ABSTRACT

Aldo-keto reductases (AKRs) are known for their ability to catalyze a wide variety of substrates and have the potential to catalyze reactions important for pharmaceutical and industrial use. Using computational software, the structure and function of AKRs 169, 304, and 308 in Saccharomyces cerevisiae, were studied with a focus on homology, active site binding, and substrate screening to determine candidate molecules for AKR catalysis. Computational docking was executed to predict binding compatibility between the substrates and the AKRs. Docking results that placed substrates near the enzyme cofactor NADPH were used to determine the active site of the proteins. Previous work has found that AKR 169, which corresponds to the protein AKR5F, is capable of detoxifying vanillin to vanillyl, a major component for yeast survival with industrial applications. It was also determined that AKR 304 is homologous to GRE3P and can act as a methylglyoxal reductase. It was also found that AKR 308, which is homologous to Ara1p, is capable of reducing ethyl acetoacetate, ethyl 4-chloroacetoacetate, and 2,3-pentanedione and exhibits enzyme activity not modeled by traditional Michaelis-Menten kinetics. Computational homology tests showed that AKRs are highly similar with present-day yeast strains, and laboratory results displayed AKR 308’s ability to catalyze the reduction of ethyl acetoacetate, ethyl 4-chloroacetoacetate, and 2,3 pentanedione suggesting that AKR 308 performs best with linear substrates.

INTRODUCTION

Aldo-Keto Reductases

Aldo-keto reductases (AKRs) are a superfamily of enzymes that reduce ketones and aldehydes with the help of NADPH (1). AKRs consist of over 190 members which are sorted into 16 families. These proteins reduce a variety of carbonyl-containing substrates, including sugar aldehydes, keto-steroids, quinones, and diketones (1). In addition, AKRs are known to catalyze a bi-bi kinetic mechanism where they rely on NADPH as a cofactor to reduce ketones and aldehydes, creating alcohols (1).

Typically, AKRs are 37 kDa monomers that contain a conserved cofactor binding domain (1) and an (α/β)8-barrel structure with large loops at the back of the barrel, which control specificity of substrates. Primary and secondary structures of AKRs are most often affected by amino acid modifications and spatial rearrangements throughout time; however, their tertiary structure tends to be conserved (2) and thus their functions are generally maintained.
The yeast *S. cerevisiae* has been extensively studied due to its non-pathogenicity, availability, and sequenced genome. Its AKRs have been investigated extensively, including in pharmaceuticals for its oxidoreductase capabilities (3). Studies have also been conducted on re-engineered *S. cerevisiae* to maximize biofuel production by manipulating the substrate docking through AKR gene deletion (4) or overexpression (5). As a result, these yeast strains and their AKRs have been promising candidates for future studies in a wide variety of scientific disciplines.

The AKRs being investigated in this study come from previous work that screened 45-million-year-old *Saccharomyces cerevisiae* yeast cells retrieved from fossilized amber (6). They identified eight AKR strains: AKR 105, 163, 169, 238, 304, 308, 437, and 456 but the definitive identities of these proteins were not indicated nor analyzed (6). In this study, five of the eight AKR strains that were identified from SC108 were cloned, namely AKR 163, 169, 304, 308, and 456 but only AKRs 169, 304, and 308 were chosen for further research through homology searching and substrate docking. Lastly, AKR 308 was additionally selected for an enzyme kinetics evaluation with multiple potential substrates.

**Computational Biochemistry**

Computational biochemistry is an essential field of study that is applicable to various disciplines such as cancer, drug therapy, and genetic sequencing of viruses and bacteria, especially due to the constant development and improvement of technologies over the past decades. This area of study allows for greater understanding of biological concepts by giving scientists the opportunity to formulate and test hypotheses that were once difficult to test or verify (7). Computational biochemistry can help scientists describe molecular functions and interactions in structures such as DNA sequences. Also, a computational biochemist may represent ligand docking in enzymes, providing the ability to document reactions with substrates with precision (8). Computational modeling, a main component of computational biochemistry, has the ability to further allow scientists to learn about other biological processes, such as the viability of chemical reactions with varying compounds or examining protein sequence similarities across species. Advancements in technologies such as protein mass spectrometry, sequence homology, and microarrays have propelled the field in constant growth (9).

Computational biochemistry has been applied to study AKRs. Studies using the homology of AKRs to other proteins allow scientists to understand the relationship and evolutionary pattern of AKRs among several species, as these proteins are found in animals, plants, and fungi (10). Furthermore, AKR studies often make use of 3-dimensional modeling softwares to study protein structure. These studies allow researchers to learn potential ligand binding sites and chemical reactions that take place within AKRs, helping to elucidate differences in protein function (11).

There are various programs and softwares that can be implemented to study AKRs. MODELLER is a program that uses a given amino acid sequence and homologous protein that has a known protein structure, then it uses both to make an estimated structure of the protein (12). SwissDock, a web-based application, allows researchers to dock small ligands into a protein of interest known as the target protein (13). After the docking is complete, different clusters and positions of the ligand within the protein can be viewed. This program showcases
various docking predictions, showing the location and favorability of the prediction (13). PyMOL is another software that showcases molecular graphics, enabling users to create 3-dimensional models of proteins and nucleic acids and manipulate, measure, and analyze these structures. The program can also be used to visualize protein-ligand interactions including the binding of the two distinct compounds (14). Furthermore, SeeSAR is a drug docking program used for high-throughput screening of compounds. SeeSAR enables estimation of binding affinities and identification of possible binding sites.

In this study, the yeast proteins corresponding to AKRs 169, 304, and 308 were identified using homology sequence identification (BLASTp) and tertiary structure homology identification (HHpred), and these homologous structures were utilized to develop their 3-dimensional protein structures through the softwares mentioned above. NADPH binding sites were found using the protein structure alignment of homologous structures, and software-base ligand docking was performed to identify each substrate’s binding site according to each AKR. Preliminary experimental data for AKR 308 was also collected to analyze the reactions with various substrates and characterize kinetic behavior.

METHODS

Homology

The National Library of Medicine’s Basic Local Alignment Search Tool (BLAST) was used to find homologous proteins to AKRs 169, 304, and 308. The Blastp algorithm was utilized and searches were confined to S. cerevisiae in the non-redundant protein sequences database. Settings for the search included default word size (6), expect threshold (0.05), and matrix (BLOSUM62) parameters. Because it was only desirable to compare the AKRs with highly homologous sequences, only proteins that had more than 75% shared identity were analyzed while hypothetical and uncharacterized proteins were disregarded.

AKR Modeling and Ligand Docking: SwissDock and PyMOL

Protein models of each AKR were created using the “Homology Detection & Structure Prediction by HMM-HMM Comparison” (HHpred) feature in the Bioinformatics Toolkit provided by the Max Planck Institute for Biology (12, 15, 16). With an HHpred search, protein structures corresponding to the proteins from the homology results were investigated. The protein sequences with modeled structures and the highest homology to AKRs 169, 304, and 308 (1QWK [17], 1QWK [17], 4IJR [18], respectively) were forwarded to MODELLER as a template to verify the folding of each AKR’s 3-dimensional model. MODELLER used the homologous proteins as template to produce a 3-dimensional model for each AKR. These models were saved for further use in ligand docking.

A ligand docking study was conducted for each AKR, but SwissDock (13) was used primarily for AKRs 169 and 304. The ligands xylose, arabinose, vanillin, NADPH, 3,4-hexanedione were docked with AKR 169 with the software SwissDock while xylose, NADPH, and methylglyoxal were docked with AKR 304 using the same software. Several clusters with low Gibbs free energy were identified for each ligand on Swissdock. To narrow down a potential active site for AKR 169, PyMOL aligned the homologous hypothetical C. elegans protein (1QWK [18]) and the NADPH-bound form of Ara1 (4IJR [18]). For AKR 304, PyMOL aligned...
AKR 304 with the APO form of codeinone reductase (7MBF [19]), *Candida tenuis* xylose reductase, and the NADPH-bound form of Ara1 (4IJR [18]). Because of the high homology of the hypothetical *C. elegans* protein to AKR 169, we translated the location of the known active site of the hypothetical *C. elegans* protein (17) to AKR 169. SwissDock then docked all of the ligands to AKR 169 and 304 with a restriction of 6 Å from the approximate center of the potential active site.

**SeeSAR**

For known substrates of AKR 308, ChemMine generated a list of 200 structurally similar compounds to acetoin, methylglyoxal, diacetyl, and pentanedione for further investigation based on their PubChem Fingerprint similarity measures (20). While SwissDock was advantageous for testing only a few ligands, the program could only dock one substrate at a time, which was especially disadvantageous given that predictions could take hours to compute. To model the hundreds of substrates found in the PubChem database, SeeSAR was utilized as it allowed for multiple compounds to be docked simultaneously. After docking the groups of compounds with the MODELLER-produced AKR 308 structure, the substrates were pared down based on their estimated affinity. Substrates that had an estimated affinity outside of the micromolar (μM) range were eliminated, and the remaining substrates were further investigated in PyMOL. PyMOL revealed the compounds of this selected group that had the closest proximity to the nicotinamide ring in AKR 308’s binding pocket, which were deemed most suitable for the protein. Acetoin, methylglyoxal, diacetyl, and pentanedione were further docked using SwissDock to calculate their thermodynamic favorability.

**Enzyme Kinetics**

AKR 308 was cloned and purified to test its reactivity with several substrates. AKR 163 was used as a positive control. Based on previous literature, 2,3-pentanedione (21), vanillin (22), ethyl acetoacetate (23), ethyl 4-chloroacetoacetate, and xylose (24) were selected as the substrates to test with AKR 308. The kinetic analysis was performed in 25 mM Tris-HCl buffer at pH 8.0. The Km values for each substrate was determined by using NADPH (0.1 mM) and varying the substrate concentrations (2,3-pentanedione from 0.5 to 15 mM, ethyl 4-chloroacetoacetate from 0.03 to 6 mM, and ethyl acetoacetate from 16 to 160 mM). In addition, xylose (2.5 mM) and vanillin (1 mM) were used with NADPH (0.1 mM) to test the catalytic activity of AKR 308 with both substrates. Similar solutions were made using AKR 163 to serve as a control, allowing for activity comparison. All experiments were performed at 25°C and data was collected using the Cary 60 UV-Vis Spectrophotometer set to a wavelength of 340 nm. Absorbance levels were taken every 15 seconds for 5 minutes unless there was low catalytic activity, in which case levels were taken every minute for 10 minutes.

**RESULTS**

**Homology**

BLASTp aligned the sequences for each AKR identified (Appendix A) to several homologous proteins with high percent identity match and query coverage contained in the National Center Biotechnology Information database. The results indicated close alignment between proteins (Table 1). The primary amino acid sequence of AKR 169 was closest to AKR5F, AKR 304 was closest to GRE3p, and AKR 308 was closest to Ara1p. It is important to
note that although AKR 304 and AKR 308 had 100% identity matches to their homologous proteins, AKR 169 was 98% identical, meaning there were slight differences in their primary structures. It is likely that AKR 169 and AKR5F’s tertiary structures would be similar as tertiary structure is highly conserved (2).

Table 1: Identification of the corresponding proteins of each AKR and the corresponding genes that produce the proteins.

<table>
<thead>
<tr>
<th>AKR</th>
<th>Homologous Protein</th>
<th>Gene</th>
<th>Identity Match (%)</th>
<th>Query Coverage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKR 169</td>
<td>AKR5F</td>
<td>YJR096W</td>
<td>98.00</td>
<td>100</td>
</tr>
<tr>
<td>AKR 304</td>
<td>GRE3p</td>
<td>GRE3</td>
<td>100.00</td>
<td>100</td>
</tr>
<tr>
<td>AKR 308</td>
<td>AKR3C1/Ara1p</td>
<td>Ara1</td>
<td>100.00</td>
<td>100</td>
</tr>
<tr>
<td>AKR 308</td>
<td>glyoxal/methylyglyoxal reductase*</td>
<td>YvgN</td>
<td>35.26</td>
<td>95</td>
</tr>
</tbody>
</table>

*Selected from *Bacillus subtilis* (strain 168)

3-Dimensional Models

HHpred results identified homologous proteins to each AKR that had significant similarities in tertiary structure (Appendix B). Intriguingly, despite having a *S. cerevisiae* protein as a primary structure homolog, AKR 169 and 304’s tertiary structures were homologous to a hypothetical oxidoreductase *C. elegans* protein (PDB code:1QWK) (17). As a result, both AKRs were modeled in MODELLER using the structure of a protein found in a separate species, while AKR 308 was modeled using the homologous *S. cerevisiae* protein Ara1p (PDB code: 4IJR). The protein models generated by matching the tertiary structure of each AKR with their respective structural homolog are shown in Fig. 1, after the results were imported into PyMOL. Each AKR model displays an alpha-beta barrel structure and the similarities between the three figures suggests that tertiary structure is highly similar across all AKRs.

Fig. 1 PyMOL visualizations of 3-dimensional protein structures. a) AKR5F (AKR 169) modeled with the hypothetical *C. elegans* protein. b) GRE3p (AKR 304) modeled with the hypothetical *C. elegans* protein. c) AKR3C1 (AKR 308) modeled with Ara1p.
**Ligand Docking**

Having the AKRs accurately modeled, docking of NADPH and other ligands with SwissDock and SeeSAR was initiated. Moreover, these structures were imported into PyMOL so that potential active sites could be investigated.

**AKR5F/AKR 169**

NADPH was docked into AKR5F with SwissDock. Several binding clusters and multiple ligand arrangements were found but the most stable docking was found to be in a region within 5Å of amino acids Tyr 22, Tyr 50, Trp 83, Ser 116, Phe 195 (Fig. 2).

![Fig. 2 AKR5F bound with NADPH in potential active site.](image)

**GRE3p/AKR 304**

GRE3p was also docked with NADPH fitting tightly into a binding pocket (Fig. 3a). Moreover, this active site was investigated and the closest amino acids interacting with the ligand were Ser 170, Asn 171, Gln 192, Asn 286, His 111, Cys 20, Trp 21, Lys 22, Gly 168, Val 169, Phe 172, Tyr 218, Ser 219, Ser 220, Phe 221, Gly 222. These amino acids were specifically examined for potential binding as they were listed in either the literature (18) or the HHpred search since NADPH binding sites for similar NADPH dependent reductases are known (18). However, there were steric clashes between residues Cys 20, Trp 21, Lys 22, and the nicotinamide ring of the NADPH (Fig. 3b).

The docking of xylose and methylglyoxal were set near the specific coordinates of the NADPH binding site because close proximity to the NADPH cofactor allows the protein to catalyze their reductions. The former docked close to the nicotinamide ring and interacted with amino acids Trp 21, Tyr 49, Asp 48, Trp 80, His 111, Phe 112, Glu 122 (Fig. 3d), while the latter docked further away from NADPH (Fig. 3c), interacting with residues Tyr 49, Asp 48, Trp 21, Lys 22 (Fig. 3e). These results suggest that these aforementioned amino acids form the active site of GRE3p.
Fig. 3 GRE3p bound to multiple ligands in a potential active site. a) Potential binding site for NADPH to AKR 304. b) Homologous protein 4IJR NADPH binding to AKR 304 c) Xylose (pink) binds with AKR 304 2.4Å away from the nicotinamide ring shown in the confirmation of its lowest ΔG (kcal/mol) value d) Active site of AKR 304 based on xylose (red) e) Active site of AKR 304 based on methylglyoxal (red).

Ara1p/AKR 308

To visualize the docking of NADPH on the MODELLER-generated AKR 308, the protein was aligned with Ara1p as the latter was available as an x-ray crystallography structure with NADPH already bound (Fig. 4).

Fig. 4 Alignment of Ara1p and AKR 308 proteins. PyMOL aligned Ara1p (PDB code: 4IJR) (orange) with AKR 308 model (light blue), and NADPH (purple) in the active site.

Docking of acetoin related molecules in SeeSAR and visualization of the results in PyMOL showed that 2-(S)-hydroxyhexan-3-one and 1-fluoro-4-hydroxypentane-3-one were located about 1Å-3Å away from the nicotinamide ring (Fig. 5a, Fig. 5b, respectively), making them sterically compatible with the NADPH binding site and suitable for reduction by Ara1p. Compounds 3-hydroxy-2,4-hexanedione and 3-(S)-hydroxyhexan-2-one were either located too
close or far away from the ring (Fig. 5c, Fig. 5d, respectively), making them unsuitable for reacting with NADPH. The remaining substances used in the screening had low estimated binding affinities, making them unlikely candidates for catalysis. When docking with similar compounds to methylglyoxal, substrates 2-oxobutanedial, propanedial, 2,3-dioxobutanal, propionaldehyde, and 1-oxopropan-2-olate appeared to be suitable for catalysis. Analysis of molecules similar to 2,3-pentanedione revealed butan-2-one to be a well-fit compound while investigation of diacetyl’s comparable compounds showed 1-iodo-2,3-butanedione and 5-Fluoropentane-2,3-dione to be best.

The substrates found suitable for reduction by preliminary SeeSAR screening were then analyzed in SwissDock, along with other linear structures. Substrates such as ethyl acetoacetate, ethyl 4-chloroacetoacetate, and 2,3-pentanedione displayed Full Fitness values below -1700 kcal/mol, while ringed structures such as vanillin and xylose had Full Fitness values below -1600 kcal/mol (Table 2). These highly negative Full Fitness values and negative ΔG values indicate a high probability of substrate binding with Ara1p and made these five compounds attractive for experimental testing. However, the less negative Full Fitness values for vanillin and xylose suggested that these substrates would exhibit weaker binding affinities with AKR 308 than the linear substrates.

**Table 2:** Thermodynamic Favorability of various substrates for AKR 308.
<table>
<thead>
<tr>
<th>Substrate</th>
<th>Full Fitness (kcal/mol)</th>
<th>Estimated ΔG (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethyl Acetoacetate</td>
<td>-1742.69</td>
<td>-6.51</td>
</tr>
<tr>
<td>Ethyl 4-Chloroacetoacetate</td>
<td>-1739.48</td>
<td>-6.68</td>
</tr>
<tr>
<td>2,3-Pentanedione</td>
<td>-1728.22</td>
<td>-6.27</td>
</tr>
<tr>
<td>Vanillin</td>
<td>-1691.53</td>
<td>-6.42</td>
</tr>
<tr>
<td>Xylose</td>
<td>-1663.99</td>
<td>-6.52</td>
</tr>
</tbody>
</table>

**Enzyme Kinetics**

For AKR 308, NADPH’s absorbance of 340 nm light was measured over time with enzyme-substrate solutions, while varying the concentrations of each substrate. AKR 308 was first tested with 4 potential substrates: 2,3-pentanedione, ethyl acetoacetate, xylose and vanillin. There was strong catalytic activity for 2,3-pentanedione and ethyl acetoacetate, while xylose and vanillin showed little to no activity as seen in Fig. 6. This is shown by the steep slope in Figs. 6a and 6c compared to a much lower slope in Figs. 6b and 6d.

**Fig. 6 Absorbance graphs of various substrates binding with AKR 308.** a) ethyl acetoacetate, b) xylose, c) 2,3-pentanedione, d) vanillin.
Following confirmation of catalytic activity with AKR 308 and ethyl acetoacetate and 2,3-pentanedione, we tested catalytic rates at varying substrate concentrations. Due to AKR 308’s strong reactivity with ethyl acetoacetate, we also tested ethyl 4-chloroacetoacetate because of its high chemical similarity. Additionally, we tested these substrates with AKR 163 as a control because it is known to catalyze ethyl acetoacetate and ethyl 4-chloroacetoacetate.

The velocity against substrate concentration was then plotted for the substrates that exhibited strong catalytic activity with AKR 308 – 2,3-Pentanedione, ethyl acetoacetate, and Ethyl 4-chloroacetoacetate – to calculate the $K_m$ constant values and $V_{\text{max}}$ for each substrate.

With increasing concentrations of substrates, Figs. 7c, 7d, and 7f seem to follow Michaelis-Menten kinetics, suggesting that speed of their respective reactions are eventually limited by the concentration of the substrate. Fig. 7b stands out from the data with the best correlation to the substrate inhibition model, while Figs. 7a and 7e have comparable $R^2$ values for each model. Further research at a wider range of concentrations is needed to clarify whether the activity better suits substrate inhibition or Michaelis-Menten kinetics. Based on this data, AKR 163 and AKR 308 are shown to act rather differently from each other since reactions with the same substrate follow different kinetic models.

![Fig. 7 Enzyme activity curves for AKR 163 and 308 relating catalytic velocity to substrate concentration](image)

Data is fitted to the Michaelis-Menten model and substrate inhibition models. $R^2$ values and calculated parameters for each model are inset. a) AKR 163 and 2,3-pentanedione, b) AKR 163 and ethyl 4-chloroacetoacetate, c) AKR 163 and ethyl acetoacetate, d) AKR 308 and 2,3-pentanedione, e) AKR 308 and ethyl 4-chloroacetoacetate, g) AKR 308 and ethyl acetoacetate.

[17-10]
DISCUSSION

AKR 169

Because the BLASTp alignment showed 98 percent identity for the sequence of AKR 169 and AKR5F, it is likely that the two proteins share many similarities in their role in *S. cerevisiae*. These two proteins are coded by the YJR096W gene, which plays an important role in vanillin reduction (22), 2-chloro-3-phenylpropanoate reduction (25), increasing the activity of xylose and arabinose reductase (Tr), and esterifying exogenous sterol (26). The ligand docking study not only confirms that xylose, arabinose, vanillin, 3,4-hexanediol, and NADPH are likely to bind with AKR 169, but it provides evidence of the potential active site where substrate-enzyme interactions are likely to occur. Nitrogen starvation (27) and exposure to methyl methanesulphonate (28) have been described as ways to induce the expression of YJR096W, indicating that AKR 169 is responsible for responding to cellular stresses. This potential role suggests that vanillin serves as a stimulator because it helps to detoxify vanillin which, when built up, causes cellular stress. Because the ligand docking study showed that xylose and arabinose are likely to bind with the potential active site, it appears that their concentration may increase during cellular stress, therefore activating AKR 169. Even though 3,4-hexanediol is reduced by AKR 169’s ortholog AKR1B1 in humans, it was unknown whether it is associated with AKR 169; however, the ligand docking study showed that 3,4-hexanediol likely binds to the potential active site.

AKR 304

It was found through the BLASTp alignment that the sequence for the AKR 304 was identical to the sequence for the protein GRE3p. This protein is widely regarded as the major xylose reducing enzyme in *S. cerevisiae* (5), despite *S. cerevisiae* being generally regarded as a non-xylose utilizing organism (29). However, an increased expression of GRE3 is induced under stress conditions, including osmotic, ionic, oxidative and heat stress, implying that the protein may aid in stress regulation in *S. cerevisiae* (30). This uptick in GRE3p production may allow the yeast to reduce xylose to xylitol for use in energy production in times of need (4). Given the function of this protein in these conditions, xylose was further studied with AKR 304. It was determined that xylose and NADPH had a highly probable docking site in the modeled protein (Fig. 3c). This corroborated the claims made in the literature about xylose docking to GRE3p given its success and close proximity to the modeled NADPH binding site. This furthers prior research by pointing to a potential structural explanation for the observed xylose reduction phenomenon in GRE3p. With this, AKR 304 could potentially be designated as a major xylose reductase in this ancient *S. cerevisiae* strain, following further in-lab studies.

In addition, some studies have pointed towards an alternate use of the protein in reducing diketones. One example of a molecule containing this functional group is methylglyoxal (31), a very reactive by-product of glycolysis and, if left unchecked, can react with parts of the cell it is not supposed to, thereby causing damage. Thus, it needs to be decomposed by methylglyoxal reductase or a chain of reactions involving glyoxalase I and II (Appendix C). When *S. cerevisiae* was observed under various stress conditions, such as oxidative stress, ionic stress, and carbon starvation, methylglyoxal levels decreased by 49% while GRE3 expression increased (31). Alluding to AKR 304’s ability to reduce methylglyoxal.
In further support of AKR 304’s function as a methylglyoxal reductase (31), a BLASTp was run comparing *Bacillus subtilis* (strain 168), against AKR 304. The top homology match was a known *B. subtilis* methylglyoxal reductase, with a 98% query coverage and 34% shared percent identity, a reasonably high match for an enzyme of a different species (Table 1). Given these results, a docking study was performed with AKR 304 and methylglyoxal to investigate possible active binding sites. Xylose and methylglyoxal docking with AKR 304 appear to share amino acids Trp 21, Tyr 49, Asp 48 within 5Å (Figs. 3d and 3e), providing evidence that they both bind near the active site, which gives us confidence that the enzyme is versatile enough to bind with methylglyoxal. Given that AKR 304 bound methylglyoxal in its active site, it appears that AKR 304 would likely reduce methylglyoxal.

**AKR 308**

BLASTp homology results show that AKR 308 has a 100% identical amino acid sequence to Ara1p, allowing us to clearly identify the binding pocket and test a wide variety of substrates using existing Ara1p structural models. Both structures exhibited an α, β barrel structure and shared active sites with the same amino acids. The high similarity between AKR 308 and Ara1p also allowed us to create an accurate model of the AKR using MODELLER for further study.

NADPH has been found to bind at the carboxyl edge of the β-strands of the barrel in Ara1p in an extended conformation, close to the nicotinamide moiety of the NADPH cofactor (18). As an essential cofactor for AKR function and hydrogen donor, NADPH bound to AKR 308 informed decisions on which binding pockets were suitable for catalysis. Since SwissDock modeling is based purely on protein structure, the NADPH bound protein structure confirmed that computational modeling was consistent with observed structures. Compounds found to clash with NADPH were eliminated because they would not be able to react appropriately with NADPH. Compounds found to bind in a different protein site were eliminated as well because the long distance from NADPH would prevent effective reduction via NADPH’s hydrogen atom.

AKR 308 has the potential to catalyze a wide variety of compounds with pharmaceutical and industrial importance. Docking results in SeeSAR showed that chemically similar compounds to known substrates could bind strongly to the enzyme with binding affinities in the μM range.

Ara1 catalyzes the reduction of compounds such as methylglyoxal, diacetyl and pentanediol, which are usually toxic metabolic by-products in yeast that could potentially cause mutagenesis and cellular damage (18). In addition, Ara1 catalyzes the reduction of acetoin in the presence of NADPH (18). This is supported by the minimal distance between the bound NADPH of AKR 308 and the acetoin substrate as shown in Fig. 5. Enzyme kinetics results for AKR 308 and substrates ethyl acetoacetate and ethyl 4-chloroacetoacetate, which are structurally similar to acetoin, revealed strong catalytic activity, confirming the results from the substrate docking (Fig. 6). This presents many interesting substrates that could be studied in the future and shows that computational methodology is capable of accurately predicting potential substrates.

Comparison of AKR 308 and AKR 163 showed significant differences in their enzyme kinetics only when catalyzing 2,3-pentanediol and ethyl 4-chloroacetoacetate while activity
was similar when catalyzing ethyl acetoacetate. \( K_m \) values for both AKR 308 and AKR 163 were similar when catalyzing ethyl acetoacetate and followed similar kinetic profiles (Fig. 7). When catalyzing 2,3-pentanedione, AKR 308 followed traditional Michaelis-Menten kinetics while AKR 163 showed peak activity around 7-10 mM, possibly suggesting substrate inhibition. AKR 308 also had a greater \( K_m \) value, suggesting a lower substrate affinity and a need for greater substrate concentrations for significant catalytic activity or a faster \( K_{\text{cat}} \). In addition, when catalyzing ethyl 4-chloroacetoacetate, AKR 163 exhibited strong substrate inhibition with peak activity at 1 mM while AKR 308 showed maximum activity at 4 mM. AKR 308’s peak enzyme activity at 4 mM of ethyl 4-chloroacetoacetate and decline in activity at 6 mM suggests possible substrate inhibition. AKR 308’s ability to continue substrate catalysis at higher concentrations presents an advantage over AKR 163, which shows dramatically decreased activity.

Additionally, AKR 163 and AKR 308 were tested with vanillin and xylose, known substrates of AKR 169 and AKR 304, respectively, to identify whether substrates for other AKRs could be catalyzed by AKR 308. Experimental results confirmed computational docking predictions, which showed low activity between the vanillin and xylose substrates with AKR 308. However, the lower concentrations of vanillin and xylose when compared to ethyl acetoacetate may have affected the enzyme kinetics results. In the control reaction with AKR 163, xylose exhibited a weak reaction as absorption decline could only be observed over a longer period of time (Fig. 6). Vanillin and xylose both exhibit a ring structure while ethyl acetoacetate and 2,3-pentanedione, which performed better in docking and lab experiments, are linear structures. Further research could investigate a rationale for why ring structures do not perform as well with AKR 308.

**CONCLUSION**

The motivation behind exploring ancient yeast strains was to determine potential substrates and how they would dock into the active sites of various aldo-keto reductases. The first step was to determine the homology of AKRs 169, 304, and 308 using the BLASTp program, which led to the conclusion that YJR09W, GRE3, and Ara1p were homologs, respectively. Upon finding these close homology matches, we looked into the published literature on these homologs to find potential substrates for further testing. Given these substrates, a ligand docking study was performed through SwissDock, which showed us the ligand clusters and elements that bound to a protein. The SwissDock results were put into PyMOL so that we could investigate the active sites and modes of binding. We modeled each AKR sequence using MODELLER based on the homologs. The MODELLER results of AKRs were aligned with SwissDock results to compare binding sites of the ligands. After careful analysis of the binding sites, it was determined that AKR 308 had the most promising docking results, that is, this protein displayed multiple ligands binding into a specific pocket near NADPH, leading us to believe it was the correct active site to carry out the reduction reactions. Therefore, we chose AKR 308 to study further in a laboratory setting. After looking at the results, it was concluded that it is more likely for substrate inhibition to occur when there are stronger electron effects on the carbon for reduction. The absorbance results of 2,3-pentanedione, ethyl acetoacetate, and ethyl 4-Chloroacetoacetate with AKR 308 matched those of AKR 163. Both displayed a negative, linear relationship, indicative of activity occurring between the substrate and enzyme. The data suggested that AKR 308 is less prone to substrate inhibition due
to the high $K_m$ values. Because the MODELLER results matched the lab results, we were able to integrate online results with the laboratory exploration that was conducted.

**Limitations**

In the ligand docking study, it is difficult to draw conclusions on the role of the substrates for all three AKRs because PyMOL only shows the ligand location where it binds. Even though location of the ligands in relation to the protein structure is important for showing associations, it does not generate clear quantitative data on whether the molecules bind to the AKRs. Unlike x-ray crystallography, PyMOL does not provide more definitive information on the exact structure and location of the AKR active sites.

Moreover, the enzyme kinetics research was limited by a lack of tested samples and by a lack of time to run the spectrophotometer for each sample. More samples with varied concentrations is necessary to provide stronger evidence of rate and intensity of catalytic activity. Although five minutes provided a good baseline for studying catalytic activity, a longer timeframe may reveal changes in the trends that were observed.

**Implications**

*S. cerevisiae* is an organism essential to fermentation processes. In biofuel production, readily available and renewable lignocellulosic materials are treated with dilute acids at high temperatures for ethanol fermentation, which results in the release of various harmful compounds (32). Vanillin, one of the released compounds, is toxic to microorganisms and inhibits fermentation. *S. cerevisiae* reduces vanillin to vanillyl alcohol, a less toxic substance. The reduction to vanillyl alcohol allows for greater *S. cerevisiae* lifespan and function in fermentation (22). Even though there was no resulting increase in vanillin reduction activity in our experiments, AKR5F’s overall cumulative contribution to vanillin detoxification should not be underestimated as it is found to respond to DNA damage (22).

AKR5F is orthologous to human aldose reductase, AKR1B1, as the two sequences share a 30–40% amino acid similarity (33). AKR1B1 is the most studied AKR in the human body; the protein is associated with diabetes mellitus, as the inhibition of the protein mitigates diabetic complications. Moreover, the AKR is related to inflammation-related diseases such as sepsis due to the regulation of oxidative stress-induced inflammation. AKR1B1 can also regulate the development of human cancers through function in lipid metabolism and detoxification of reactive carbonyls. The protein has also been found in several human tumors (34). Therefore, studying these AKRs, their structures, and their reactivity with substrates can have applications in modern medicine.

**Areas of Future Research**

AKR 308 has practical applications in the reduction of compounds, such as methylglyoxal. With further study, more can be learned about the molecular structure of AKR 308 and its interactions with various substrates. Such methods can be applied to AKR 308 homologs to understand chemical reactions that occur in other proteins and AKRs. This can aid in the discovery of other AKR uses and the relationship between AKRs in several organisms.

**REFERENCES**


## Appendix A: Sequence homology results

<table>
<thead>
<tr>
<th>AKR Name</th>
<th>Protein Sequence</th>
<th>Homologous Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKR 169</td>
<td>MVPKFYKLSNGFKIPSIALGTGYDIPR SQTAEIVYEYGKCGYRHFDTAVLY GNEKEVGDGIKWLNEDPGNHKRE EFIYTTKLNQSNGYKRAKAARQ CLNEVSGLQYIDLILLIPSLEGSKLR LETWRAMQEAVIDEGLVKSIGNVSN GKKHIDELNLWPELKKPVVNVQIEI SPWiMRQELADYCKSKGL- WEAFAPLCHGKMTNPDDLLKVCK EVDRNPQVLRWSLQHQGYPPLPNT KTVKRLEGNLAYNFELSDEQMKFLDHPDAYEPDWDCTEAD</td>
<td>MVPKFYKLSNGFKIPSIALGTGYDIPR SQTAEIVYEYGKCGYRHFDTAVLY GNEKEVGDGIKWLNEDPGNHKRE EFIYTTKLNQSNGYKRAKAARQ CLNEVSGLQYIDLILLIPSLEGSKLR LETWRAMQEAVIDEGLVKSIGNVSN GKKHIDELNLWPELKKPVVNVQIEI SPWiMRQELADYCKSKGL- WEAFAPLCHGKMTNPDDLLKVCK EVDRNPQVLRWSLQHQGYPPLPNT KTVKRLEGNLAYNFELSDEQMKFLDHPDAYEPDWDCTEAD</td>
</tr>
</tbody>
</table>

| AKR 304 | MSSLVTLNGLKMLPVLGLGCWKID KKCVCANQYEAIKLYRLFDGACD YGNEKEVGEGRKIASEGVSRLKDIF VSVKLWNNFHHPDHVKLAKKTL DMLGDLNYLDYYHFPIAFKYPFEE KYPGFTGADDEKKGHITAEHPI IDTYRALEECVDEGLIKSIGSVNSFG SLIQDLLRGRCKPVALQIEHHIPYLT QEHLEFCKNLQIQVAYSSFQPS FIEMDLQAKTPTTFENDVIKKVS QNHGSTTSQVLLRWTQRGIAVIP KSSKERRLANLEIKKFLTEQEL KDISALNANIRFNPWTWLDGKFTP FAWSSQ | MSSLVTLNGLKMLPVLGLGCWKID KKCVCANQYEAIKLYRLFDGACD YGNEKEVGEGRKIASEGVSRLKDIF VSVKLWNNFHHPDHVKLAKKTL DMLGDLNYLDYYHFPIAFKYPFEE KYPGFTGADDEKKGHITAEHPI IDTYRALEECVDEGLIKSIGSVNSFG SLIQDLLRGRCKPVALQIEHHIPYLT QEHLEFCKNLQIQVAYSSFQPS FIEMDLQAKTPTTFENDVIKKVS QNHGSTTSQVLLRWTQRGIAVIP KSSKERRLANLEIKKFLTEQEL KDISALNANIRFNPWTWLDGKFTP FAWSSQ |

| AKR 308 | MSSSVASTENIVENLHPKTEIYFS LNNGVRIPALGLGTANPEHALET QAVKAAIAKAGYRHIDTAWAYETEP FVGEAIKELLEDGSIREDLFITTKW VPIWLDEVRSLNKLAKALGLEYVDDLQLHWPCLFEKIKDKPGISGLVTTPVD- DSGRTMYAADDGYLETYKQLEK XLDNPDHRVAIGVSNFSIERYLERLIK ECVRVKPTVQNVETFHPHLQPMLRKF CFMDHIDILTASPYLGLSHGAPNKLIP LVKLAEKYNVTGNDLHIYHRQ TIVIPRSSNLNPRISSSSIEASLKDELQ ELNDFGEKYPVFIDEPAAILPE | MSSSVASTENIVENLHPKTEIYFS LNNGVRIPALGLGTANPEHALET QAVKAAIAKAGYRHIDTAWAYETEP FVGEAIKELLEDGSIREDLFITTKW VPIWLDEVRSLNKLAKALGLEYVDDLQLHWPCLFEKIKDKPGISGLVTTPVD- DSGRTMYAADDGYLETYKQLEK XLDNPDHRVAIGVSNFSIERYLERLIK ECVRVKPTVQNVETFHPHLQPMLRKF CFMDHIDILTASPYLGLSHGAPNKLIP LVKLAEKYNVTGNDLHIYHRQ TIVIPRSSNLNPRISSSSIEASLKDELQ ELNDFGEKYPVFIDEPAAILPE |
Appendix B: PyMOL visualizations of 3-dimensional protein structures

a) Model of the hypothetical oxidoreductase *C. elegans* protein that was found to be analogous to AKR 169 and AKR 304
b) Ara1p model analogous to AKR 308 with NADPH in gray.

Appendix C: Methylglyoxal in glycolysis

*Hogan, Biermann, Demers (2020) Fig. 1*