an extended conformation, with the nicotinamide ring of the cofactor held in the central cavity of the barrel in close proximity to a catalytic tetrad of residues. Tyr 48 and His 110 residues, together with the nicotinamide ring form an anion hole in which the aldehyde or ketone oxygen binds to permit reduction. The role of Asp 50 and Lys 84 is proposed to facilitate proton donation by decreasing the pKa of the catalytic tyrosine (reviewed in [1]). The substrate specificity is modulated by amino acid residues at the edge of the active site, particularly those directly before Tyr 48 and after His110. In addition the active site cleft is formed from three loops which have become known as the A, B and C-terminal loops, each of which contribute residues (Figure 1). The early AKR structures described were monomeric, in contrast to the functionally related SDR and MDR superfamilies that form

dimers or tetramers. This gave rise to the view that AKRs are a monomeric class of proteins with the odd exception such as the tetrameric AKR6 family of proteins involved in the gating of voltage-sensitive potassium (K_v) channels [7] and dimeric AKR7 family proteins which are notably able to reduce 2-carboxybenzaldehyde [8].

A systematic nomenclature has been devised to divide individual AKRs into discrete families having >40% amino acid identity and then subfamilies having >60% sequence identity [9]. An AKR superfamily homepage [10] has been established as a database of existing and potential members of this diverse group of proteins, last updated in 2009 [10]. There are 15 families of enzymes, AKR1-AKR15, named in the sequential fashion as they are identified, with subfamilies denoted by a letter and each member having its own unique identifier. The majority of mammalian enzymes belong to the AKR1 family, with AKR6 and AKR7 being specific exceptions. The explosion in sequenced genomes [11] has highlighted the existence of AKR-related open reading frames (ORFs) in all branches of living organisms. There is also considerable sequence diversity of microbial enzymes, many of which have not been included in the AKR nomenclature because there is no data available for endogenous expression or function. A significant number of microbial AKRs have been researched and several are being exploited in biotransformation reactions (reviewed in [12]).

In this paper we will review the status of nine ORFs encoding putative AKR identified in the genome sequence of *Escherichia coli* K-12 MG1655 [13], and discuss their structure and function in the light of recent structural, biochemical and kinetic characterisation (Table 1). The ORF gene product name will be used (e.g. YdjG) as the main identifier

together with the systematic protein name if available (e.g. AKR11B2) and the gene locus ID which is a unique identifier for the *E. coli* K-12 MG1655 genome (e.g. B1771).

2. Classification of *E.coli* AKRs by Quaternary structure

2.1. Monomeric AKRs

As might be expected of the nine ORFs encoding AKRs in the *E.coli* genome, at least three of those characterised have been shown experimentally to be monomeric. The two proteins YqhE (B3012, AKR5C2) and YafB (B0207, AKR3F2) are related as both have been shown to catalyze the reduction of 2,5-diketo-D-gluconate (25DKG) to 2-keto-L-gluconate [14], an important step in the commercial biosynthesis of ascorbic acid and commonly referred to as DkgA and DkgB respectively. YdjG (B1771, AKR11B2) has been purified and characterised as a monomeric AKR with activity against methylglyoxal and a requirement for NADH, rather than NADPH as co-factor [15]. Of the other six *E.coli* AKR ORFs it is likely that a number are also monomeric such as YeaE (B1781), by sequence similarity to YqhE and YafB, however there are no experimental data to confirm this and in some cases misinterpretation of results has led to incorrect assignments as is the case of YgdS described later.

2.1.1 25DKG reductases AKR5C2 (YqhE, B3012) and AKR3F2 (YafB, B0207)

From protein based sequence searches using pBlast program on a database of non-redundant sequences at the National Center for Biotechnology Information (NCBI) [16], it is clear that the majority of Enterobacteriaceae seem to have both YqhE and YafB related sequences, while other bacterial families do not. It is striking that YqhE

sequences are also found in α proteobacteria, high GC gram positive bacteria and firmicutes, while YafB like sequences are found in β and χ proteobacteria (Figure 2). Looking at individual bacterial species from these groups it becomes clear that most contain two or more 25DKG reductase-like sequences which share significant sequence similarity. However, these sequences are either similar to YafB or YqhE, suggestive of gene duplication events within these bacteria, while enterobacteria must have acquired a second copy by horizontal gene transfer.

YqhE (B3012, DkgA) is a 275 amino acid residue protein (AKR5C2) which shows the highest activity towards a range of aldehydes and ketones of all the *E. coli* AKRs that have been kinetically characterised [17]. It has also been shown to catalyse the stereoselective reduction of β -keto ester ethyl acetoacetate and other 2-substituted derivatives [18].

YafB (B0207, DkgB) is a 267 amino acid protein sharing only 39% sequence identity (60% sequence similarity) to YqhE and thus belonging to a different AKR family of which it is member AKR3F2. Despite this, YafB has a similar activity profile to YqhE, having on average ~ 65% lower specific activities on aldehydes and ketones [17].

Structurally both enzymes are very similar to the first characterised prokaryote AKR, 2,5-diketo-D-gluconic acid reductase from *Corynebacterium sp.*, in complex with NADP⁺ (PDB: 1A80) [19], which is effectively a YqhE homologue. The apo structure of YqhE from *E. coli* has been determined [20] (PDB 1MZR) and more recently an apo structure of a YafB homolog (with 44% amino acid sequence identity) from

Sinorhizobium meliloti 1021(PDB 3UP8) from the New York Structural Genomics Research Consortium (Figure 3). The structures are very similar and can be superimposed with a root mean squared difference (RMSD) of 1.52Å over 262 residues. The folded structure is characterised by particularly short loops which effectively removes the significant active site cavity generated by loops A and C in aldose reductase (Fig. 1). The functional implication of a short loop B, which is involved in binding of the co-factor, is a 3-fold less tight binding of cofactor compared with human aldose reductase [19]. The C-terminal loop is a major difference between the two 25DKG reductases, being between 14-17 residues long, contributing in the case of YqhE, Pro 269 and Phe 274 to the active site pocket and Val264, Pro266 and Leu 269 in the case of YafB. The C-terminus has a significant effect on the activity of the *Corynebacterium sp.* enzyme, deletion of the last eight amino acids via mutagenesis results in a 40-fold decrease in Vmax [19], which is similar to that seen for a deletion of the C-terminus in aldose reductase [21]. The C-terminus does not affect cofactor binding but rather seems to play a crucial role in substrate orientation within the active site. Therefore the different residues contributed in YqhE and YafB are probably responsible for differences in activity against D-xylose (Table 1) for example [17].

Both the apo structures when compared the *Corynebacterium sp.* holo structure show some conformational changes in Loop 1 and the adjacent Loop 8 which are both involved in the binding of the NADP cofactor. In particular the aromatic amino acid residue in Loop 1 that packs against the nicotinamide ring of NADP adopts a conformation which occupies the cleft unoccupied by the cofactor. Both structures

however resemble the holo enzyme and not the high resolution *Corynebacterium sp* apo structure (PDB: 1HW6) where large scale changes and disorder are present in the first 4 loops, the active site and the C-terminus [22], which in the light of these structures may be considered an artefact of crystallization.

2.1.2 A NADH specific AKR, AKR11B2 (YdjG, B1771)

YdjG (B1771) is a 326 amino acid protein which is a member of the AKR11 family of proteins, member AKR11B2. YdjG related sequences are not widely found in enterobacteria, as other *E.coli* AKRs, but widely in α proteobacteria, CFB group and firmicutes (Fig. 2). It has been characterised as a monomeric AKR [17], with enzyme active against methylglyoxal and DL-glyceraldehyde but not 4-nitrobenzaldehyde. From comparison of kinetic parameters it would seem to have enzyme activity comparable to YghZ or YeaE for methylglyoxal (Table 1) [17]. What is unusual is the specificity of YdjG for NAD(H) as a cofactor, which represents the first example of an AKR which is not able to use NADP(H). The *ydjG* gene is located in an uncharacterised operon which includes a putative sugar kinase (*ydjH*), aldolase (*ydjI*) and a major facilitator superfamily transporter (*ydjK*) [23] (Figure 4a). These types of enzyme are consistent with a sugar metabolism pathway (exemplified by rhamnose degradation) with the AKR called upon to reduce the aldehyde produced by the aldolase.

There are a number of related AKR11 family structures determined, the two most similar in sequence to YdjG are from *Bacillus subtilis*, YhdN, AKR11B1 (PDB: 1PZ1) which

shares 39% sequence identity (59% similarity), a conventional AKR with significant activity against a range of aldehydes and ketones [24], and IoIS, AKR11A1 (PDB: 1PZ0), 31% sequence identity (54% similarity) which has limited activity against erythrose and D/L glyceraldehyde and methyl glyoxal (which also irreversibly inactivated the enzyme) [24]. Both enzymes use NADPH as a cofactor and have a similar fold, but while AKR11B1 has a regular 19 residue C-terminal loop which contributes to the active site, AKR11A1 has a short 3 residue loop which does not. Based on the AKR11B1 structure (1PZ1) we can see that the fold of YdjG would have notably a long 15 residue loop 1, an unusually long loop 3 (24 residues, only 14 residue shown in Fig. 4b), a long Loop B (~32 residues) and effectively no C-terminal loop (Fig. 4b). It seems likely that the β -hairpin like structure of loop 3 is extended as is seen partially in the glycerol dehydrogenase AKR11B4 structure from *Gluconobacter oxydans* (PDB: 3N2T)[25].

Given the limited kinetic characterisation of YdjG it is not really possible to assess the presumed absence of a C-terminal loop contributing to the active site. The highly unusual preference for NAD(H) over NADP(H) of YdjG is readily explained in relation to the mutation of Asp 232 to Ala which is then able to use both co-factors [15]. This residue is in loop B, far from the 2' phosphate position where it would need to act. What is clearly more important is the role of positively charged residues to bind the phosphate group, which are contributed by loop 8 and loop B in both *B. subtilis* and the *G. oxydans* structures but are absent in the YdjG sequence. There are two critical residues, namely Thr228 and Thr293 in the YdjG sequence which are either lysine or arginine in the sequence of proteins that should bind NADP(H). Looking specifically for these residues

in homologous sequences to YdjG shows that there is a fairly even distribution of sequences, suggesting that both NAD(H) and NADP(H) specific AKRs are found in this group.

2.2. A Dimeric AKR, YgdS (B2834, TAS protein)

The *ygdS* gene was identified by its ability to permit slow growth of the tyrosine auxotroph strain *E. coli* WU3610 under tyrosine starvation conditions [26], and is referred to *tas* (tyrosine auxotrophy suppressor). In a significant number of experiments it was clear that *ygdS* was not the target of mutation under tyrosine starvation nor was it able to complement a related phenylalanine auxotroph strain. When *ygdS* was cloned into a lower-copy-number vector and introduced into *E. coli* WU3610, it allowed a faster growth on minimal media plates lacking tyrosine. It is therefore likely, that *ygdS* directly substitutes for the prephenate dehydrogenase activity which is absent in the Tyr⁻ auxotroph strain [26].

YgdS (B2834) is a 346 amino acid residue protein and based on a blast search of non-redundant sequences, YgdS related sequences are found widely in enterobacteria, β and χ proteobacteria, CFB group, cyanobacteria, and to a lesser extent in green algae and some plants. The problem of whether an amino acid sequence was similar or not was based on the conservation of sequence flanking the active site residues and the length of the loop sequences.

The crystal structure of *E. coli* YgdS has been reported in complex with NADP(H) (PDB: 1LQA) at 1.6Å resolution, to assist with the functional assignment of the protein [27]. However, analysis of the crystal structure and X-ray data shows that this AKR is dimeric a fact ignored in the structure report [27], and contrary to the monomeric assignment by the authors in the PDB entry. Refinement and correction of the structure using the deposited X-ray data permitted an improvement in the R and R-free to 0.138 and 0.167 respectively with the addition of several glycerol molecules, chloride and Mg²⁺ ions. This includes one glycerol in the active site cleft which gives an indication of substrate binding (Figure 5). The tertiary structure is very different from that of the 25DKG reductases, with long loops 1-4, a long (32 residue) loop B, and a short 4 residue C-terminus. The loop B is more reminiscent of that found in AKR7 family enzymes, playing a significant role in the tight binding of co-factor. The short C-terminus does not interact directly with the active site, as is the case in other active AKRs. However, the terminal proline residue forms a salt bridge with an arginine in loop B and the R-group packs against loop 4 which contributes an invariant phenylalanine residue (Phe 141) into the active site in a position analogous to that of C-terminal loops in other AKR structures. The active site cleft is narrower and deeper in YgdS when compared to other AKRs, due to the length of loops 2, 3, and 4 and loop B.

In forming a stable dimer, YgdS buries 1716 Å² of solvent accessible surface area, ~12% of the total solvent exposed surface of the protein (calculated using the program PISA [28]). This is the third reported class of dimeric AKR, after rat aflatoxin reductase, AKR7A1[8] and *Candida tenui* xylose reductase, AKR2B5 [29]. The YgdS dimer formed

is structurally most similar to that of AKR7A1 with a rather slender interface ($\sim 20\text{\AA}$ x $\sim 35\text{\AA}$) involving almost exclusively helices $\alpha 5$ and $\alpha 6$ with 8 hydrogen bonds and a salt bridge per subunit. However in the Tas protein a long loop (22 residues) between strand $\beta 4$ and helix $\alpha 4$ is also involved, forming a sequence-conserved intermolecular salt bridge interaction between the invariant residues Arg 136 and Glu 226. The dimer interface helps stabilize this loop which contributes Phe 141 to the active site of the enzyme. Unfortunately, at present there is no experimental evidence of AKR activity for this protein although the catalytic tetrad is present and the active site is sufficiently large to accommodate a range of aldehydes and ketones.

2.3 A Trimeric AKR, *YdhF* (B1647)

YdhF is a 298 amino acid protein which, based on sequence comparison, is found widely in enterobacteria, α and χ proteobacteria, CFB group bacteria and firmicutes (Fig. 2.). This hypothetical AKR has been a successful target for a Marseille-based structural genomics consortium [30]. The structures of both the apo (PDB 1UR3) and holo (PDB 1OG6) with NADP⁺ enzymes has been determined to 2.57 \AA and 2.8 \AA resolution respectively. The active site catalytic tetrad is present with a defined anion hole for the binding of an aldehyde or ketone. Unfortunately there are no kinetic data available to confirm that it is a functioning aldo/keto reductase or to provide some insight into the substrate specificity of this enzyme. Analysis of the protein monomer reveals a standard AKR fold with a long loop 3 (16 residues) and a short Loop B (~ 20 residues) and C-terminal loop (5 residues) which points away from the active site of the monomer. What makes this AKR structure unique is that it is the first reported trimeric

AKR, the C-terminus of one monomer contributing to the active site of a neighbouring monomer (Figure 6.). Analysis of the interface using the program PISA [28], shows that $\sim 5,900 \text{ \AA}^2$ of solvent accessible surface area is buried by trimer formation, $\sim 15\%$ of total protein surface area of the protein. The interface involves loops 3 and 4 and α -helix 5 on one subunit and α -helix 6 and α -helix H2 together with the C-terminal loop on the other, each burying $\sim 980 \text{ \AA}^2$ of solvent accessible surface area. There are 7 hydrogen bonds formed between monomers, and two salt bridges, the conserved Asp 131 to Arg 282 and the less conserved Asp 137 to Arg 203, which are located at opposite ends of the interface. Trimeric, α - β barrel proteins are known, for example KDPG aldolases, however they have a different mode of association compared with YdhF. It is of note that in the YdhF trimer, all active sites are present on one side of the oligomer, leaving a side that could act as a recognition site for other protein/proteins. In this way it is reminiscent of the tetrameric AKR6 family members that have been found to associate with, and control the function of, mammalian voltage-sensitive potassium (K_v) channels [31, 32]. A 6 \AA radius channel at the centre of the trimer is slightly larger than the $\sim 4 \text{ \AA}$ radius channel found in the AKR6A2 (PDB: 1EXB). Against this hypothesis, no binding partners have been identified from two separate pull-down experiments using his-tagged protein [33, 34]. Also the presence of well defined active site architecture as found in other AKRs, suggests a broad range of substrates which would not be consistent with YdhF playing a role in the gating of a potential protein channel.

2.4. An Octameric AKR, AKR14A1 (*YghZ*, B3001)

Based on sequence similarity, YghZ-like ORFs are found in enterobacteria, α , β and χ proteobacteria and high GC content gram positive bacteria and form a distinct grouping of AKRs together with YajO (Fig. 2.). YghZ is a 346 amino acid protein first cloned and characterised due to its sequence similarity to the unusual mammalian AKR6 and AKR7 family enzymes [35]. The enzyme was shown to be active against a range of substrates but with highest activity against methylglyoxal and diketones such as isatin. Over expression of the recombinant gene in *E. coli* leads to increased resistance to methylglyoxal [35]. Ko and co workers [17] independently showed that while YafB, YqhE and YeaE have higher specific activities for methylglyoxal than YghZ (Table 1). However a $\Delta yghZ$ and $\Delta gloA$ (glyoxylase A) double mutant leads to reduced cell viability in response to methylglyoxal. More recently, over expression of *yghZ* has been shown to complement (sluggish growth) for *tpiA*⁻ (triosephosphate isomerase deficient) strain of *E. coli* which is unable to produce dihydroxyacetone phosphate (DHAP) during growth on M9 minimal medium containing L-lactate [36]. The authors show that YghZ is able to selectively reduce L-glyceraldehyde 3-phosphate (L-GAP) to glycerol 3-phosphate (L-G3P), an activity identified previously by Tropp and co-workers [37]. The product L-G3P can be oxidised by glycerol 3-phosphate dehydrogenase to DHAP which is the route glycerol is metabolised. However, there is no known enzymatic route to L-GAP which is toxic to *E. coli* cells [38], and the accumulated metabolite D- glyceraldehyde 3-phosphate will chemically break down to methylglyoxal and phosphate via a beta elimination [39].

Recently the crystal structure of the apo form of YghZ has been solved [40] (PDB: 3N6Q) and the apo and tertiary complex (NADP⁺ and citrate) by our own group (PDB:

4AST and 4AUB) at 2.38Å and 2.05Å resolution respectively. As might be expected from sequence similarity, YghZ is structurally most similar to AKR6 family of proteins. Analysis of the protein monomer reveals a standard AKR fold with an unusual 12 residue N-terminus loop preceding the β -hairpin that starts almost all AKR structures, a long loop 3 (13 residues) and a long loop B (~38 residues) and in contrast to AKR6 proteins a normal length C-terminus (13 residues) which loops into the active site of the monomer (Figure 7).

YghZ forms a structurally equivalent tetramer to the AKR6 family, the monomers are arranged end-to-side, in which the N-terminal β strands and the β -hairpin at the bottom of the α/β barrel of one monomer packed against the α -helices α 3 and α 4 and Loop 3 of an adjacent monomer. Between two monomers the interface involves 12 hydrogen bonds and two salt bridges which buries ~1800 Å² of solvent accessible surface area of each monomer. The octamer (Fig. 7) is formed by a much smaller interface involving the first four residues of the novel N-terminus and residues from loops 1 and 2 burying only ~600 Å² of solvent accessible surface area of each monomer. Although 600 Å² is quite small for a stable interface, we have confirmed by gel filtration and electron microscopy that an octamer is formed.

2.5 Structurally uncharacterised *E.coli* AKRs

Where an *E.coli* AKR structure has not been determined, a suitably homologous structure has been available, as in the case of both YafB and YdjG. It is clear in the cases of YajO, YdbC and YeaE that they belong to different families of AKR than those

characterised so far and therefore cannot be described as the previous six *E. coli* AKRs. Modelling can give some basic insight into the kind of structure that we might expect and two out of the remaining three AKRs have been kinetically characterised.

2.5.1 *YeaE* (B1781)

YeaE (B1781) is a 284 amino acid protein with sequence similarity to ORFs from enterobacteria, α -, β - and χ proteobacteria and firmicutes. The gene has been over expressed and the protein characterised as having significant activity against 3- and 4-nitrobenzaldehyde, and the oxaldehydes, methylglyoxal and phenylglyoxal with NADPH as a cofactor [17]. Growth of a *yeaE* and glyoxylase A (*gloA*) double mutant is inhibited by 0.3 mM methylglyoxal. Despite sharing ~30 sequence identity (50% similarity) to *YqhE* or *YafB*, the protein has no appreciable activity against 25DKG or sugars other than weak activity for D/L glyceraldehyde [17]. The structure of *YeaE* is predicted to be similar to *YqhE* and *YafB*, with short loops 1-4, a short loop B and a 13 residue C-terminal loop. The C-terminus is not sequence similar to any other AKR structures but based on the amino acid sequence conservation amongst *YeaE* related sequences it is likely that Leu 281 is the residue contributed to the active site.

2.5.2 *YajO* (B0419)

YajO (B0419) has high sequence homology to ORFs found in a wide range of bacteria from enterobacteria, various proteobacteria groups, high GC gram positive bacteria to cyanobacteria (Fig. 2). The 324 amino acid protein has been shown to be

NADPH dependent with rather low enzyme activity against a range of aldehydes and ketones [17], notably only marginal activity against methylglyoxal. However of the AKRs tested YajO shows high activity against 2-carboxy benzaldehyde, which was not a substrate for the other 4 AKRs in that study. This is a property shared with the mammalian AKR7 family of proteins [41], which have a carboxylate binding pocket within the active site [8] to permit the reduction of acid aldehydes such as succinic semialdehyde. Structurally, YajO is most similar to AKR14A1 (YghZ) and AKR11 proteins with long loop 1 and 3, a long loop B and a C-terminal loop that contributes to the active site. It is peculiar in having a very long loop 1 (~19 residues), which must adopt a slightly larger loop structure similar to that seen in AKR11 proteins (figure 4b). The aggregation state of the protein has not been characterised and in terms of sequence is in a grey area between monomeric AKR11 family proteins and tetrameric AKR6 like proteins.

2.5.3 *YdbC* (B1406)

YdbC (B1406) is a 286 amino acid protein annotated as an AKR based on sequence homology alone and has not been kinetically characterised to date. Related sequences are found in a wide range of bacteria from enterobacteria, α - β - and χ proteobacteria and high GC gram positive bacteria. Of all the AKRs described the $\Delta ydbC$ knockout is the only one which has an observed phenotype, in that it is not able to respire when grown on acetate [42]. Structurally, *YdbC* shows some similarity to AKR11 proteins with a long loop 1 and 3, however it has a short loop B as is seen in the 25DKG reductases and no significant C-terminal loop. From modelling [43], It would seem the protein has

retained all the necessary residues to bind co-factor, which is probably NADPH although the 2' phosphate recognition is not entirely standard. However, it seems that the protein would lack an active site histidine which would be replaced with an arginine. It is interesting that the histidine should be missing from the catalytic tetrad as this is replaced by Asn in the AKR6 family and Glu in the AKR1D family. Both these unusual classes of AKR have been shown to be active; AKR6 proteins turn over a limited range of substrates and regulate inactivation and membrane localization of Kv1 and Kv4 channels [44], while AKR1D are Δ^4 -3-ketosteroid-5 β -reductases which perform the selective reduction of the steroid 4-5 double bond instead of the standard ketone reduction [45]. Therefore YdbC is a rather intriguing AKR which may be inactive or alternatively perform a different chemical reaction.

3. Conclusions

The *E.coli* AKRs based on sequence comparison are overall related to other enterobacteria, α , β and χ proteobacteria, high GC content gram positive bacteria, CFB group bacteria and firmicutes and are therefore representative of bacterial AKRs. This hides the fact that the distribution of individual AKRs is in many cases very group- and species-specific, and that the nine *E.coli* enzymes do not represent the full diversity of bacterial AKRs. Based on the rules established for the AKR super family [9], where families are distinguished by having >40% amino acid identity and subfamilies have >60%, each of the *E.coli* AKRs would be a member of an individual family of AKRs, although only four have been assigned.

E. coli AKRs provide ample evidence of the structural diversity of microbial AKRs. The 25DKG reductases YqhE and YafB, have the shortest loops 1-4 and loop B of the AKRs described while retaining a regular microbial C-terminal loop and display high activity to a range of aldehydes and ketones. This minimal AKR fold shares significant sequence similarity with mammalian AKR1 family proteins and could be considered a typical microbial AKR structure. However, other very different structures are found in *E.coli*, for example the YgdS structure has long loops 1-4, a long loop B and a short C-terminal loop, creating a very different active site cavity. It is the first AKR to have the normal C-terminal residues from the active site contributed by loop 4, however we have no idea what effect this has on the enzyme activity as it has not been kinetically characterised. While the majority of the *E.coli* AKRs are probably monomeric, examples of dimeric (YgdS), trimeric (YdhF) and octameric (YghZ) forms show that novel quaternary structures are possible given the basic AKR fold. In the case of YgdS, the dimer interface stabilizes the conformation of loop 4 which contributes to the active site instead of the C-terminus. While in the trimeric AKR YdhF, the C-terminus of symmetry-related monomer contributes to the active site. In both cases the quaternary structure has permitted an unusual active site architecture, the effect on substrate specificity and ultimately the function of these proteins have not been investigated.

The AKR superfamily has been shown to require NADP(H) over NAD(H) as co-factor, with dual specificity relatively rare (reviewed in [15]). A triple mutant of 25DKG reductase from *Corynebacterium sp.* engineered to have activity with NAD still retains significant activity with NADP [46]. Therefore it is surprising to find one *E.coli* AKR, YgdS is specific for NAD(H) [15] a first for the AKR superfamily. The specificity of the

enzyme could be altered by site-specific mutation of a residue in loop B, confirming its importance in cofactor binding [15].

The function within the organism of the nine *E.coli* AKRs is far from clear (Table 1.). The genes have been systematically knocked out as part of larger systematic studies [42, 47], none are essential, and only $\Delta ydbC$ displaying a discernable phenotype [29]. The difficulty of this approach is illustrated in *Saccharomyces cerevisiae*, where only a particular triple knockout of AKR genes, resulted in a glucose-dependent heat shock phenotype which could be rescued by expression of human aldose reductase [48]. Similarly it was necessary to use a glyoxylase-deficient strain to show the relative sensitivity of *E.coli* AKR knockouts to methylglyoxal [17]. Only two genes, *yghZ* and *ygdS* have been identified under specific growth conditions as complementing for *tpiA* and *tyrA14(oc)* mutants of *E. coli* respectively [26, 36]. In neither case is the mechanism by which they complement the deficiency satisfactorily resolved. So rather like the assignment of a number of AKRs as methylglyoxal reductases and 25DKG reductases from a given enzyme activity (not always the highest), the precise function(s) of these genes remains unresolved.

Overall the the study of *E. coli* AKRs extends our understanding of the diversity in form and aggregation state which is possible based around the core AKR fold. However, where we have highly unusual structural information, namely in the cases of YdhF and YgdS we have no kinetic data available. Equally, where we have an AKR with a defined phenotype and an unusual sequence (YdbC), no characterisation has been carried out to date. Clearly there is scope to fill these gaps in our understanding and hopefully better understand the functions performed by these diverse AKRs.

References.

- [1] J.M. Jez, M.J. Bennett, B.P. Schlegel, M. Lewis, T.M. Penning, Comparative anatomy of the aldo-keto reductase superfamily, *Biochem J* 326 (Pt 3) (1997) 625-636.
- [2] K. Kavanagh, H. Jornvall, B. Persson, U. Oppermann, The SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes, *Cell Mol Life Sci* 65(24) (2008) 3895-3906.
- [3] B. Persson, J. Hedlund, H. Jornvall, Medium- and short-chain dehydrogenase/reductase gene and protein families : the MDR superfamily, *Cell Mol Life Sci* 65(24) (2008) 3879-3894.
- [4] H. Esterbauer, R.J. Schaur, H. Zollner, Chemistry and biochemistry of 4-hydroxynonenal, malonaldehyde and related aldehydes, *Free radical biology & medicine* 11(1) (1991) 81-128.
- [5] J.M. Rondeau, F. Tete-Favier, A. Podjarny, J.M. Reymann, P. Barth, J.F. Biellmann, D. Moras, Novel NADPH-binding domain revealed by the crystal structure of aldose reductase, *Nature* 355(6359) (1992) 469-472.
- [6] D.K. Wilson, K.M. Bohren, K.H. Gabbay, F.A. Quiocho, An unlikely sugar substrate site in the 1.65 Å structure of the human aldose reductase holoenzyme implicated in diabetic complications, *Science* 257(5066) (1992) 81-84.
- [7] J.M. Gulbis, S. Mann, R. MacKinnon, Structure of a voltage-dependent K⁺ channel beta subunit, *Cell* 97(7) (1999) 943-952.
- [8] E. Kozma, E. Brown, E.M. Ellis, A.J. Laphorn, The crystal structure of rat liver AKR7A1. A dimeric member of the aldo-keto reductase superfamily, *The Journal of biological chemistry* 277(18) (2002) 16285-16293.
- [9] J.M. Jez, T.G. Flynn, T.M. Penning, A new nomenclature for the aldo-keto reductase superfamily, *Biochemical pharmacology* 54(6) (1997) 639-647.
- [10] D. Hyndman, D.R. Bauman, V.V. Heredia, T.M. Penning, The aldo-keto reductase superfamily homepage, *Chemico-biological interactions* 143-144 (2003) 621-631.
- [11] K. Liolios, K. Mavromatis, N. Tavernarakis, N.C. Kyrpides, The Genomes On Line Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata, *Nucleic acids research* 36 (2008) D475-D479.
- [12] E.M. Ellis, Microbial aldo-keto reductases, *FEMS microbiology letters* 216(2) (2002) 123-131.

- [13] F.R. Blattner, G. Plunkett, 3rd, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. Collado-Vides, J.D. Glasner, C.K. Rode, G.F. Mayhew, J. Gregor, N.W. Davis, H.A. Kirkpatrick, M.A. Goeden, D.J. Rose, B. Mau, Y. Shao, The complete genome sequence of *Escherichia coli* K-12, *Science* 277(5331) (1997) 1453-1462.
- [14] D.Y. Yum, B.Y. Lee, J.G. Pan, Identification of the *yqhE* and *yafB* genes encoding two 2, 5-diketo-D-gluconate reductases in *Escherichia coli*, *Appl Environ Microbiol* 65(8) (1999) 3341-3346.
- [15] E. Di Luccio, R.A. Elling, D.K. Wilson, Identification of a novel NADH-specific aldo-keto reductase using sequence and structural homologies, *Biochemical Journal* 400 (2006) 105-114.
- [16] E.W. Sayers, T. Barrett, D.A. Benson, S.H. Bryant, K. Canese, V. Chetvernin, D.M. Church, M. DiCuccio, R. Edgar, S. Federhen, M. Feolo, L.Y. Geer, W. Helmberg, Y. Kapustin, D. Landsman, D.J. Lipman, T.L. Madden, D.R. Maglott, V. Miller, I. Mizrachi, J. Ostell, K.D. Pruitt, G.D. Schuler, E. Sequeira, S.T. Sherry, M. Shumway, K. Sirotkin, A. Souvorov, G. Starchenko, T.A. Tatusova, L. Wagner, E. Yaschenko, J. Ye, Database resources of the National Center for Biotechnology Information, *Nucleic acids research* 37(Database issue) (2009) D5-15.
- [17] J. Ko, I. Kim, S. Yoo, B. Min, K. Kim, C. Park, Conversion of methylglyoxal to acetol by *Escherichia coli* aldo-keto reductases, *J Bacteriol* 187(16) (2005) 5782-5789.
- [18] M. Habrych, S. Rodriguez, T.D. Stewart, Purification and identification of an *Escherichia coli* beta-keto ester reductase as 2,5-diketo-D-gluconate reductase *YqhE*, *Biotechnol Progr* 18(2) (2002) 257-261.
- [19] S. Khurana, D.B. Powers, S. Anderson, M. Blaber, Crystal structure of 2,5-diketo-D-gluconic acid reductase A complexed with NADPH at 2.1-Å resolution, *Proceedings of the National Academy of Sciences of the United States of America* 95(12) (1998) 6768-6773.
- [20] S. Jeudy, V. Monchois, C. Maza, J.M. Claverie, C. Abergel, Crystal structure of *Escherichia coli* *DkgA*, a broad-specificity aldo-keto reductase, *Proteins* 62(1) (2006) 302-307.
- [21] K.M. Bohren, C.E. Grimshaw, K.H. Gabbay, Catalytic effectiveness of human aldose reductase. Critical role of C-terminal domain, *The Journal of biological chemistry* 267(29) (1992) 20965-20970.

- [22] G. Sanli, M. Blaber, Structural assembly of the active site in an aldo-keto reductase by NADPH cofactor, *Journal of molecular biology* 309(5) (2001) 1209-1218.
- [23] B.K. McIntosh, D.P. Renfro, G.S. Knapp, C.R. Lairikyengbam, N.M. Liles, L. Niu, A.M. Supak, A. Venkatraman, A.E. Zweifel, D.A. Siegele, J.C. Hu, EcoliWiki: a wiki-based community resource for *Escherichia coli*, *Nucleic acids research* 40(Database issue) (2012) D1270-1277.
- [24] A.H. Ehrensberger, D.K. Wilson, Structural and catalytic diversity in the two family 11 aldo-keto reductases, *Journal of molecular biology* 337(3) (2004) 661-673.
- [25] N. Richter, K. Breicha, W. Hummel, K. Niefind, The three-dimensional structure of AKR11B4, a glycerol dehydrogenase from *Gluconobacter oxydans*, reveals a tryptophan residue as an accelerator of reaction turnover., *Journal of molecular biology* 404(3) (2010) 353-362.
- [26] A.R. Timms, B.A. Bridges, Reversion of the tyrosine ochre strain *Escherichia coli* WU3610 under starvation conditions depends on a new gene *tas*, *Genetics* 148(4) (1998) 1627-1635.
- [27] G. Obmolova, A. Teplyakov, P.P. Khil, A.J. Howard, R.D. Camerini-Otero, G.L. Gilliland, Crystal structure of the *Escherichia coli* *Tas* protein, an NADP(H)-dependent aldo-keto reductase, *Proteins* 53(2) (2003) 323-325.
- [28] E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state, *Journal of molecular biology* 372(3) (2007) 774-797.
- [29] K.L. Kavanagh, M. Klimacek, B. Nidetzky, D.K. Wilson, The structure of apo and holo forms of xylose reductase, a dimeric aldo-keto reductase from *Candida tenuis*, *Biochemistry* 41(28) (2002) 8785-8795.
- [30] C. Abergel, B. Coutard, D. Byrne, S. Chenivresse, J.B. Claude, C. Deregnaucourt, T. Fricaux, C. Ganesini-Boutreux, S. Jeudy, R. Lebrun, C. Maza, C. Notredame, O. Poirot, K. Suhre, M. Varagnol, J.M. Claverie, Structural genomics of highly conserved microbial genes of unknown function in search of new antibacterial targets, *Journal of structural and functional genomics* 4(2-3) (2003) 141-157.
- [31] S.B. Long, E.B. Campbell, R. Mackinnon, Crystal structure of a mammalian voltage-dependent Shaker family K⁺ channel, *Science* 309(5736) (2005) 897-903.
- [32] S.B. Long, E.B. Campbell, R. Mackinnon, Voltage sensor of Kv1.2: structural basis of electromechanical coupling, *Science* 309(5736) (2005) 903-908.

- [33] M. Arifuzzaman, M. Maeda, A. Itoh, K. Nishikata, C. Takita, R. Saito, T. Ara, K. Nakahigashi, H.C. Huang, A. Hirai, K. Tsuzuki, S. Nakamura, M. Altaf-UI-Amin, T. Oshima, T. Baba, N. Yamamoto, T. Kawamura, T. Ioka-Nakamichi, M. Kitagawa, M. Tomita, S. Kanaya, C. Wada, H. Mori, Large-scale identification of protein-protein interaction of *Escherichia coli* K-12, *Genome research* 16(5) (2006) 686-691.
- [34] P. Hu, S.C. Janga, M. Babu, J.J. Diaz-Mejia, G. Butland, W. Yang, O. Pogoutse, X. Guo, S. Phanse, P. Wong, S. Chandran, C. Christopoulos, A. Nazarians-Armavil, N.K. Nasserli, G. Musso, M. Ali, N. Nazemof, V. Eroukova, A. Golshani, A. Paccanaro, J.F. Greenblatt, G. Moreno-Hagelsieb, A. Emili, Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins, *PLoS biology* 7(4) (2009) e96.
- [35] A.W. Grant, G. Steel, H. Waugh, E.M. Ellis, A novel aldo-keto reductase from *Escherichia coli* can increase resistance to methylglyoxal toxicity, *FEMS microbiology letters* 218(1) (2003) 93-99.
- [36] K.K. Desai, B.G. Miller, A metabolic bypass of the triosephosphate isomerase reaction, *Biochemistry* 47(31) (2008) 7983-7985.
- [37] M.K. Kalyananda, R. Engel, B.E. Tropp, Metabolism of L-glyceraldehyde 3-phosphate in *Escherichia coli*, *J Bacteriol* 169(6) (1987) 2488-2493.
- [38] C.T. Tang, R. Engel, B.E. Tropp, L-Glyceraldehyde 3-phosphate, a bactericidal agent, *Antimicrobial agents and chemotherapy* 11(1) (1977) 147-153.
- [39] J.P. Richard, Restoring a metabolic pathway, *ACS chemical biology* 3(10) (2008) 605-607.
- [40] M. Totir, N. Echols, M. Nanao, C.L. Gee, A. Moskaleva, S. Gradia, A.T. Iavarone, J.M. Berger, A.P. May, C. Zubieta, T. Alber, Macro-to-micro structural proteomics: native source proteins for high-throughput crystallization, *PloS one* 7(2) (2012) e32498.
- [41] L.S. Ireland, D.J. Harrison, G.E. Neal, J.D. Hayes, Molecular cloning, expression and catalytic activity of a human AKR7 member of the aldo-keto reductase superfamily: evidence that the major 2-carboxybenzaldehyde reductase from human liver is a homologue of rat aflatoxin B₁ aldehyde reductase., *Biochem J* 332(Pt1) (1998) 21-34.
- [42] M. Ito, T. Baba, H. Mori, H. Mori, Functional analysis of 1440 *Escherichia coli* genes using the combination of knock-out library and phenotype microarrays, *Metab Eng* 7(4) (2005) 318-327.

- [43] A. Roy, A. Kucukural, Y. Zhang, I-TASSER: a unified platform for automated protein structure and function prediction, *Nature protocols* 5(4) (2010) 725-738.
- [44] S.M. Tipparaju, O.A. Barski, S. Srivastava, A. Bhatnagar, Catalytic Mechanism and Substrate Specificity of the β -Subunit of the Voltage-Gated Potassium Channel. , *Biochemistry* 47(34) (2008) 8840-8854.
- [45] L. Di Costanzo, J.E. Drury, T.M. Penning, D.W. Christianson, Crystal structure of human liver Delta4-3-ketosteroid 5 β -reductase (AKR1D1) and implications for substrate binding and catalysis, *The Journal of biological chemistry* 283(24) (2008) 16830-16839.
- [46] G. Sanli, S. Banta, S. Anderson, M. Blaber, Structural alteration of cofactor specificity in *Corynebacterium* 2,5-diketo-D-gluconic acid reductase, *Protein science : a publication of the Protein Society* 13(2) (2004) 504-512.
- [47] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K.A. Datsenko, M. Tomita, B.L. Wanner, H. Mori, Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection, *Molecular systems biology* 2 (2006) 1-11.
- [48] Q. Chang, T.A. Griest, T.M. Harter, J.M. Petrash, Functional studies of aldo-keto reductases in *Saccharomyces cerevisiae*, *Biochimica et biophysica acta* 1773(3) (2007) 321-329.
- [49] T.P. Flores, D.S. Moss, J.M. Thornton, An algorithm for automatically generating protein topology cartoons, *Protein engineering* 7(1) (1994) 31-37.

Table 1A summary of the biochemical data for the 9 *E. coli* AKR ORFs.

Protein	Genome locus	Systematic name	X-ray structures and homologues (pdb code)	Quaternary state	Co-factor specificity	Specific activity of selected substrates (nanomoles min ⁻¹ mg ⁻¹)		Protein description	Conditions causing up regulated gene expression
YqhE	B3012	AKR5C2	Apo enzyme (1MZR)	Monomer	NADPH	4-NBA DL-GA	10,300 ± 600 2210 ± 10	25DKG reductase, beta-keto ester reductase	Stationary phase
YafB	B0207	AKR3F2	Apo enzyme homologue (3UP8)	Monomer	NADPH	25DKG MG D-Xylose	51.9 ± 1.5 21,000 ± 400 94.9 ± 9.7	25DKG reductase	Cold stress. Exposure to model apple juice
YdjG	B1771	AKR11B2	Holo enzyme homologues (1PZ0, 1PZ1), Apo enzyme homologues (1PYF, 3NT2)	Monomer	NADH	4-NBA DL-GA MG	6530 ± 280 1360 ± 70 40.1 ± 1.1 14,900 ± 500	MG reductase	None significant
YgdS	B2834	-	Holo enzyme (1LQA) reinterpreted as tertiary complex, NADP ⁺ and glycerol	Dimer	NADPH*	-	-	Tyrosine auxotrophy suppressor (tas) protein	None significant
YdhF	B1647	-	Apo enzyme (1UR3) Holo enzyme (1OG6)	Trimer	NADPH*	-	-	-	High temperature adapted <i>E. coli</i> strain
YghZ	B3001	AKR14A1	Apo enzyme (3N6Q, 4AST) tertiary complex, NADP ⁺ and citrate (4AUB)	Octamer	NADPH	4-NBA DL-GA 25DKG MG D-Xylose	3680 ± 130 850 ± 37 11.9 ± 1.9 2200 ± 20 13.8 ± 1.9	MG reductase, L-GAP reductase	Non glucose carbon source. Recovery from stationary phase in minimal media
YeaE	B1781	-	-	-	NADPH	4-NBA DL-GA MG	6,230 ± 140 77.7 ± 1.6 1890 ± 30	MG reductase	Cold stress
YajO	B0419	-	-	-	NADPH	4-NBA 2-NBA DL-GA	39.7 ± 5.0 22,900 ± 827 23.6 ± 5.0	2-Carboxy benzaldehyde reductase	Stationary phase
YdbC	B1406	-	-	-	-	-	-	-	Switch from glucose to acetate or lactose, Stationary phase

Abbreviations: 2-NBA, 2-nitrobenzaldehyde; 4-NBA, 4-nitrobenzaldehyde; DL-GA, DL-glyceraldehyde; MG, methylglyoxal; 25DKG, 2,5-diketo-D-gluconate; L-GAP, L-glyceraldehyde 3-phosphate. Specific activities taken from [17], experiments carried out at 25 °C, 100 mM potassium phosphate (pH 7.0), 10 mM KCl, and 0.1 mM NADPH with the exception of * which is calculated using the Michaelis constants derived from experiments carried out at 25 °C in 50 mM potassium phosphate (pH 7.0) with 0.5 mM EDTA from [15]. - no data available. Missing specific activities from the table, e.g. xylose, for some enzymes means the measured specific activity is <10 nmol min⁻¹ mg⁻¹, with the exception of YdjG for which only limited activities are reported [15]. # no biochemical data, presumed from cofactor bound in holo enzyme X-ray structure. Significant gene up-regulation (≥3-fold) was obtained from visual inspection and interpretation of data from multiple microarray studies from the *E. coli* Community's Gene Expression Database (GenExpDB) [23].

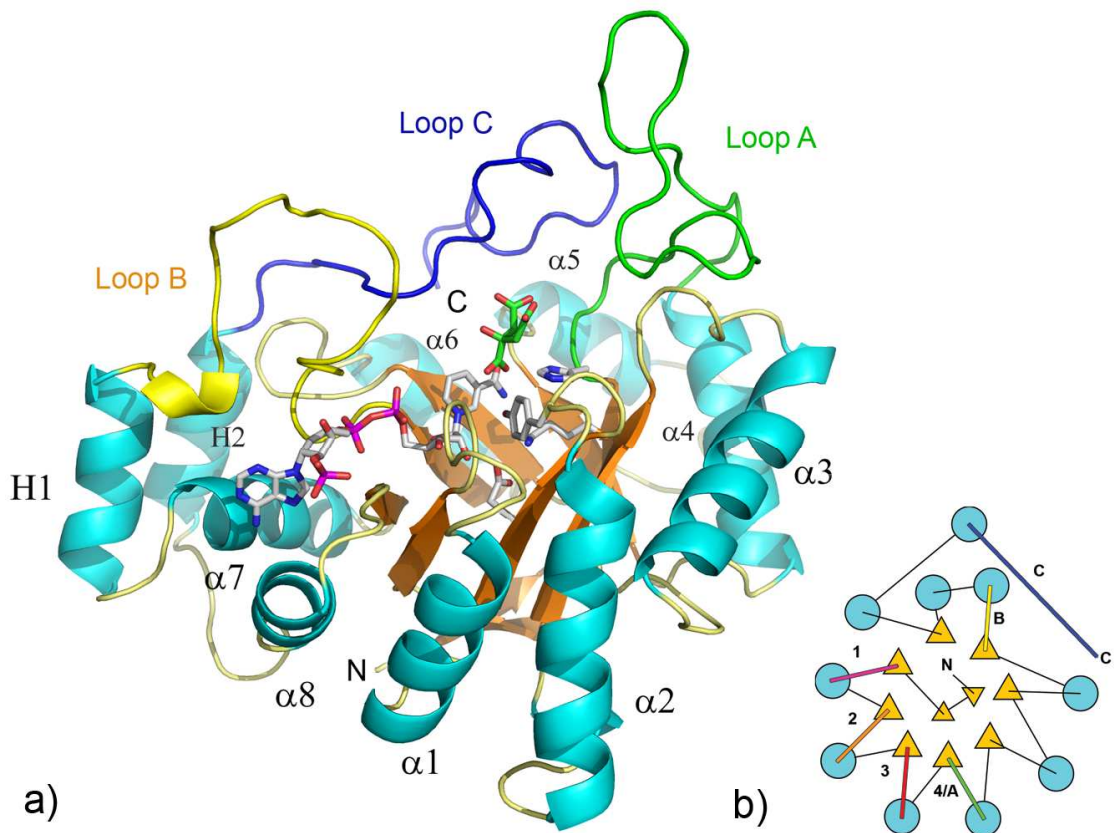


Fig. 1. A) A cartoon representation of a typical mammalian AKR, human aldose reductase AKR1B1 (PDB: 2ACS). A typical α/β barrel, β -strands coloured orange, α -helices coloured cyan and labelled $\alpha 1$ -8, with the two additional helices labelled H1-2. The main loops defined in the early AKR papers are loop A coloured green, loop B or seatbelt coloured yellow and loop C coloured blue. The NADP^+ is represented in stick as is a citrate molecule in the active site and the catalytic tetrad Asp 43, Tyr 48, Lys 77 and His 110. This and subsequent figures were generated using PyMol. B) A Topology diagram of AKR1B1 generated using TOPS [49] showing the main secondary structure elements of the AKR fold and highlighting the naming and colour scheme of the loop structures.

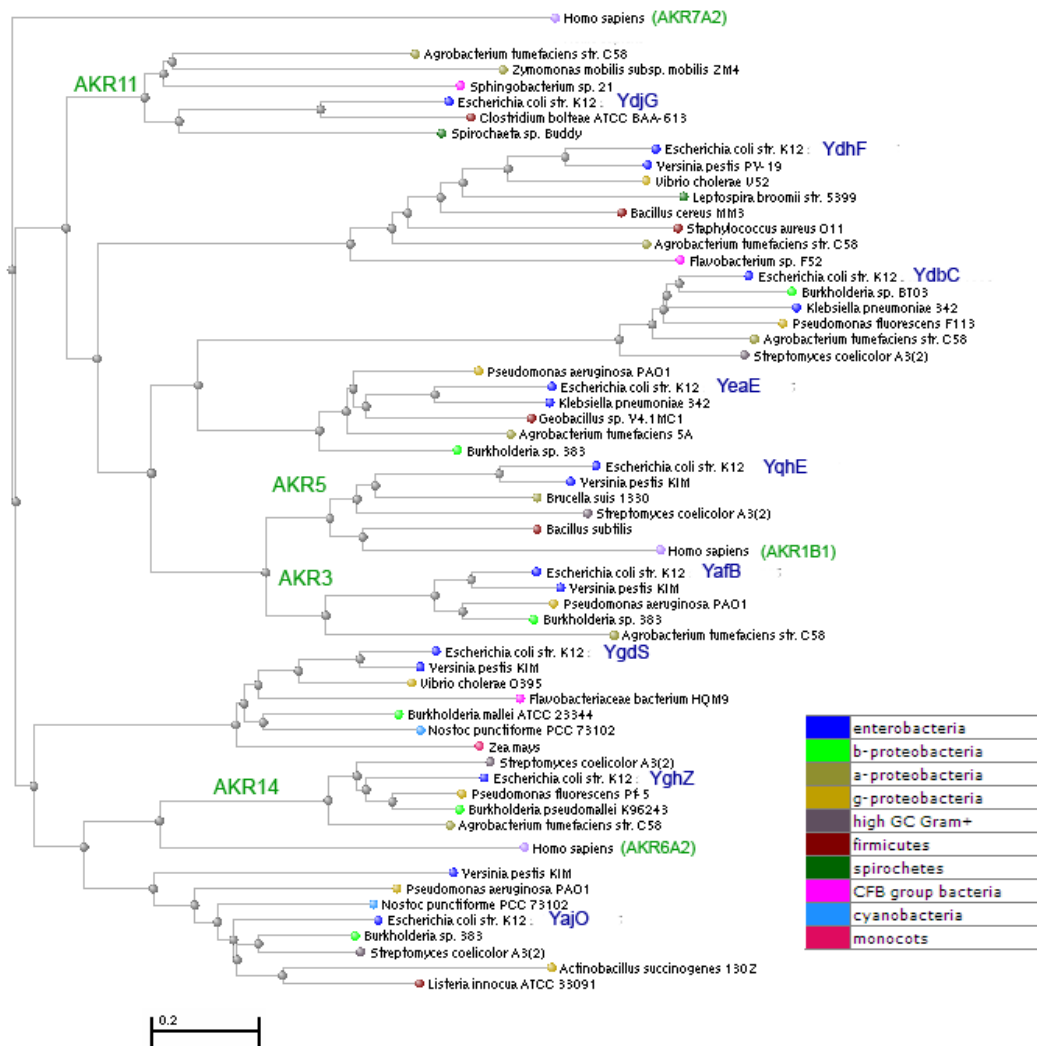


Fig. 2. A phylogenetic tree of the AKR proteins from *Escherichia coli* strain K12 sub strain MG1655 including representative sequences from various bacterial group. Three human AKR sequences from AKR families 1, 6 and 7 are included for reference. The diagram aims to show the predominant bacterial groups which share a related AKR sequence with an *E. coli* AKR and where possible the same bacterial strains are used for convenience. In particular the tree shows the sequence diversity within a given group of bacterial AKRs; for example, AKR14 (YghZ) related sequences are rather tightly clustered whilst AKR11 (YdjG) related sequences are more sequence diverse.

The phylogenetic tree was constructed using COBALT [30] using sequences from the NCBI Blast server.

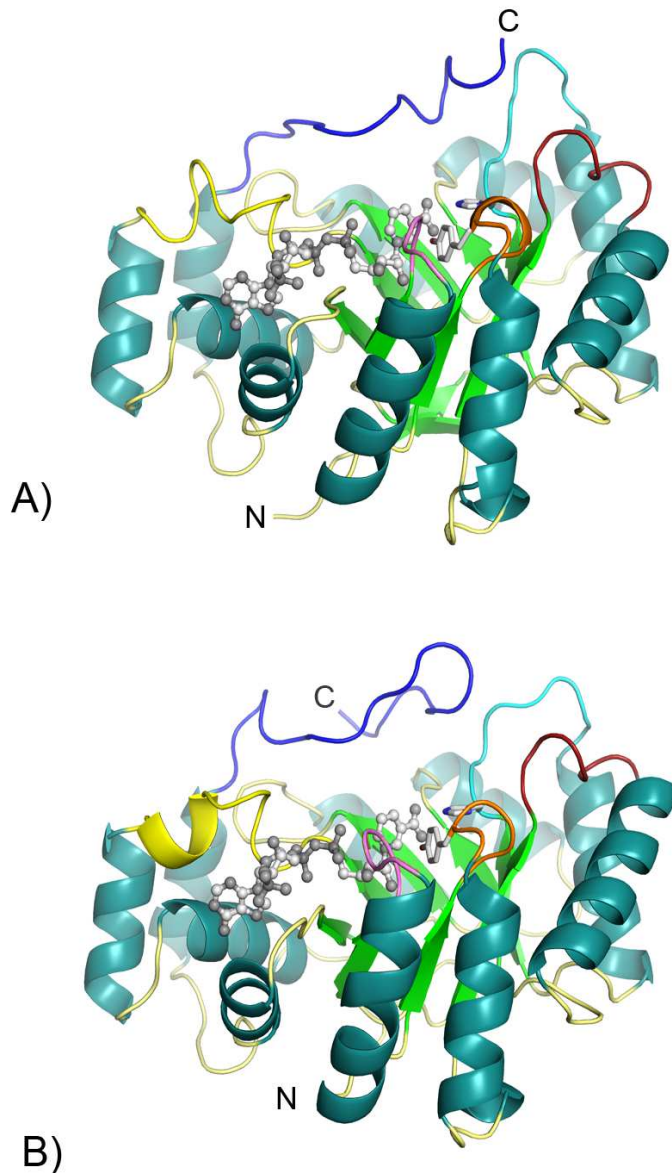
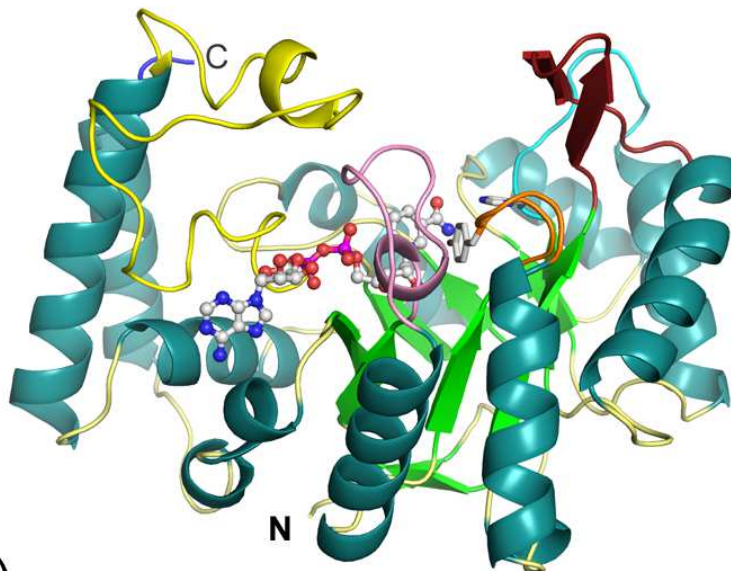


Fig. 3. A cartoon representation of A) YqhE (DkgA, PDB: 1MZR) and B) YafB homolog from *S. meliloti* 1021 (DkgB, PDB: 3UP8) with loop1 coloured lilac , loop2 coloured orange, loop3 coloured red, loop 4 coloured cyan , the NAD(P) binding loop coloured yellow and the C-terminal loop coloured blue. The likely position of the NADP+ is shown in grey ball & stick representation superimposed from the related structure 25DKG

reductase from *Corynebacterium* sp. (PDB: 1A80) for clarity. The active site tyrosine and histidine that form the oxyanion hole are shown in stick representation.



A)



B)

Fig. 4. A) schematic representation of the operon containing YdjG (blue). B) A ribbon representation of YdjG based on the X-ray structure of AKR11B1 from *B. subtilis* (PDB: 1PZ1) with loop1 coloured lilac, loop2 coloured orange, loop3 coloured red, loop 4 coloured cyan , the NAD binding loop coloured yellow and the C-terminal loop coloured blue. The position of the NAD⁺ is shown in ball and stick representation based on the position of NADP⁺ in 1PZ1. The active site tyrosine and histidine that form the oxyanion hole are shown in stick representation

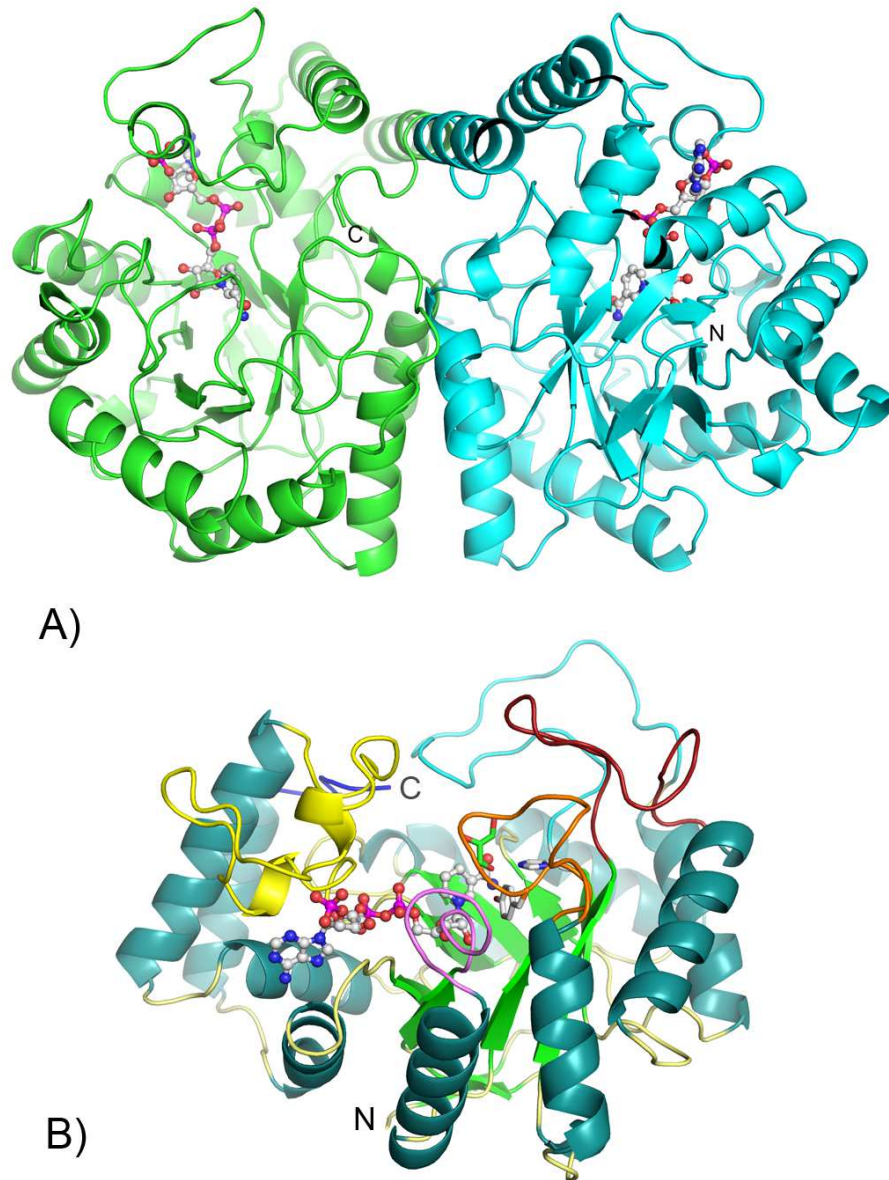


Fig. 5. A cartoon representation of A) the YgdS dimer (PDB: 1LQA), each monomer coloured uniquely and with the NADP+ shown in ball and stick. B) YdhF monomer with loop1 coloured lilac , loop2 coloured orange, loop3 coloured red, loop 4 coloured cyan, the NAD(P) binding loop coloured yellow and the C-terminal loop coloured blue. The active site tyrosine and histidine that form the oxyanion hole are shown in stick as is a glycerol in the active site pocket.

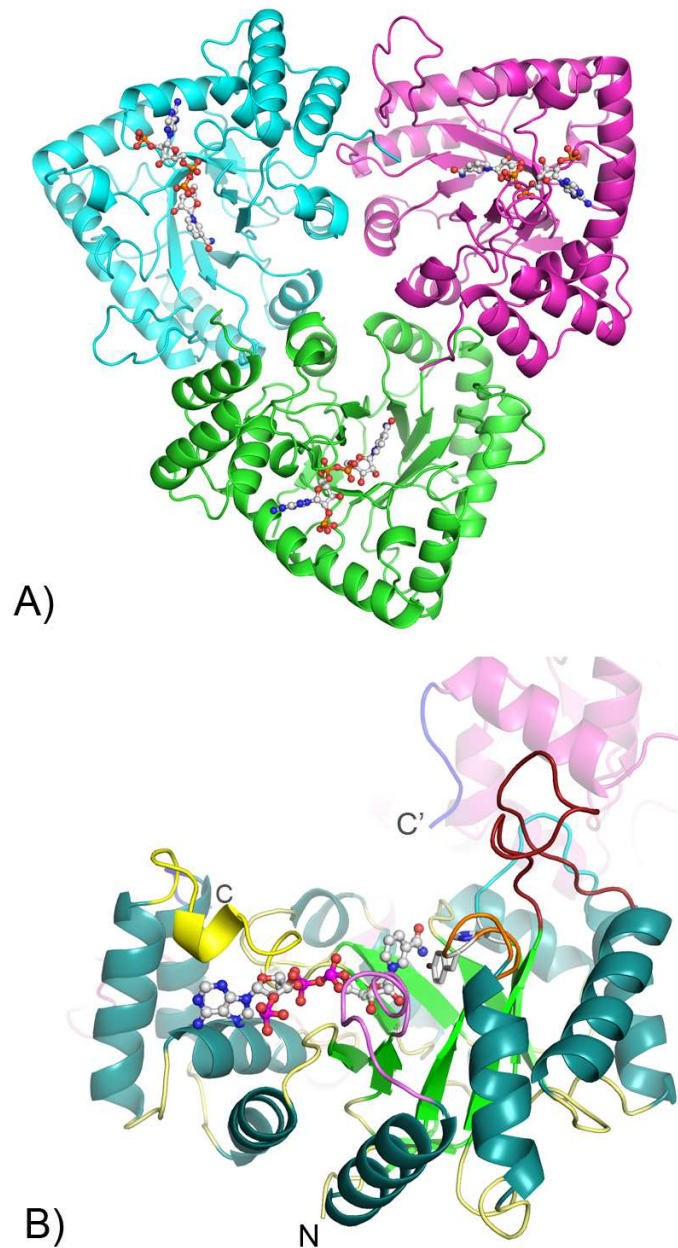


Fig. 6. A cartoon representation of A) the YdhF trimer (PDB: 1OG6) each monomer coloured uniquely and with the NADP+ shown in ball and stick. B) YdhF monomer with loop2 coloured orange, loop3 coloured red, loop 4 coloured cyan, the NAD(P) binding loop coloured yellow and the C-terminal loop coloured blue. The active site tyrosine and histidine that form the oxyanion hole are shown in stick. A related monomer is included coloured magenta (faded for clarity), showing how the C-terminal loop is contributed to the active site by another monomer.

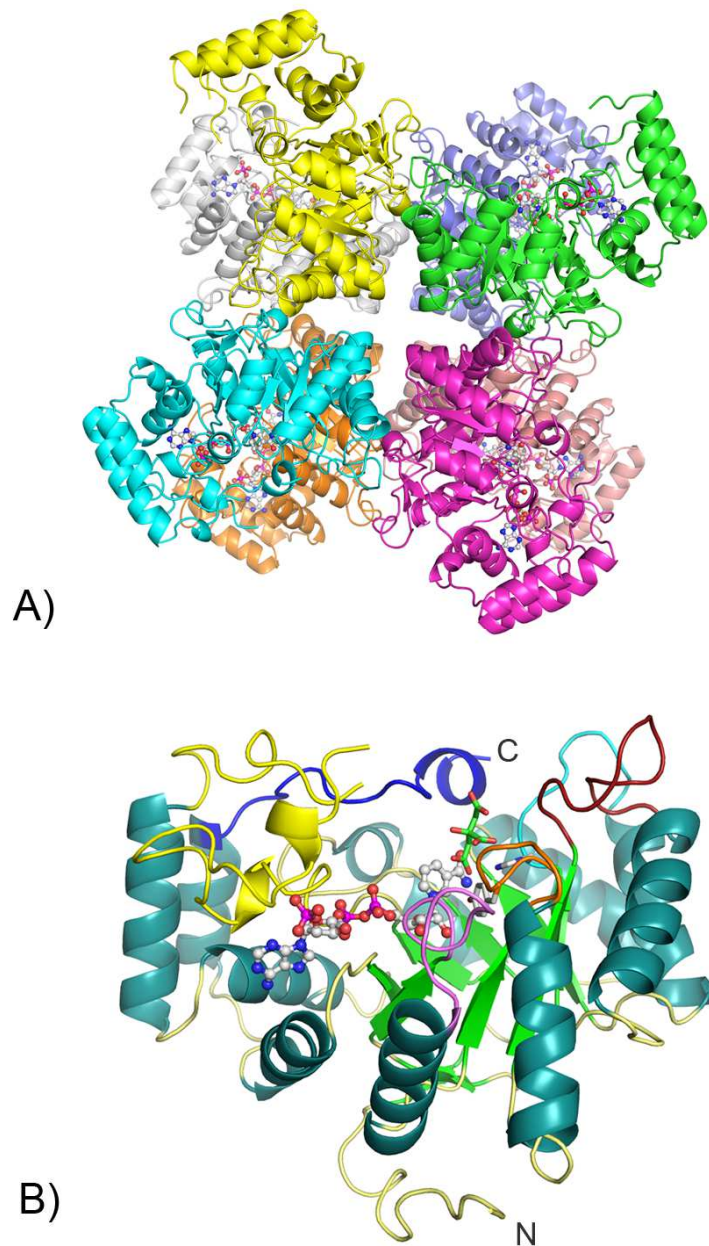


Fig. 7. A cartoon representation of A) the YghZ octamer (PDB: 4AUB) viewed down the 4-fold axis, each monomer coloured uniquely and with the NADP+ shown in ball and stick. B) YghZ monomer with loop1 coloured lilac , loop2 coloured orange, loop3 coloured red, loop 4 coloured cyan, the NAD(P) binding loop coloured yellow and the C-terminal loop coloured blue. The active site tyrosine and histidine that form the oxyanion hole are shown in stick, as is a citrate in the active site pocket.