

Structures of alcohol dehydrogenases from *Ralstonia* and *Sphingobium* spp. reveal the molecular basis for their recognition of ‘bulky-bulky’ ketones

Henry Man,^[a] Kinga Kędziora,^[b] Justyna Kulig,^[c] Annika Frank,^[a] Iván Lavandera,^[b]
Vicente Gotor-Fernández,^[b] Dörte Rother,^[c] Sam Hart,^[a] Johan P. Turkenburg^[a] and Gideon
Grogan*^[a]

^[a] Department of Chemistry, University of York, Heslington, York, YO10 5DD U.K

^[b] Universidad de Oviedo, Facultad de Química, Química Orgánica e Inorgánica,

Julian Clavería 8, C.P. 33006 Oviedo, Asturias, Spain

^[c] Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich
GmbH, 52425 Jülich, Germany

*Author to whom correspondence should be addressed

Tel: 44 1904 328256

Fax: 44 1904 328266

e-mail: gideon.grogan@york.ac.uk

Abstract

Alcohol dehydrogenases (ADHs) are applied in industrial synthetic chemistry for the production of optically active secondary alcohols. However, the substrate spectrum of many ADHs is narrow, and few, for example, are suitable for the reduction of prochiral ketones in which the carbonyl group is bounded by two bulky and/or hydrophobic groups; so-called 'bulky-bulky' ketones. Recently two ADHs, RasADH from *Ralstonia* sp. DSM 6428, and SyADH from *Sphingobium yanoikuyae* DSM 6900, have been described, which are distinguished by their ability to accept bulky-bulky ketones as substrates. In order to examine the molecular basis of the recognition of these substrates the structures of the native and NADPH complex of RasADH, and the NADPH complex of SyADH have been determined and refined to resolutions of 1.5, 2.9 and 2.5 Å respectively. The structures reveal hydrophobic active site tunnels near the surface of the enzymes that are well-suited to the recognition of large hydrophobic substrates, as determined by modelling of the bulky-bulky substrate *n*-pentyl phenyl ketone. The structures also reveal the bases for NADPH specificity and (*S*)-stereoselectivity in each of the biocatalysts for *n*-pentyl phenyl ketone and related substrates.

Introduction

Alcohol dehydrogenases (ADHs) have been employed for many years in both academic and industrial groups for the asymmetric reduction of prochiral carbonyl groups to optically active secondary alcohols [1,2]. The natural diversity of ADHs has ensured that enzymes encompassing a broad range of catalytic characteristics have been made available for applications. Hence there are ADHs that offer (*R*)- [3] or (*S*)-[4] selectivity, thermostability [5,6] and an impressive tolerance to organic solvents that allows the enzymes to be used in the presence of high substrate concentration [4] or even neat substrate itself [7]. One such niche application of ADHs is served by a subgroup of these enzymes that transforms prochiral ketones that feature large hydrophobic groups on either side of the carbonyl group; ‘bulky-bulky’ ketones.

As part of a culture collection screening programme, Kroutil and co-workers described the cloning, expression and application of an ADH from *Ralstonia* sp. DSM 6428, which was able to accept these bulky-bulky ketones as substrates [8]. The *Ralstonia* ADH (RasADH from hereon) enzyme was applied to the reduction of *n*-butyl phenyl ketone and *n*-pentyl phenyl ketone **1** with (*S*)-enantioselectivity and with enantiomeric excesses (e.e.s) of up to >99% (**Figure 1**) [8]. The excellent enantioselectivity was also extended to bulky-small ketones such as acetophenone. RasADH has subsequently been characterised in some detail by one of our groups [9]. The enzyme was reported to be of the short chain dehydrogenase (SDR) family, with a subunit molecular weight of 26.7 kDa (249 amino acids) and it was suggested that three subunits of the enzyme associated to form a trimer in solution. RasADH exhibited a broad pH optimum for the reductive transformation of benzaldehyde and had a half-life of 80 h at 25°C. The stability of the enzyme could be augmented by the addition of calcium ions. RasADH was subsequently also applied to the reduction of α -hydroxy ketones

[10] and in the dynamic kinetic resolution (DKR) of α -alkyl- β -keto esters such as **3** (**Figure 1**) to give (2*S*, 3*S*)- products of type **4** [11].

Further screening of microbial culture collections for the reduction of bulky-bulky ketones revealed a strain of *Sphingobium yanoikuyae* DSM 6900 that was also able to reduce *n*-pentyl phenyl ketone **1** to the (*S*)-alcohol **2** in up to 97% e.e. [12]. This short chain dehydrogenase, named SyADH, was isolated and the relevant gene cloned and expressed, and subsequently used for the non-selective oxidation of some small prochiral alcohols [13]. Like RasADH, SyADH has also recently been employed in the DKR of α -alkyl- β -keto esters such as **5** [11], in this case to give (2*R*, 3*S*)- diol products **6** with excellent d.e.s and e.e.s

RasADH and SyADH therefore present distinctly useful biocatalysts as they catalyse the transformation of sterically challenging substrates and with, in some cases, complementary stereo- and diastereo-selectivities to established ADHs. In the case of each of these enzymes, knowledge of their three-dimensional structure would not only provide information for the first time on the determinants of bulk-bulky ketone recognition in such ADHs, but also serve to inform the engineering of other ADHs for transformation of those substrates, or for the expanded substrate specificity of RasADH and SyADH themselves. In this report, we present the X-ray crystal structures of the *apo*- and NADPH-complex of RasADH at 1.5 Å and 2.9 Å resolution respectively, and also the structure of the NADPH complex of SyADH at 2.5 Å resolution. These structures reveal the hydrophobic characteristics of the active site(s) that allow accommodation of large hydrophobic substrates, and also shed light on the determinants of cofactor specificity in these 'bulky-bulky' ADH enzymes.

Results and Discussion

Structure of RasADH

RasADH was expressed from a strain of *E. coli* BL21 (DE3) that had been transformed with the gene encoding RasADH ligated into the pET-YSBLIC-3C vector [14]. After purification using nickel affinity and gel filtration chromatography, the pure protein was initially concentrated to 8 mg mL⁻¹. During this process problems with precipitation were encountered, but these were successfully addressed through the inclusion of Ca²⁺ ions, 500 mM sodium chloride and glycerol in the cell resuspension and protein purification buffers (see Experimental Section). Gel filtration studies on the protein derived in this way were suggestive of a tetramer of RasADH monomers in solution, rather than the trimer suggested by earlier studies [9].

Crystals of RasADH were obtained in two forms, each of which was soaked with 10 mM NADPH before testing diffraction quality. It was found that the first form diffracted to a resolution of 1.5 Å with a C₂ space group. Data collection and refinement statistics can be found in **Table 1**. The structure of RasADH was solved by molecular replacement using a monomer of a 'probable dehydrogenase protein' from *Rhizobium etli* CFN42 (55% sequence identity with RasADH; PDB code 4FGS) as a model. In the structure solution there were four monomers A-D in the asymmetric unit, which made up a tetramer (**Figure 2a**), a quaternary association common to some other short-chain ADHs such as the carbonyl reductase from *Candida parapsilosis* (PDB code 3CTM, [15]) and levodione reductase from *Corynebacterium aquaticum* (1IY8, [16]).

Each monomer of RasADH (**Figure 2b**) was made up of a classical Rossmann fold with a central β-sheet composed of seven β-strands: β1 (residues 8-13); β2 (33-37); β3 (54-58); β4 (83-87); β5 (131-135); β6 (174-180) and β7 (239-242). The β-sheet was surrounded by six alpha helices: α1 (residues 17-30); α2 (40-50); α3 (64-77); α4 (101-126); α5 (148-168) and

$\alpha 7$ (217-228). A further helix, the sixth in sequence order $\alpha 6$ (195-208, as revealed by the NADPH complex structure – *vide infra*), was largely missing from the density in the *apo* structure of RasADH. This helix was part of a sequence of residues in the region of Thr185 to Phe205 that in each subunit could not be modelled (although residues Glu200 to Lys204 were present in subunit ‘C’). A possible role for this helix is discussed below. Notwithstanding helix $\alpha 6$, the overall fold of the RasADH monomer (**Figure 2b**) is well-conserved amongst SDR structures; a superimposition of the ‘A’-chain of RasADH with that of a monomer of 4FGS gave an r.m.s.d. of 0.71 Å over 240 C α atoms. The RasADH monomer is most similar to 4FGS and a putative oxidoreductase protein from *Sinorhizobium meliloti* 1021, (PDB code 4ESO; with 41% sequence identity to RasADH), as determined by analysis using the DALI server [17], with r.m.s.d values for the latter structure 1.4 Å over 250 residues. In this first, higher resolution, structure of RasADH no residual density in the omit map corresponding to the NADP(H) cofactor could be observed in the putative active site of the enzyme.

A monomer of RasADH (for example, subunit ‘A’) closely associates with its neighbours ‘C’ and ‘B’, and to a lesser extent, ‘D’, to form the tetramer. Subunit ‘A’ makes a contact surface area of 1540 Å² with subunit ‘C’, as calculated using PISA [18], with interactions dominated by the reciprocal association of helices $\alpha 7$, including stacking interactions between Phe226 residues of each subunit, and strands $\beta 7$, which are largely governed by hydrophobic side-chain interactions. ‘A’ also closely interacts with subunit ‘B’, with a calculated shared interface of 1670 Å². This pair of monomers interacts most closely through reciprocal interactions between helices $\alpha 4$ and $\alpha 5$, including the indole nitrogen of Trp164 of ‘A’ with

the carboxylate side chain of 'B' Asp148 and a salt bridge between the side-chains of 'A' Asp105 and 'B' Arg 113.

The other crystal form of RasADH diffracted more poorly, to a resolution of 2.9 Å, and in the same $C2_1$ space group as the *apo*-enzyme. Using the monomer of the *apo*-structure as a model, the structure solution in this instance yielded six monomers in the asymmetric unit, representing one full tetramer and one half-tetramer. In this case, each monomer featured extensive density in the omit map in the active site that was representative of the cofactor NADPH at full occupancy. In addition, density for the helix α_6 in the region of Thr185 to Phe205 was now largely complete, and could be modelled, save for one or two residues in the region of Val191. Chain 'F' was complete, however. The loop density in the NADPH complex confirmed that the loop appears to act as lid for the active site, closing over the cofactor in the *holo*-form of the enzyme (**Figure 2b**).

Structure of SyADH

SyADH was also subcloned into the pET-YSBLIC-3C vector, and was again expressed well in the soluble fraction of the BL21 (DE3) strain of *E. coli*. In this case however, gel filtration analysis suggested that SyADH was predominantly a dimer in solution. Crystallisation of SyADH followed cleavage of the hexahistidine tag as described in the Experimental Section. Crystals were of the $P2_1$ space group. The structure was solved using the programme BALBES [19], which selected a monomer of the NADPH-dependent blue fluorescent protein from *Vibrio vulnificus* (3P19; 25% sequence identity with SyADH) as a model. The solution featured five dimer pairs in the asymmetric unit (**Figure 3a**).

The monomer of SyADH (Subunit 'A' will serve as the model for the description below), again features the characteristic Rossmann-type fold, again featuring a central β -sheet of seven strands: β 1 (residues 6-11); β 2 (31-34); β 3 (56-58); β 4 (82-88); β 5 (134-138); β 6 (177-181) and β 7 (224-225). In SyADH, the sheet is surrounded by seven α -helices: α 1 (residues 15-27); α 2 (38-52) α 3 (66-78); α 4 (102-130); α 5 (151-171) α 6 (190-194) and α 7 (207-220). A last helix, α 8, stretching from residues 234-238, is at the C-terminus of the protein and participates in pronounced monomer-monomer ('A'/'B') interactions, helping to form the dimer (**Figure 3b**). Electron density allowed the modelling of residues Thr3 or Leu 4 through Tyr258 in all subunits, with only residue Asn199 not modeled in subunits 'I' and 'J', owing to poor electron density. The monomer of SyADH was most similar to monomers of clavulanic acid dehydrogenase from *Streptomyces clavuligerus* (PDB code 2JAH; 26% sequence identity with SyADH; r.m.s.d 2.2 Å over 245 C α atoms) [20] and sepiapterin reductase from *Chlorobium tepidum* (PDB code 2BD0; 25% sequence identity with SyADH; r.m.s.d 2.4 Å over 240 C α atoms) [21]. The overall structural fold of the monomer of SyADH was very similar to that of 2JAH, save for a loop region Phe218-Leu223 between α 7 and β 7 in SyADH which was longer in 2JAH (Ala227 –Val236). Monomer 'A' was calculated to form an interface of 3700 Å² with its dimer partner, subunit 'B', as calculated using PISA [18]. The dimer is stabilised by reciprocal interactions between residues in helices α 8, including a salt bridge between the side-chain of residues Asp248 and Arg234 in each subunit, but also reciprocal interactions between α 4 and α 5, which are structurally homologous with helices α 4 and α 5 in RasADH, described above.

The substrate binding sites

The active site of RasADH is a hydrophobic tunnel that is formed at the top of the central β -sheet of the Rossmann fold as seen in **Figure 2b** and **Figure 4**, and is partially covered by helix α_6 in the NADPH-complex structure. The nicotinamide ring of NADPH sits at the base of this tunnel, lined by residues Tyr150 (the likely proton donor to a nascent alcohol in the reductive ADH-catalysed reaction [22]), Ser137, Phe205, Leu144, Leu142, Leu201, Val138, Ile91, His147 and Gln191. The strict specificity for the phosphorylated cofactor NADPH is governed by interactions of adjacent arginine residues Arg38 and Arg39 and Asn15, with Arg38 also participating in π -stacking interactions with the adenine ring. The interactions with the phosphate explain the preference of RasADH for NADPH over the non-phosphorylated cofactor NADH.

The active site tunnel in SyADH is situated between two loops formed by Gly195-Val204 and Phe144-Met150, the former in place of helix α_6 in RasADH. The nicotinamide ring again occupies the base of the tunnel, the lining of which shares significant homology with RasADH, with probable proton donor Tyr153 (Ras-ADH-Tyr150), Ser140 (Ser137), Met150 (His147), and Val141 (Val138), conserved in respect of steric bulk, but with three larger hydrophobic residues from RasADH replaced by alanine: Ala92 (equivalent to RasADH Ile91) and Ala145 (Leu142) and Ala194 (Gln191), suggestive, at least superficially, of a greater active site volume in SyADH overall. Additional steric bulk is provided in the active site by Trp191 (Ile188) and Phe148 (Leu144). The 2'-ribose phosphate group of NADPH is less constrained in SyADH than in RasADH; Arg36 mirrors the location and function of RasADH-Arg38, but RasADH-Arg39 is not present and is replaced by an aspartate (Asp37) that points away from the phosphate binding site. Ser13 replaces RasADH Gln15 in the other direct interaction with phosphate. The side-chain of an additional arginine, Arg40, which is not present in RasADH is close to the phosphate in SyADH, but, with both terminal N-atoms

of the side chain at 3.6 Å and 3.8 Å from the phosphate oxygen, too far away to suggest a major role in phosphate binding.

The nicotinamide rings of the NADPH cofactors in each case are presented to the active site cavity. The C4 atom of the nicotinamide ring, from which hydride will be delivered to the substrate carbonyl, is in each case 4.2 Å (RasADH) or 4.3 Å (SyADH) from the phenolic hydroxyl of the tyrosine proton donor. In SyADH, the region extending from the space between the nicotinamide C4 and the phenolic hydroxyl of Tyr153, and out into the hydrophobic cavity, contained electron density although no obvious ligands that featured in the crystallization conditions could be modelled.

Modelling bulky-bulky ketones in the active sites

RasADH and SyADH are distinctive in their ability to reduce ketones in which the prochiral carbonyl group is bounded on either side by hydrophobic groups, such as *n*-pentyl phenyl ketone **1**, which is transformed by each enzyme into the (*S*)-alcohol with excellent enantiomeric excess, thus exhibiting a Prelog selectivity [23]. With the structures of the enzymes in hand, it was now possible to model the structure of **1** into their active sites, in the knowledge that, given the known (*S*)-stereospecificity of the enzymes with this substrate, only the *re*-face of the prochiral ketone would be presented to the C4-nicotinamide atom that delivers hydride, and also that the carbonyl group must be situated within H-bonding distance of the relevant conserved serines (Ser137 in RasADH and S140 in SyADH) and tyrosine proton donors (Tyr150 in RasADH and Tyr153 in SyADH) that are involved in catalysis [22]. **1** was modelled into the active sites of RasADH and SyADH using the programme

AUTODOCK-VINA [24] applying a procedure described in the Experimental Section. A comparison of these models is shown in **Figure 5**.

In the model of SyADH with substrate **1**, the hydrophobic tunnel can be described in terms of a proximal region nearer the surface of the enzyme, in which the phenyl ring is bound, and a distal region towards the enzyme interior, which binds the alkyl chain. The *re*-face of the carbonyl in **1** is situated against the nicotinamide ring of NADPH, with the C=O oxygen positioned between the side chains of putative proton donor Tyr153 and Ser140. As a consequence, the phenyl ring is accommodated in the proximal region of the tunnel, bounded by hydrophobic residues Val141, Ala184, Leu201 and Phe148. The pendant *n*-pentyl group penetrates the distal region of the tunnel and is bounded by further hydrophobic residues Trp191, Ile190, Ala92 and Met150. In RasADH, the requisite interactions of the carbonyl group with the C4 of NADPH and the side chains of Tyr150 and Ser137 are still observed, but the phenyl ring of the substrate is pressed further into the active site in relation to the SyADH model complex, as a result of steric repulsion by residue Phe205, which is part of the 'lid' helix α_6 . In the distal region of the hydrophobic tunnel, penetration of the *n*-pentyl chain is prevented by Gln191 and is displaced into the region of the active site occupied by Ile188, which would not be permitted in SyADH because of the presence of bulky Trp191 (seen behind the substrate in **Figure 5a**). The alkyl chain now makes favourable hydrophobic interactions in this position with Phe205 and Ile188 (seen behind the substrate in **Figure 4b**). The models suggest that although each accepts **1**, SyADH is better suited to accommodate the substrate. The models are useful in revealing the hydrophobic characteristics of the active sites that permit binding of hydrophobic substrates but also the distinctive substrate ranges of SyADH and RasADH. The model suggest that while the flexible alkyl chain of **1** can be

accommodated in the distal region of hydrophobic tunnel through contortion in RasADH, less flexible substituents, such as aromatics, would not be accommodated. This is supported by substrate specificity studies [9], which show that compounds such as benzoin are poor substrates for RasADH. For SyADH, however, the larger distal portion of the hydrophobic tunnel may allow for the accommodation of such bulky groups. The large capacity of the hydrophobic tunnel of SyADH also suggests a basis for its non-selectivity towards smaller prochiral ketones such as acetophenone [13] with space available for these substrates to be accommodated in such a way that presents the *si*- or the *re*- face of the carbonyl group to the NADPH hydride at C4 of the cofactor nicotinamide ring.

Interestingly, from **Figure 5** the models suggest that, in both cases, NADPH provides its pro-*S* hydride to attack the *re*-face of the carbonyl moiety of *n*-pentyl phenyl ketone, suggesting that the reaction proceeds by the ‘E4 pathway’ [25]. Whilst the confirmation of pro-*S* hydride transfer in RasADH and SyADH awaits analysis using experiments with C4-deuterated cofactors, the structural observation is in agreement with the observed stereospecificity of hydride delivery by other short chain ADHs with Prelog selectivity [26, 27]. Other ADHs or Prelog selectivity such as yeast [YADH, 28], horse liver [HLADH, 29], *Rhodococcus erythropolis* [30], or *Thermoanaerobacter ethanolicus* [31] ADHs, supply the NADPH pro-*R* hydride, thus following the E3 pathway [25], but in these cases the enzymes are of the zinc-dependent Medium Chain Reductase (MDR) family, in which the opposite face of the nicotinamide ring (of NADH in these cases) is presented to the active site, through rotation around the bond between the nicotinamide ring and the ribose, relative to the conformation observed in the SDR, and as seen in PDB structures 2HCY (YADH, [32]) and 1LDE (HLADH).

Conclusion

The structures of RasADH and SyADH have revealed the structural basis for the recognition of bulky-bulky ketones in a hydrophobic tunnel near the surface of the enzymes. The structures of enzymes with potential use in biocatalytic applications can be valuable in identifying active site determinants of substrate specificity and enantioselectivity that may inform rational engineering, or randomised mutagenesis at targeted residues, for improved or altered specificity.

Experimental Section

Gene cloning, expression and protein purification

The genes encoding RasADH and SyADH were obtained from the laboratory of Professor Wolfgang Kroutil, University of Graz, Austria. Each gene was sub-cloned into the pET-YSBLIC-3C vector in York. RasADH was amplified by PCR from the template plasmid using the following primers: 5'-CCAGGGACCAGCAATGTATCGTCTGCTGAATAAAACCGCAGTTATTACCG-3' (Forward) and 5'-GAGGAGAAGGCGCGTTATTAAACCTGGGTCAGACCACCATCAACAAACAG-3' (Reverse). SyADH was amplified using primers 5'-CCAGGGACCAGCAATGACCACCCTGCCGACCGTTCTG-3' (Forward) and 5'-GAGGAGAAGGCGCGTTATTATTATTTTCAAACCTGCGGATGTGACCATGC-3' (Reverse). Following agarose gel analysis of the PCR products, the relevant bands were eluted from the gels using a PCR Cleanup kit® (Qiagen). The genes were then sub-cloned into the pET-YSBL-LIC-3C vector using previously published techniques [14]. The resultant plasmids were then used to transform *E. coli* XL1-Blue cells (Novagen), yielding colonies

which in turn gave plasmids using standard miniprep procedures that were sequenced to confirm the identity and sequence of the genes.

The recombinant vector(s) containing the RasADH and SyADH genes were used to transform *E. coli* BL21 (DE3) cells using 30 $\mu\text{g mL}^{-1}$ kanamycin as antibiotic marker on Luria Bertani (LB) agar. Single colonies from agar plate grown overnight were used to inoculate 5 mL cultures of LB broth, which were then grown overnight at 37°C with shaking at 180 r.p.m. The starter cultures served as inocula for a 1 L cultures of LB broth in which cells were grown until the optical density (OD_{600}) had reached a value of 0.8. Expression of either RasADH or SyADH were then induced by the addition of 1 mM isopropyl β -D-1-thiogalactopyranoside (IPTG). The cultures were then incubated at 18°C in an orbital shaker at 180 r.p.m. for approximately 18 h. The cells were harvested by centrifugation for 15 min at 4225 g in a Sorvall GS3 rotor in a Sorvall RC5B Plus centrifuge. At this point, considerations specific to each protein applied and purification strategies diverged.

For RasADH, cell pellets were resuspended in 20 mL of 50 mM Tris-HCl buffer pH 7.5 containing 500 mM NaCl, 20 mM imidazole, 1 mM CaCl_2 and 10% glycerol (v/v) (buffer 'A') per 1 L of cell culture. The cell suspensions were then sonicated for 10 x 45s bursts at 4°C with 30 s intervals. The soluble and insoluble fractions were separated by centrifugation for 30 min at 26,892 g in a Sorvall SS34 rotor. The crude cell lysate from 1L cell culture was loaded onto a 5 mL HisTrap FF crude column (GE Healthcare), which was then washed with buffer 'A') and then the protein eluted with a linear gradient of 20-300 mM imidazole over 20 column volumes at a flow rate of 2.5 mL min^{-1} . Fractions containing RasADH were combined and concentrated using a 10 kDa cut-off Centricon® filter membrane. The

concentrated RasADH was then loaded onto a pre-equilibrated S75 Superdex™ 16/60 gel filtration column, which was then eluted with 120 mL of buffer 'A' without imidazole, using a flow rate of 1 ml min⁻¹. Fractions containing pure RasADH, as determined by SDS-PAGE analysis were combined and stored at 4°C. For SyADH, an equivalent procedure was employed using, in place of buffer 'A', a 50 mM Tris-HCl buffer containing 300 mM NaCl and 20 mM imidazole only (buffer 'B'). All subsequent procedures were equivalent to the strategy employed for RasADH, with purification performed using Ni-NTA chromatography followed by gel filtration. In the case of SyADH, prior to crystallisation, the histidine tag was cleaved using C3 protease using a procedure described previously [14].

Protein Crystallisation

Crystallization conditions for both enzymes were determined using commercially available screens in the sitting-drop format in 96-well plates using 300 nL drops (150 nL protein plus 150 nL precipitant solution). Positive hits were scaled up in 24 well Linbro dishes using the hanging-drop method of crystallisation, with crystallisation drops containing 1 µL of protein solution and 1 µL of precipitant reservoir. For the apo-RasADH, the best crystals were obtained in 0.1M Bis-tris propane pH 7.0 containing 20% (w/v) PEG 3350 and 0.02 M sodium-potassium phosphate. The protein concentration was 24 mg mL⁻¹. For the Ras-ADH NADPH complex, the best crystals were obtained in 0.1M Bis-tris propane pH 8.0, containing 16% (w/v) PEG 3350, 0.2M potassium isothiocyanate and 5% (w/v) ethylene glycol. The protein concentration in each case was 24 mg mL⁻¹. In each case, the crystals were picked and transferred to a solution of the mother liquor containing 20% (v/v) ethylene glycol as cryoprotectant and 10 mM NADPH, and left for 5 min. The crystals were then flash cooled in liquid nitrogen prior to diffraction analysis.

For the SyADH-NADPH complex, the best initial crystals were obtained in conditions containing 0.1M MES buffer pH 5.5 with 0.3 M sodium acetate and 15% (w/v) PEG 4K, with a protein concentration of 20 mg mL⁻¹. As with RasADH, crystals were picked and transferred to a solution of the mother liquor containing 20% (v/v) ethylene glycol as cryoprotectant and 10 mM NADPH, and left for 5 min. The crystals were then flash-cooled in liquid nitrogen prior to diffraction analysis.

Crystals were tested for diffraction using a Rigaku Micromax-007HF fitted with Osmic multilayer optics and a Marresearch MAR345 imaging plate detector. Those crystals that diffracted to greater than 3 Å resolution were retained for full dataset collection at the synchrotron.

Data Collection, Structure Solution, Model Building and Refinement of Q1EQE0

Complete datasets for the *apo*-RasADH and its NADH complex, and the SyADH complex with NADPH were collected on beamlines I04, I02 and I04 respectively, at the Diamond Light Source, Didcot, Oxfordshire, U.K. Data were processed and integrated using XDS [33] and scaled using SCALA [34] as part of the Xia2 processing system [35]. Data collection statistics are given in **Table 1**. The structure of the *apo*-RasADH was solved using MOLREP [36], using a monomer of structure PDB code 4FGS as a search model. The solution contained four molecules in the asymmetric unit, representing one tetramer. The structure of the RasADH NADPH complex was then solved using the monomer of the *apo*-RasADH. The structure of SyADH was solved using the programme BALBES [19], which selected a monomer of the PDB code structure 3P19 as a model. The solution contained five dimers in the asymmetric unit. All structures were built and refined using iterative cycles of Coot [37] and REFMAC [38] employing local NCS restraints. For NADPH complexes of

both RasADH and SyADH, the omit maps, after building and refinement of the proteins, revealed residual density at the active sites, which could in each case be successfully modelled and refined as NADPH. The final structures of *apo*-RasADH, Ras-ADH (NADPH) and SyADH (NADPH) exhibited R_{cryst} and R_{free} values of 15.8/18.8%, 26.8/29.3% and 23.4/25.1 respectively. The structures were validated using PROCHECK [39]. Refinement statistics are presented in **Table 1**. The Ramachandran plot for *apo*-RasADH showed 97.8% of residues to be situated in the most favoured regions, 1.6% in additional allowed and 0.5% outlier residues. For the Ras-ADH-NADPH complex, the corresponding values were 95.3%, 4.6% and 0.1% respectively. For the SyADH-NADPH complex, the corresponding values were 99.0% and 1% with no residues in outlier regions. Coordinates and structure factors for *apo*-RasADH Ras-NADPH and SyADH-NADPH have been deposited in the Protein Data Bank with the accession codes 4bmn, 4bms and 4bmv respectively.

Docking

Automated docking was performed using AUTODOCK VINA 1.1.2 [25]. Structures for RasADH, SyADH were prepared using AUTODOCK utility scripts. Coordinates for substrate **1** were prepared using PRODRG [40]. A monomer model was used for RasADH and a dimer model was used for SyADH with the appropriate pdbqt files prepared in AUTODOCK Tools. The active site of RasADH and SyADH was contained in a grid of 20 x 30 x 20 and 24 x 18 x 14 respectively with 0.375Å spacing, centred around the catalytic centre which was generated using AutoGrid in the AUTODOCK Tools interface. The number of runs for genetic algorithm was set to 10 and the rest of the docking parameters were set to default parameters. The dockings were performed by VINA, therefore the posed dockings were below 2Å rmsd. The results generated by VINA were visualised in AUTODOCK Tools

1.5.6 where the ligand conformations were assessed upon lowest VINA energy, but also the following criteria: The known mechanisms of short-chain ADHs [24] and experimentally-determined enantioselectivity of both RasAH and SyADH for ketone **1** dictated that only poses in which the carbonyl of **1** was observed to make appropriate interactions with the phenolic hydroxyl of the catalytic tyrosine residue (Tyr150 or Tyr153 for RasADH or SyADH respectively); the conserved active site serine [22] (Ser137 or Ser140 for RasADH or SyADH respectively) and which presented the (*re*)-face of the carbonyl to the nicotinamide ring of NADPH (resulting in the (*S*)-alcohol product), were considered.

References

- [1] Hollman F, Arends IWCE, Holtmann D (2011) Enzymatic Reductions for the Chemist. *Green Chem* 13:2285-2313.
- [2] Musa MM, Phillips RS (2011) Recent Advances in Alcohol Dehydrogenase-catalysed Asymmetric Production of Hydrophobic Alcohols. *Catal Sci Technol* 1:1311-1323.
- [3] Schlieben NH, Niefind K, Müller J, Riebel B, Hummel W, Schomburg D (2005) Atomic Resolution Structures of R-specific Alcohol Dehydrogenase from *Lactobacillus brevis* Provide the Structural Bases of its Substrate and Cosubstrate Specificity. *J Mol Biol* 349: 801-813.
- [4] Karabec M, Lyskowski A, Tauber KC, Steinkellner G, Kroutil W, Grogan G, Gruber, K (2010) Structural insights into substrate specificity and solvent tolerance in alcohol dehydrogenase ADH-'A' from *Rhodococcus ruber* DSM 44541. *Chem Commun* 46:6314-6316.

- [5] Höllrigl V, Hollmann F, Kleeb A, Buehler K, Schmid, A (2008) TADH, the thermostable alcohol dehydrogenase from *Thermus* sp. ATN1: a versatile new biocatalyst for organic synthesis Appl Microbiol Biotechnol 81:263-273.
- [6] Machielsen R, Uria AR, Kengen SWM, van der Oost J. (2006) Production and Characterization of a Thermostable Alcohol Dehydrogenase That Belongs to the Aldo-Keto Reductase Superfamily. Appl Environ Microbiol 72:233-238.
- [7] Jakoblinnert A, Mladenov R, Paul A, Sibilla F, Schwaneberg U, Ansorge-Schumacher MB, de Maria PD (2011) Asymmetric reduction of ketones with recombinant *E. coli* whole cells in neat substrates, Chem Commun 47:12230-12232.
- [8] Lavandera I, Kern A, Ferreira-Silva B, Glieder A, de Wildeman S, Kroutil W (2008) Stereoselective Bioreduction of Bulky-Bulky Ketones by a Novel ADH from *Ralstonia* sp J Org Chem 6003-6005.
- [9] Kulig J, Frese A, Kroutil W, Pohl M, Rother D. (2013) Biochemical characterization of an alcohol dehydrogenase from *Ralstonia* sp. Biotechnol. Bioeng. (2013) doi: 10.1002/bit.24857.
- [10] Kulig J, Simon RC, Rose CA, Husain SM, Hackh M, Ludeke S, Zeitler K, Kroutil W, Pohl M, Rother, D (2012) Stereoselective synthesis of bulky 1,2-diols with alcohol dehydrogenases. Catal. Sci. Technol. 2:1580-1589.
- [11] Cuetos, A, Rioz-Martínez A, Bisogno FR, Grischek B, Lavandera I, de Gonzalo G, Kroutil W, Gotor V (2012) Access to Enantiopure α -Alkyl- β -hydroxy Esters through Dynamic Kinetic Resolutions Employing Purified/Overexpressed Alcohol Dehydrogenases Adv Synth Catal 354: 1743-1749.

- [12] Lavandera I, Oberdorfer G, Gross J, de Wildeman S, Kroutil, W (2008) Stereocomplementary Asymmetric Reduction of Bulky–Bulky Ketones by Biocatalytic Hydrogen Transfer. *Eur J Org Chem* 2008: 2539-2543.
- [13] Lavandera I, Kern A, Resch V, Ferreira-Silva B, Glieder A, Fabian WMF, de Wildeman S, Kroutil W (2008) One-Way Biohydrogen Transfer for Oxidation of sec-Alcohols *Org Lett* 10: 2155-2158.
- [14] Atkin KE, Reiss R, Turner NJ, Brzozowski AM, Grogan G (2008) Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of variants of monoamine oxidase from *Aspergillus niger*. *Acta Crystallogr. Sect. F* 64: 182-185.
- [15] Zhang R, Zhu G, Zhang W, Cao S, Ou X, Li X, Bartlam M, Xu Y, Zhang XC, Rao Z (2008) Crystal structure of a carbonyl reductase from *Candida parapsilosis* with anti-Prelog stereospecificity, *Protein Sci* 17:412-1423.
- [16] Sogabe S, Yoshizumi A, Fukami TA, Shiratori Y, Shimizu S, Takagi H, Nakamori S, Wada M (2003) The Crystal Structure and Stereospecificity of Levodione Reductase from *Corynebacterium aquaticum* M-13. *J Biol Chem* 278:19387-19395.
- [17] Holm L, Sander C (1996) Mapping the Protein Universe. *Science*. 273:595-560.
- [18] Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372: 774-797.
- [19] Long F, Vagin AA, Young P, Murshudov GN (2008) BALBES: a Molecular Replacement Pipeline. *Acta Crystallogr Sect D Biol Crystallogr* 64:125-132.
- [20] MacKenzie AK, Kershaw NJ, Hernandez H, Robinson CV, Schofield CJ, Andersson I (2007) Clavulanic Acid Dehydrogenase: Structural and Biochemical Analysis of the Final

Step in the Biosynthesis of the β -Lactamase Inhibitor Clavulanic Acid. *Biochemistry* 46:1523-1533.

[21] Supangat S, Seo KH, Choi YK, Park YS, Son D, Han C-D, Lee KH (2006) Structure of *Chlorobium tepidum* Sepiapterin Reductase Complex Reveals the Novel Substrate Binding Mode for Stereospecific Production of l-threo-Tetrahydrobiopterin *J Biol Chem* 281:2249-2256.

[22] Filling C, Berndt KD, Benach J, Knapp S, Prozorovski T, Nordling E, Ladenstein R, Jörnvall H, Oppermann U (2002) Critical Residues for Structure and Catalysis in Short-chain Dehydrogenases/Reductases. *J Biol Chem* 277:25677-25684.

[23] Prelog V (1964) Specification of the Stereospecificity of some Oxido-Reductases by Diamond Lattice Sections. *Pure Appl Chem* 9:119-130.

[24] Filling C, Berndt KD, Benach J, Knapp S, Prozorovski T, Nordling E, Ladenstein R, Jörnvall H, Oppermann U (2002) Critical Residues for Structure and Catalysis in Short-chain Dehydrogenases/Reductases. *J Biol Chem* 277:25677-25684.

[25] Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comp Chem* 31:455-461.

[26] Bradshaw, CW, Fu, H, Shen, GJ, Wong, CH (1992) A *Pseudomonas* sp. Alcohol Dehydrogenase with Broad Substrate Specificity and Unusual Stereospecificity for Organic Synthesis. *J Org Chem* 57:1526-1532.

[27] Pennacchio, A, Giordano, A, Esposito, L, Langella, E, Rossi, M, Raia, CA (2010) Insight into the Stereospecificity of Short-Chain *Thermus thermophilus* Alcohol

Dehydrogenase Showing pro-*S* Hydride Transfer and Prelog Enantioselectivity. *Protein Pept Lett* 17:437-443.

[28] Pennacchio A, Giordano A, Pucci B, Rossi M, Raia CA (2010) Biochemical Characterization of a Recombinant Short-Chain NAD(H)-Dependent Dehydrogenase/Reductase from *Sulfolobus acidocaldarius* *Extremophiles* 14:193-204.

[29] You K-s (1984) Stereospecificity for Nicotinamide Nucleotides in Enzymic and Chemical Hydride Transfer Reactions. *Crit Rev Biochem* 17:313-451.

[30] Groman, EV, Schultz, RM, Engel, LL, Orr, JC (1976) Horse Liver Alcohol Dehydrogenase and *Pseudomonas testosteroni* 3(17) β -Hydroxysteroid Dehydrogenase Transfer Epimeric Hydrogens from NADH to 17 β -Hydroxy-5 α -androstan-3-one. An Exception to One of the Alworth-Bentley Rules. *Eur J Biochem* 63:427-429.

[31] Zelinski T, Kula M-R (1994) A Kinetic Study and Application of a Novel Carbonyl Reductase Isolated from *Rhodococcus erythropolis* *Bioorg Med Chem* 2:421-428.

[32] Musa MM, Ziegelmann-Fjeld KI, Vieille, C, Zeikus, JG, Phillips, RS (2010) In *Practical Methods for Biocatalysis and Biotransformations*, Whittall J, Sutton, P, Eds., John Wiley & Sons, Chichester, pp 284-287.

[33] Ramaswamy S, Scholze M, Plapp BV (1997) Binding of formamides to liver alcohol dehydrogenases. *Biochemistry* 36:3522-3527.

[34] Kabsch W (2010) XDS. *Acta Crystallogr Sect D Biol Crystallogr* 66:125-132.

[35] Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr Sect D Biol Crystallogr* 62:72-82.

- [36] Winter G (2010) *xia2*: an expert system for macromolecular crystallography data reduction. *J Appl Cryst* 3:186-190.
- [37] Vagin A, Teplyakov A (1997) *MOLREP*: an Automated Program for Molecular Replacement. *Appl Crystallogr* 30:1022-1025.
- [38] Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr Sect D: Biol. Crystallogr.* 60:2126-2132.
- [39] Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood method. *Acta Crystallogr Sect D Biol Crystallogr* 53:240-255.
- [40] Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291.
- [41] Schüttelkopf AW, van Aalten DMF (2004) PRODRG - a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr Sect D Biol Crystallogr* 60:1355–1363.

Acknowledgements

This research was supported by a Marie Curie Network for Initial Training fellowships to K.K., J.K. and A.F. in the project BIOTRAINS (FP7-PEOPLE-ITN-2008-238531). We would also like to thank Prof. Wolfgang Kroutil of the University of Graz for genes encoding both RasADH and SyADH.

Table 1. Data Collection and Refinement Statistics for RasADH (*apo*- and in complex with NADPH) and SyADH in complex with NADPH. Numbers in brackets refer to data for highest resolution shells.

| | RasADH (<i>apo</i>) | RasADH (NADPH complex) | SyADH (NADPH complex) |
|---|--|--|--|
| Beamline | Diamond I02 | Diamond I04 | Diamond I02 |
| Wavelength (Å) | 0.97950 | 0.97950 | 0.97949 |
| Resolution (Å) | 60.86-1.52 (1.57-1.52) | 74.54-2.89 (2.93-2.89) | 139.0 - 2.5 (2.56-2.50) |
| Space Group | $C2_1$ | $C2_1$ | $P2_1$ |
| Unit cell (Å) | a = 136.5; b = 52.5; c = 151.5; $\alpha = \gamma = 90.0$; $\beta = 116.6$ | a = 192.3; b = 135.6; c = 93.6; $\alpha = \gamma = 90.0$; $\beta = 100.1$ | a = 144.9; b = 86.8; c = 155.6; $\alpha = \gamma = 90.0$; $\beta = 106.4$ |
| No. of molecules in the asymmetric unit | 4 | 6 | 10 |
| Unique reflections | 149646 (45160) | 52465 (3880) | 133601 (10900) |
| Completeness (%) | 97.7 (96.8) | 99.2 (98.8) | 99.7 (99.8) |
| R_{merge} (%) | 0.03 (0.25) | 0.18 (0.73) | 0.14 (0.54) |
| $R_{\text{p.i.m.}}$ | 0.03 (0.23) | 0.16 (0.64) | 0.12 (0.46) |
| Multiplicity | 3.3 (3.5) | 4.2 (4.2) | 4.2 (4.3) |
| $\langle I/\sigma(I) \rangle$ | 15.6 (4.9) | 7.4 (2.0) | 8.2 (3.0) |
| $CC_{1/2}$ | 1.00 (0.95) | 0.98 (0.74) | 0.99 (0.82) |
| Overall B factor from Wilson plot (Å ²) | 24 | 41 | 43 |
| $R_{\text{cryst}}/R_{\text{free}}$ (%) | 15.8/18.8 | 26.8/29.3 | 23.4/25.1 |
| No. protein atoms | 6942 | 11081 | 18948 |
| No. water molecules | 846 | 37 | 290 |
| r.m.s.d 1-2 bonds (Å) | 0.02 | 0.01 | 0.01 |
| r.m.s.d 1-3 angles (°) | 1.69 | 1.60 | 1.60 |
| Avge main chain B (Å ²) | 20 | 35 | 35 |
| Avge side chain B (Å ²) | 23 | 36 | 37 |
| Avge water B (Å ²) | 33 | 14 | 20 |
| Avge ligand B (Å ²) | - | 32 | 30 |

Figure Legends

Figure 1. Representative asymmetric reductions of bulky-bulky ketones catalysed by alcohol dehydrogenases from *Ralstonia* sp. DSM 6428 (RasADH) and *Sphingobium yanoikuyae* DSM 6900 (SyADH).

Figure 2. Structure of RasADH. **a:** The asymmetric unit of the *apo*-enzyme contained four monomers A-D that constituted one tetramer, which is shown in ribbon format and coloured by chain. **b:** Monomer of RasADH apoenzyme (green ribbon) superimposed with monomer of RasADH *holo*-enzyme (coral ribbon) in complex with NADPH. The r.m.s.d for the two structures over 201 C-alphas was 0.42 Å. The monomer(s) display the typical Rossmann fold, with a central β -sheet surrounded by six helices, four of which, α 3, α 4, α 5 and α 6. Helix α 6, which was absent in the *apo*-enzyme structure, appears to act as a lid to the active site, closing over the NADPH molecule when cofactor is bound. The N and C termini of the *apo*-enzyme monomer are also indicated. The cofactor NADPH is shown in cylinder format with carbon atoms shown in grey.

Figure 3. Structure of SyADH. **a:** The asymmetric unit contains ten monomers that constitute five dimers shown in ribbon format and coloured by chain. **b:** Each dimer consists of two monomers, coloured green and coral, each of which exhibits the Rossmann $(\alpha\beta)_8$ fold, with eight helices (numbered α 1-4 and α 6-8 for the coral monomer; α 5 is obscured) forming the central bundle and a ninth helix α 9 that associates closely with the neighbouring monomer. The N and C termini of the coral monomer are also indicated. The cofactor NADPH is shown in cylinder format with carbon atoms shown in grey.

Figure 4. Superimposition of active sites of RasADH and SyADH illustrating side-chain components of the hydrophobic tunnel in both active sites. RasADH is shown in ribbon format in coral with side chains projecting from the backbone and the nicotinamide ring of NADPH (centre) shown (carbon atoms in coral). SyADH is shown in light blue ribbons with NADPH and side chains (carbon atoms in light blue).

Figure 5. Ketone **1** modelled into the active site of a) SyADH and b) RasADH using the programme AUTODOCK VINA [24]. In a), the carbon atoms of the side chains of SyADH are shown in light blue in cylinder format; **1** is shown in ball-and-stick format with the carbon atoms in black. In b), the carbon atoms of the side chains of RasADH are shown in coral in cylinder format; **1** is again shown in ball-and-stick format with the carbon atoms in black. Distances are given in Ångstroms.

Figure 1

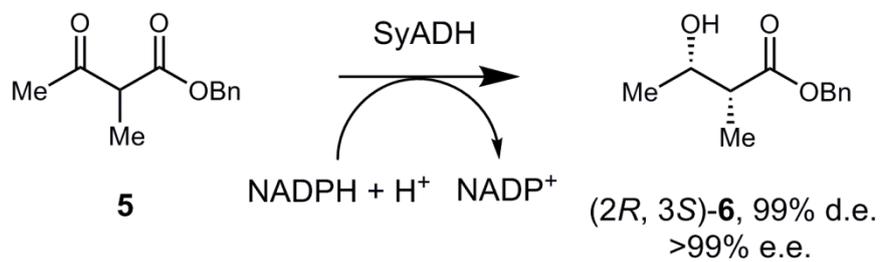
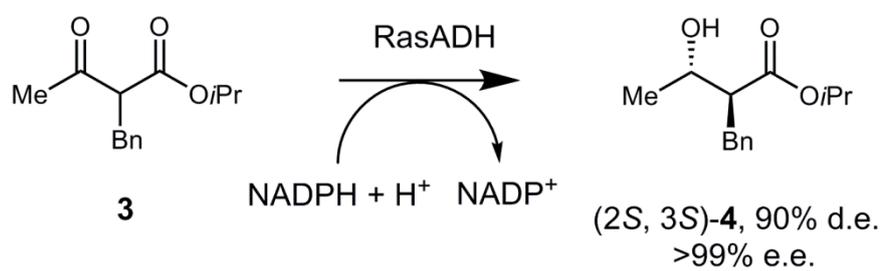
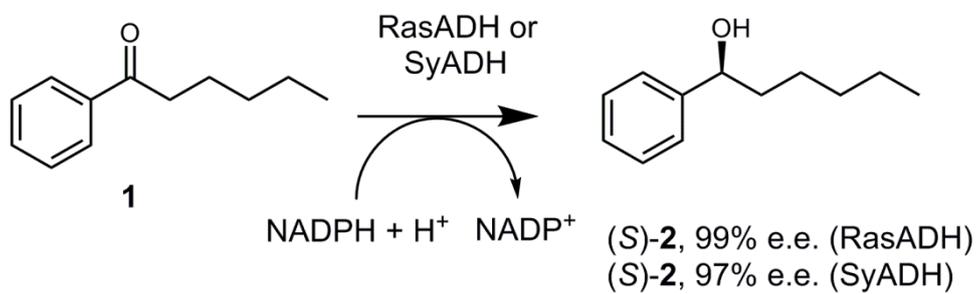
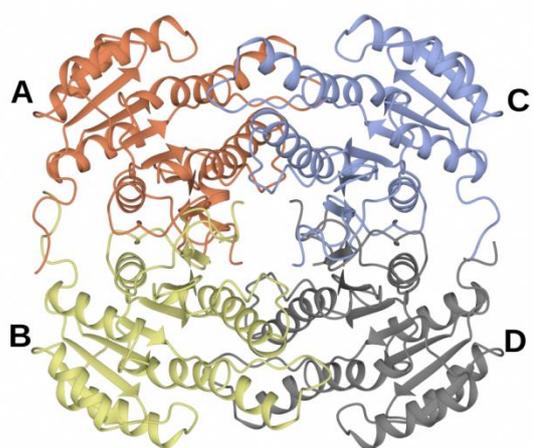


Figure 2

a



b

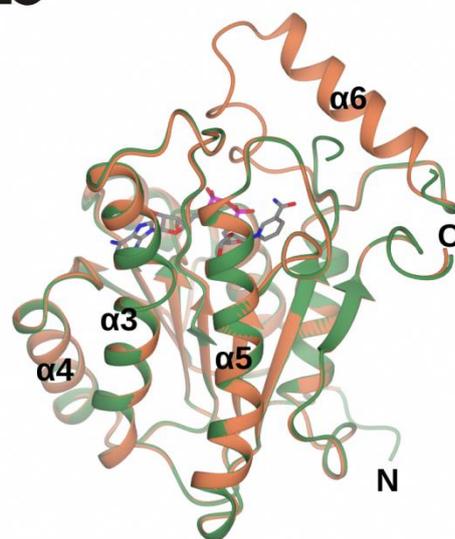
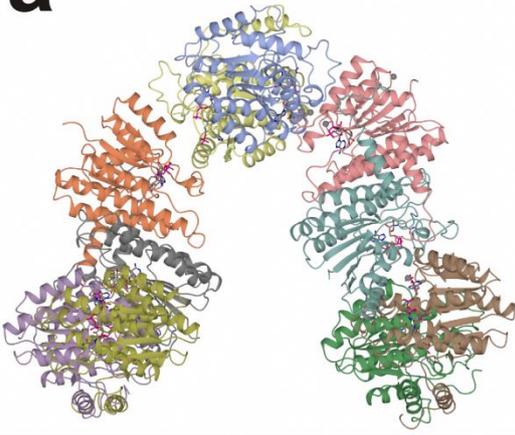


Figure 3

a



b

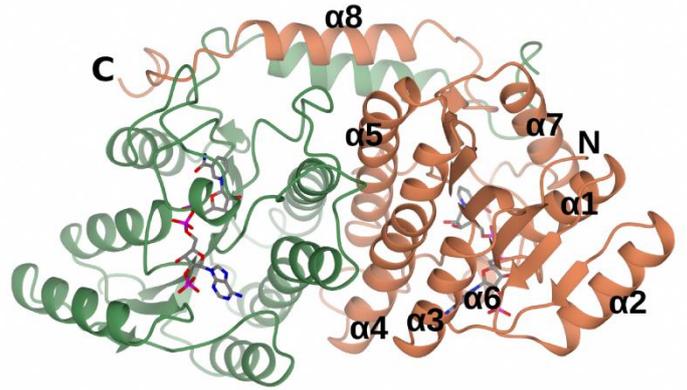


Figure 4

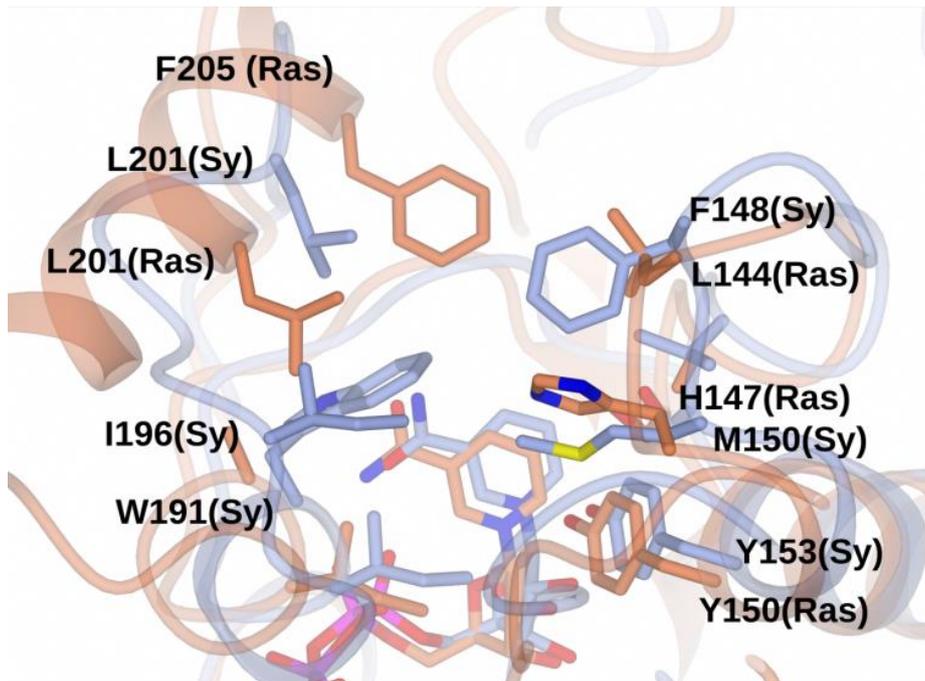


Figure 5

