



Applying structural biology to understand the reactivity of different biocatalysts



Simone Savino

**Dottorato di Ricerca in
Genetica, Biologia molecolare e Cellulare
XXX Ciclo – A.A. 2014-2017**



UNIVERSITA' DEGLI STUDI DI PAVIA
Dipartimento di Biologia e Biotecnologie
“L. Spallanzani”

Applying structural enzymology to understand the reactivity of different biocatalysts

PhD candidate Simone Savino

**Supervised by Prof. Andrea Mattevi
and Prof. Claudia Binda**

**Dottorato di Ricerca in
Genetica, Biologia molecolare e cellulare
XXX Ciclo – A.A. 2014-2017**

Acknowledgements

Even if there are many people who should be acknowledged in this thesis, I will assume most of them are perfectly aware of it. In the same way, I am just going to explicit some names, without pointing out the reasons they are mentioned for.

I would like to thank Andrea, Claudia, Federico and Francesca in the high ranks of the laboratory. Amongst my colleagues I thank Filippo, Martina and Thai, together with my student Francesca. My collaborators in Graz, and especially Alexander, deserve a remark, and the Molecular Enzymology group in Groningen must be mentioned as well. A huge thank goes also to the external reviewers of this thesis, Silvia Garavaglia and Willem van Berkel. Out of the research context, I would like to thank Foppy, Kappa and Zane. At last I thank my family and Vanessa.

Abbreviations

University of Pavia (UNIPV)

Austrian Centre for Industrial Biotechnology (ACIB)

Technical University of Graz (TU Graz)

Rijksuniversiteit Groningen (RUG)

UDP-apiose/UDP-xylose synthase (AXS)

UDP-xylose synthase (UXS)

Nicotinamide adenine dinucleotide (NAD)

Uridine diphosphate (UDP)

Glucuronic acid (GlcUA)

Apiose (Api)

Xylose (Xyl)

Lipopolysaccharide (LPS)

Rhamnogalacturonan II (RG-II)

Small-angle X-ray scattering (SAXS)

Surface Entropy Reduction prediction (SERp)

Polycyclic ketone monooxygenase (PockeMO)

Flavin adenine dinucleotide (FAD)

Nicotinamide adenine dinucleotide phosphate (NADP)

7 β -hydroxysteroid dehydrogenase (7 β -HSDH)

Ursodeoxycholic acid (UDCA)

Chenodeoxycholic acid (CDCA)

List of contents

Abbreviations	page 1
List of contents	page 2
Abstract	page 3
Chapter I: Enzymes and structural biology	page 4
I.1 Properties of enzymes	page 4
I.2 Physiological roles of enzymes	page 6
I.3 Biocatalysis	page 7
I.4 Structural enzymology	page 9
I.5 Protein crystallization	page 10
I.6 X-ray crystallography	page 17
I.7 References	page 22
Chapter II: The crystal structure of UDP-apiose/UDP-xylose synthase (AXS) enlightens the mechanism for alternative products formation	page 25
II.1 Introduction	page 25
II.2 Materials and methods	page 31
II.3 Results and discussion	page 43
II.4 Conclusions and perspectives	page 53
II.5 References	page 57
Chapter III: PockeMO: small evolutionary differences in protein folding drastically change substrate acceptance	page 61
Chapter IV: Hydroxysteroid dehydrogenases (HSDH): similar active sites can host molecules in different orientations to modify stereoselectivity	page 86
List of original manuscripts and poster sessions	page 99

Abstract

This thesis will present my research project carried out during the three-year program as a PhD student in the Structural Biology Laboratory at the University of Pavia. As better explained below, I worked on the structural elucidation of different enzymes performing reactions on disparate substrates that are of paramount relevance for biocatalytic applications. Along the thesis I will mainly focus on the experimental work performed by myself, although I will also refer to other studies performed by collaborators involved in each project to provide a comprehensive description of the results and their scientific implications. Part of this work led to the publishing of two articles that are attached at the end of this thesis.

Chapter I is introductory to the field of structural enzymology, with emphasis on the importance of studying biological catalysts in their fine molecular mechanisms. In Chapter II the main topic of my PhD will be extensively presented, which concerns the enzyme UDP-apiose/UDP-xylose synthase (AXS). As the name suggests, this plant enzyme is involved in the synthesis of two different activated sugars, used by plants in the cell wall formation. I will describe in detail the methodologies applied to investigate the structural properties of this enzyme and the results obtained, including protein purification, biochemical profiling of wild type and mutant enzymes, structure determination and analysis. All these data will be used to explain the enzyme reaction mechanism, which has been long awaited and remained rather obscure for the last 40 years. In Chapters III and IV I will report on polycyclic ketones monooxygenase (PockeMO), a fungal NADP-dependent Baeyer-Villiger monooxygenase, and on the bacterial enzyme 7 β -hydroxysteroid dehydrogenase (7 β -HSDH), respectively. As these works were published in two articles (Fürst MJLJ *et al.* 2017, JACS and Savino S *et al.* 2016, Proteins), I have attached the original manuscripts, including the experimental session, and in Chapters III and IV I will focus on the biological significance of the results considering the biocatalytic properties of the enzymes under study. This kind of exposition has been chosen to avoid repetitions, while still highlighting and commenting the results obtained during these three years of formative experience.

Chapter I:

Enzymes and their structure

I.1 Properties of the enzymes

Being mainly proteins, enzymes are represented by long polypeptides, which, by virtue of how they are synthesised and due to the physical-chemical constraints dictated by their specific amino acid sequence, fold into a unique structure that is strictly related to their biological function. Even if protein folding, and the pathways leading to the acquisition of protein structure, will not be explored in detail in this thesis, the role of such events is critical in acquisition of enzymatic properties. In a similar way to how misfolding events may cause prion diseases and amyloidoses, also enzymes can be affected by these kind of “mistakes”, which lead to inactive molecules, altered activity or elimination of the misfolded enzyme [1]. Protein folding is strictly entangled to the evolution of these molecules and to the different ways Nature has undertaken to develop many different reactivities by mutagenesis, fusion of different folds, truncations, rearrangements of the overall architecture and localization at the cellular level. Enzymes are typically globular monomers or oligomers when they serve as water-soluble catalysts. Membrane enzymes evolved a peculiar structural organization intended to stay anchored to the lipid bilayer, and depending on how they are attached, the activity can be defined as intra or extracellular. Some enzymes are known to interact between each other or with other ancillary factors to combine in macromolecular assemblies [2] [3]. We will here focus on globular soluble enzymes of bacterial, fungal or plant origin which are suited to be used for biocatalytic applications. In this respect, studying biochemical and structural properties of these enzymes relatively to their biological role in cellular pathways [4] provides important elements for catalytic optimization.

Enzymes have higher reaction rates than chemical catalysts by several orders of magnitude and they can perform their physiological reactions in much milder conditions. In particular, enzymes generally work below 100°C (i.e. at temperatures compatible with cell life), with important – and industrially relevant – exceptions for those ones coming from extremophilic microbes. Moreover, atmospheric pressure and nearly neutral pH values are perfectly

suitable for enzymatic reactions. These conditions make enzymes valuable tools in industrial applications from the safety perspective. Moreover, enzymes are more selective than chemical catalysts both for substrates and products, which results in reduced amounts of starting material, by-products and waste. This high level of specificity is strictly related to the enzyme structure, which provides an active site architecture where the substrate perfectly fits and establishes favourable low-energy interactions that dynamically regulate enzyme functions. Enzymes perform chemical reactions at physiological conditions because they are able to overcome the energy barrier associated to them (Fig. I-1). By doing so they make chemical reactions suitable for the biological systems they are part of. The formation of a transition state that takes from the initial to the final state of the reaction is a necessary step. Enzymes can lower the energy required to pass this transition state in a considerable manner. As the reaction can proceed in both ways, a catalyst accelerates forward and reverse reactions equally. Other driving forces control the direction of the reaction, such as concentration of reagents and products, stability, reaction temperature, pH values that stabilize intermediates or products and availability of salts, which cooperate in the reaction.

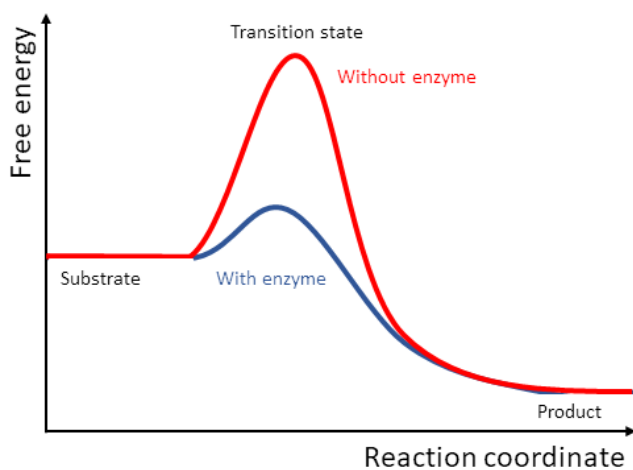


Figure I-1: formation of the transition state corresponds to an increase in free energy. In the case of catalysed reaction, the delta energy is much lower, while the energy levels of substrate and product remain the same in both cases.

Cofactors are often involved in enzymatic reactions. Despite the high potentiality of functional groups present in the active site of enzymes, some reactions can only be efficiently performed in presence of other molecules that assist proteins in the catalytic process by means of their redox potential. Cofactor is a very broad term that includes both metal ions and coenzymes (Table I-1). Coenzymes are divided in co-substrates, which are transiently bound, and prosthetic groups, permanently associated to the protein. The active complex of cofactor and enzyme is defined as holoenzyme, while an apoenzyme is referred to as the protein scaffold dissociated from the cofactor, which therefore exists in an inactive form. While metal ions participate to reactions in a similar way as they would do for a chemical process – and so they just need to be maintained at a certain level of availability in the cell – coenzymes must be regenerated, since they participate like substrates to the reaction. Coenzyme regeneration can occur during the catalytic process itself when they are still bound to the enzyme and by performing a reverse reaction. Otherwise the coenzyme can be released and regenerated in a reverse reaction performed by another enzyme. This latter strategy is often used in biocatalysis in large-scale production.

Cofactors		
Metal ions	Co-substrates	Prosthetic groups
Zn ²⁺	NAD(P)	Heme
Fe ³⁺ /Fe ²⁺	FAD	Bound FAD
Cu ²⁺ /Cu ⁺	FMN*	Pyridoxal phosphate**

Table I-1: some examples of cofactors belonging to the different categories already mentioned. The cofactors in bold are involved in the reactions of the enzymes I have been working with. *FMN is flavin mononucleotide, quite similar to FAD. **Pyridoxal phosphate is commonly known as vitamin B6.

I.2 Physiological roles of enzymes

The most relevant feature of enzymes is to make possible life as we know it. Enzymes are essentially molecular machines, performing a variety of tasks. In many cases their role is extremely specific, as they perform a single reaction on a small class of compounds or even on a unique molecule, while in others they are meant to act on a panel of chemicals as broad as possible

[5]. Not only the substrate selectivity can vary, but also the reactivities catalysed by a single enzyme can be multiple, as we will discuss in depth in Chapter II of this thesis.

Enzymes are involved in cellular activities at disparate levels, ranging from the replication of the information carriers, to the synthesis of proteins. They are also the main actors of metabolism, by building, combining and disassembling nutrients and structural components of the cell. Many signalling molecules are derived from their catalytic action over available metabolites, and enzymes can be even released outside the cell to exert some functions, for example food digestion.

The regulation of enzymes is extremely fine from the mechanistic point of view and includes allosteric mechanisms, covalent modifications and pH/ionic strength dependency. Moreover, regulation of enzyme synthesis is controlled at the level of gene expression, for example by inducible systems, often inserted into large operons. Inhibition by substrate and product concentrations provides an additional layer of control. In the end, protein stability and half-life are also important in determining the overall efficiency of enzymes *in vivo*, together with cofactors availability, programmed senescence and degradation of the proteins.

I.3 Biocatalysis

The potentiality of enzymes to catalyse reactions also outside a cellular context is the concept at the base of biocatalysis. Despite the unaware biotechnological usage of enzymes dates back to the human origins, only in the second half of the XX century enzymes could be produced in large-scale as recombinant proteins to be intentionally used as biocatalytic tools. The very first application of biotransformation in prehistorical times was the fermentation, based on empirical practice. This process, spontaneously performed by microorganisms, allowed for millennia the production of cheese, beer and wine across Europe, Asia and Africa. By the end of the XVIII century the first attempts to produce amylases on large scale followed the discovery that these enzymes were produced by fungi during aerobic growth. Of course, the existence of enzymes was not yet known at that time. The rational utilization of microorganisms, and hence of enzymes, is strictly

bound to the beginning of microbiology and to the work of Louis Pasteur and Robert Koch. Once the fermentation was eventually described as a biological process caused by microorganisms, numerous new ways to exploit the process were developed. Lactic acid, ethanol, butanol, acetone and glycerine started to be produced in mass. In 1894 the first enzyme to be commercialized was a fungal amylase, patented by Jokichi Takamine.

Nowadays the choice to use an isolated enzyme or entire cells is based on the productive process to be performed. A major limitation in the usage of cells consists in the fact that, while growing, they undergo different metabolic phases, and this is reflected in fluctuations of the yield of target molecules and in their more difficult purification. Instead, modern biocatalysis based on employing a purified enzyme is limited by the need for immobilization of the protein molecules. This step is required, in bioreactors, to increase efficiency, and separate the precious catalyst from the product-containing media that must be collected for extraction. Another concern, consists in enzyme stability in the different conditions required for industrial applications. Heat stability represents a common problem, which is often tackled by enzyme optimization through structural studies and site-directed mutagenesis approaches. For example, removal or rigidifying of flexible loops has been done to address this problem [6], and different bioinformatics methods can be used to tackle stabilization through enzymes sequences knowledge [7]. Enzymes are also supposed to be solvent resistant when applied in productive steps. This requirement is essential to solubilize them in the proper medium, usually an organic solvent, from which the final product can be efficiently extracted [8]. Rational enzyme engineering has been used also to turn substrate specificity of enzymes toward different molecules or even to introduce different activities [9], [10].

Despite these issues, which must be fixed in the research and development phase of a biocatalytic process, from the economic and ecological perspective the usage of renewable catalysts like enzymes is very convenient. Together with the possibility to finely tune and control the productive phase, the expenses related to energy supply are reduced compared to the chemical paths. Moreover, another positive aspect is represented by the production of waste and by-products, which can be reduced or avoided by coupling the

reaction to another enzymatic function, which in turn may even be used to produce other high value compounds.

I.4 Structural enzymology

To investigate the fine aspects of enzymes reactivity, it is necessary to use different tools. Complementary to biochemical and enzymological characterization, structural biology is of paramount relevance to study biocatalysts. Amongst the number of techniques now available for structural analysis of proteins, biocrystallography is still the most powerful and reliable one since high-resolution atomic models of macromolecules can be obtained.

Structural biology applied to enzymes allows to describe in detail the architecture of their active sites and uncover the catalytic mechanisms and the ligand binding modes. Comparative studies are usually informative with respect to the investigation of the active sites and of allosteric effects, and most often they make use of homologs of the enzyme present in other species. In general, structure is more conserved than sequence, therefore enzyme homologs that share low sequence identity may have structural features in common which can be relevant for biotechnological purposes. By inspecting enzyme structures some recurrent elements can be found to help categorize a protein and understand ligand binding, solubility and interactions with other proteins. The most typical secondary elements present in a protein are the α -helix and the β -strand. These structural elements help the organization of the protein by establishing inter- and intra-motif interactions. α -helices have an extremely regular pitch thanks to the N-O hydrogen bonds, which are present along the entire helix axis; thanks to such regularity this element can extend for long traits of protein, conferring rigidity and behaving as a rigid body. β -strands also can form semi-rigid structures by arranging in sheets of parallel or anti-parallel strands, to which a certain torsion can be applied. These two elements can be combined in many ways, forming super-secondary structures such as $\beta\alpha\beta$ motives, β hairpins, $\alpha\alpha$ motives, Greek keys and β barrels. The boundary between supersecondary and tertiary structure is quite labile, if one considers that such elements are also involved in the formation of folds with recurrent function, such as the Rossmann fold. Proteins bearing the typical Rossmann fold are supposed to host a dinucleotide cofactor in their core;

whenever such cofactor is not retained by the protein, using bioinformatic sequence-based predictions, the molecule binding can be still hypothesized, based on the position of the Rossmann fold and on the aminoacidic environment of the putative binding pocket. A degree of flexibility is provided by the presence of loops with random arrangement, Ω loops and β bends.

A higher level of organization that occurs in some enzymes is represented by the quaternary structure, i.e. the association of identical (or even distinct) subunits is required to reach the final functional form. This is a convenient way for enzymes to exert their functions *in vivo*, as each subunit maintains its active site, and, from an evolutionary point of view, it is easier than extending the overall length of the polypeptide chain [11]. This level of organization is entirely based on the exposed portions of each protomer, which interact with others by means of hydrogen bonds, hydrophobic and van der Waals interactions, and even disulphide bridges. Such forces can be more or less prominent, meaning that the oligomeric state can be subject to alterations depending, for example on pH or ionic strength of cell compartments *in vivo* and of buffer composition *in vitro*.

The aforementioned complementarity features between active site and ligands are the base of the lock-and-key model. Thanks to structural studies it is possible to rationalize these interactions and to see conformational changes that can occur when they are established. Usually the induced fit of ligands in an active site is a phenomenon involving few residues, placed on flexible loops or on rigid elements, which can be roto-translated to enfold the molecule and allow the reaction to occur, by moving close the active amino acids.

I.5 Protein crystallization

Determination of enzyme structures by X-ray diffraction methods (Fig. I-2) have of course advantages and limitations with respect to other techniques. The necessity to grow crystals of proteins (Fig. I-3 left) is the main bottleneck of this methodology. Crystals are regular solids characterized by a periodic ordered structure; this means that the molecules composing crystals are

repeated in the three-dimensional space, theoretically maintaining the same shape and with the same reciprocal distances between atoms (Fig. I-3 right).

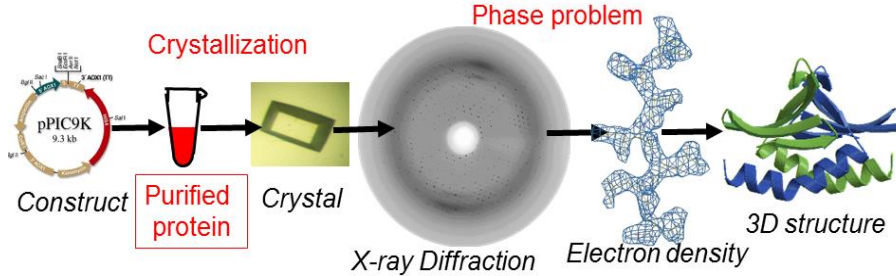


Figure I-2: pipeline of a protein crystallography project: the target macromolecule is generally produced as recombinant protein, followed by crystallization screening and optimization and X-ray data collection; intensities of each reflection can be measured and combined with phases to calculate the electron density map by the Fourier transform, which allows to trace the atomic model of the protein structure. While the protein expression and purification generally do not represent limiting steps (at least for bacterial and fungal proteins, with mayor exceptions for membrane proteins, glycosylated proteins and complexes), crystallization is a well-known bottleneck. Moreover, even when obtained, crystals might feature issues like limited reproducibility, bad diffraction and twinning that affect the data quality and usability. In this respect, availability of models for molecular replacement can be critical for the structure determination. Insufficient resolution, lack of electron density for flexible portions, model quality and out-of-register assignment of aminoacids are always possible and deserve special attention. In my projects I sometimes started from purified protein samples shipped by collaborators and most often established the expression and purification processes by myself.

Salt and organic small-molecule crystals are kept together by strong interactions and they are resistant to mechanical forces. In protein crystals interactions are much weaker, including hydrogen bonds and van der Waals forces, which make them much more fragile. Such fragility is utterly increased by the very high amount of water solvent present in biological crystals. The presence of solvent, despite increasing the fragility and susceptibility of protein crystals to mechanical shocks when manipulated, is extremely important for the maintenance of the functional biological conformation of the protein of interest. The presence of solvent is also important to perform experimental procedures on the crystalline sample, such as heavy metal and ligand soaking. The need for a periodical structure (Fig.

I-3 right) is intrinsically linked to the X-ray diffraction method, as periodicity is required to amplify the signal of the diffraction pattern, thanks to constructive interference. A single protein molecule, other than being hard to manipulate and irradiate, could not give a diffraction pattern.

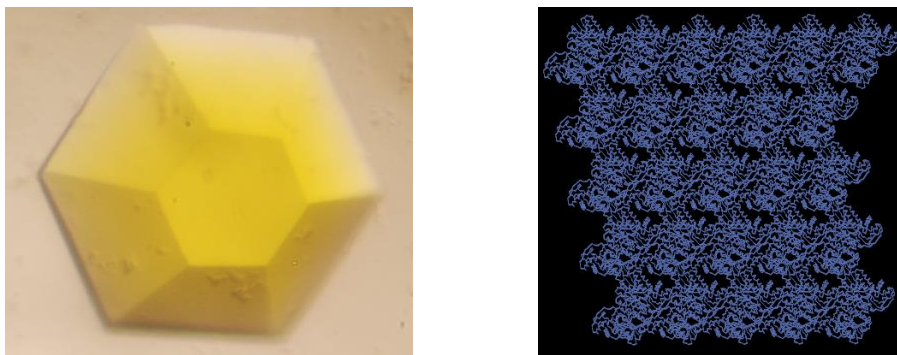


Figure I-3: on the left a nicely shaped protein crystal of PockeMO, one of the enzymes on which I have worked in my PhD project. The yellow colour is due to the Flavin (in the oxidized state) associated to the protein. On the right an example of molecules packed in a crystal (of another protein). This is P1 space group that does not contain crystallographic symmetry except for translation. Molecules are tightly packed, which results into a very low solvent content of 41%.

In crystallization the key point is to modulate protein solubility, even if many other factors are critical for crystal growth. This is achieved by adding to the protein solution a precipitant agent (frequently a salt or an organic polymer) that modifies the solvation state of the protein molecules and favours protein-protein interactions. The transition from protein molecules in solution to crystals involves different steps which can be represented in the phase diagram (Fig. I-4). These range from ordered aggregation of few molecules (nucleation) to addition of more of them to nuclei to form the final crystal. Nucleation is a kinetically unfavoured process, because of the high flexibility of biological macromolecules. On the other hand, crystallization is also thermodynamically favoured [12] as it satisfies the formation of weak interactions on the extended protein surfaces by forming an ordered pattern. Super-saturation is the physical state of solubility where crystallization can occur. The super-saturation region is also divided into two sub-regions: in the

nucleation (or labile) region, the crystal is formed, while crystal enlargement occurs in the metastable zone (Fig. I-4).

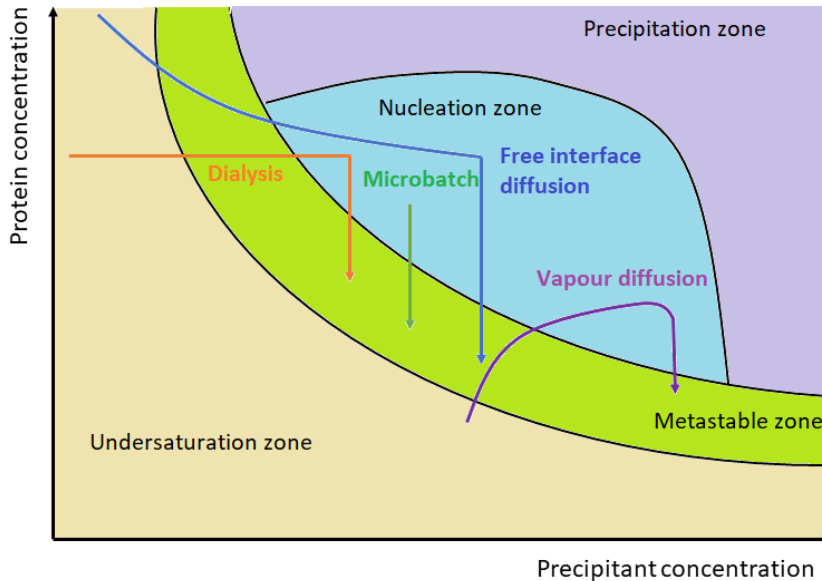


Figure I-4: phase diagram of crystallization experiments. All methods must transit from the nucleation to the metastable zone in order to have proper crystals and avoid precipitate or formation of too small objects.

Protein crystals are however extremely difficult to obtain, and their growth is still considered the major bottleneck in the structural biology pipeline. This statement is not surprising if we consider the number of variables that can contribute to a successful crystallization. The main responsible factors are the chemical composition of precipitant, its concentration, the presence of additive salts and their concentration, the pH value of the solution, the temperature of incubation and the protein integrity [13]. Several other factors can be added to this list, like pressure, gravity and electromagnetic forces; but their contribution is rarely considered, as it would overcomplicate experimental setup. All these factors can give high variability in obtaining crystals. Even the propensity to degradation of the protein sample and its tendency to crystallize only at precise concentrations makes the experiments hard to be reproduced. In short, the number of variables to be considered in a

protein crystallization experiment is simply too high. Is not possible to design systematic crystallization methodologies without knowing anything *a priori* about the crystallization behaviour of the designated protein target. Therefore, the best strategy structural biologists have come to develop is massive screening of different semi-random crystallization conditions.

Protein crystallization is therefore a trial and error process in which many conditions are tested by mixing them with the purified protein sample at a precise concentration. The presence of contaminants in the protein sample must be avoided, as it may impair crystallization by interacting with the target molecule, or simply the contaminant itself may crystallize in place of the target. Knowledge of biochemical features of the protein and all the available literature on the topic is essential to provide the sample with ligands, cofactors and stabilizing agents. To maximize success rate, reagents that have shown to give crystals should be used, and then the mixture should be optimized through a trial and error process to give better diffracting crystals. For this reason, commercial kits of sparse matrix screens are available to test many experimental conditions. Other kits are instead based on a systematic approach of multidimensional grids, which are meant to test precipitants, salts and buffers concentrations.

Canonical precipitants are for example polyethylene glycol (PEG) of different lengths, and ammonium sulphate ($(\text{NH}_4)_2\text{SO}_4$), but there are also more “exotic” ones, such as 2-methyl-2,4-pentanediol (MPD), various alcohols and poly- γ -glutamic acid (PGA). The role of the precipitant is to sequester water molecules from the solution by competing with the protein for them. Salts are additives that can help crystallization thanks to their charges. Of course, different salts can exert different actions on the sample, and usually just a few salt species are prone to help the reaction. They can do this by masking charges of side chains or making salt bridges inter or intra-molecularly.

The control of pH with a buffer solution is essential not only to eventually obtain a reproducible result, but also to keep the biomolecule at a stable pH value. Whenever the pH value is not controlled in a drop that gave a positive hit, a quest begins to find the proper pH value, which is usually in a quite narrow range. Crystallization experiments can have different outcomes. In the worst case, the protein is damaged by reagents used in the crystallization

condition or because of pH, which leads to protein denaturation and produces an amorphous precipitate. Most drops will not give any kind of result, meaning that the protein is not affected by anything present in the condition and just remains in solution.

In practice, a crystallization experiment requires a pure protein sample, crystallization plates and screening kits, or chemicals to be mixed at use. The subsequent optimization steps will require preparation of customised crystallization mixtures, with an appropriate concentration of each compound, and with the right pH in order to obtain the best possible crystals. Automation of crystallization by robots [14] has made possible to increase the number of individual conditions to be screened per unit of time with respect to manual work, and reduces significantly sample consumption. This approach makes possible to find crystallization hits much more efficiently, but very often the crystals obtained are too small for diffraction experiments, thereby requiring additional manual optimization steps before proceeding further with the X-ray measurements. Several techniques are available for crystallization. Most of these techniques can be performed nowadays also using automated robots.

Vapour diffusion is the most widespread method to get crystals: it consists in preparation of a reservoir with the right crystallization mix and a small drop of protein to which an equal volume of reservoir solution is added. The droplet and the reservoir are left to equilibrate in a sealed chamber, allowing vapour to diffuse. As the droplet contains half of the total concentration of precipitant in the reservoir, the slow diffusion of water vapour from the drop to the reservoir leads to reduction of drop size and, therefore, increase of protein and precipitant concentration that may evolve to crystal formation. Vapour diffusion can be performed in sitting drop, hanging drop and sandwich drop configurations. In sitting drop, a microbridge with room to host a drop is put over the reservoir, while a transparent tape isolates the environment; in hanging drop a siliconized glass cover is used to seal the well using vacuum grease and the drop just hangs from the cover over the reservoir. The big difference between sitting and hanging methods is the volume of drop, which can be larger in sitting, making it the most used method for optimization. Manual vapour diffusion is the procedure where the operator can give the best contribution to obtain the result, in particular when

a protein has shown to need special conditions, like addition of cofactors just before the experiment or short exposure time to open air and room temperature. Even pipetting and mixing the sample can be crucial, and in some cases the deposition of an oil film above the reservoir may be necessary to slow down the process. All these procedures cannot be realized by a nanolitre-dispensing robot and have the potential to make the difference between successful crystallization and umpteenth fail. On the other hand a robot can dispense very small volumes in a very precise and reproducible way, saving precious sample.

Batch methods require the deposition of protein drops mixed with crystallization condition, and then covered by a layer of paraffin oil; in this case super-saturation is reached immediately and the risk of precipitation increases. In dialysis, or liquid diffusion method, the crystallization solution passes through a dialysis membrane, not altering protein concentration. This method is not so widespread anymore but has a great advantage over others. It allows changing the crystallization solution in which the dialysis membrane is put, making possible to re-use the protein sample. Seeding consists in taking a crystal grown in a condition and placing it in another one, together with an amount of protein in solution. The pre-existing crystal behaves as a nucleus on which a larger crystal grows. Even twinned crystals can be used thanks to micro-seeding [15]. In this case, the twinned crystal is broken using a vortex, then the resulting micro-fragments are dispensed in drops of screening kits together with protein in solution.

To obtain a protein structure in complex with a ligand (which may be a substrate, an inhibitor or a regulatory molecule), two methods can be used. In co-crystallization experiments [16] the ligand is added at high concentration in molar excess compared to the protein stock solution and used for setting up crystallization plates. In this way, a crystal containing the complex of protein and ligand may grow; *de facto* the complex formation itself can help to get an ordered structure, as it is more stable than the apoenzyme. Another strategy to obtain structural data of protein complexes is to perform soaking experiments in which, prior to cryo-freezing, the crystal is transferred to a drop containing the ligand, with incubation times that may vary depending on the specific case. This technique takes advantage of the high solvent content of protein crystals to let molecules diffuse in the solvent channels. Controlled

exposure will identify the minimal concentration of molecule required for efficient binding, minimizing the possibility of distortions in the crystal lattice during soaking operation. A variant of soaking is the competitive replacement of ligands [16].

Once crystals have grown, they must be harvested from the droplets using a nylon or plastic loop of the proper diameter, and then stored in liquid nitrogen. This procedure is performed to protect the crystal from degradation that may occur because of temperature fluctuations during transportation. The theory behind this step is that the crystal is quickly taken to -196°C , avoiding formation of crystallized ice, which would destroy the crystal, as the solvent expands to reach the solid state. Since amorphous ice is formed during flash freezing, the expansion does not occur. In addition, by avoiding crystalline ice formation the final diffraction pattern is enhanced. Before putting it in liquid nitrogen, the crystal is protected from possible cryogenic damages through a brief immersion in a solution, identical to the crystallization mix, but containing 20% v/v cryo-protecting agents, such as low molecular weight PEG or glycerol. Formation of ice around the crystal must also be avoided during following transfer step, for example during crystal mounting on the goniometer of the beamline.

I.6 X-ray crystallography

The basic element of a crystal that represents its translational periodicity is the unit cell, defined by the lattice constants: three axes a , b , c and the angles α , β , γ . Crystal periodicity, which results from the geometric arrangement of the molecules, can be represented by translation of the unit cell that generates the so-called Bravais lattice. The unit cell can contain one or more molecules that are related by crystallographic symmetry operators. In protein crystals, rotation and screw axis operators exist only, because others such as symmetry planes are incompatible with the chemistry of biological macromolecules that contain chiral centres. In some cases, also non-crystallographic symmetry may be present when more than one molecule occupies the asymmetric unit that generates the unit cell by application of crystallographic operators. It has been demonstrated that there is a finite number of combinations of a molecule in a crystal. By combining the 14 possible Bravais lattice and the 32 point

groups (resulting from the combination of symmetry operators), 230 theoretical space groups are defined, of which only 65 are compatible with protein crystals because of the chiral nature of biological molecules.

Many series of infinite sets of parallel planes, defined by the Miller indices $h k l$ corresponding to their interceptions with the unit cell axes, have been used to mathematically describe crystals and their diffraction properties [16]. X-ray diffraction is based on scattering phenomena that occur to waves whenever they are deflected. In our case, the object causing deflection in a direction different from the incident one, and with loss of part of the energy, is an atom. The portion of the atom that is relevant in X-ray diffraction is not the dense nucleus, but the electron cloud. This means that, depending on the number of electrons an element has, it will produce X-ray scattering in a more or less intense way. In protein crystals, the combination of large unit cells, high solvent content and prevalence of poorly electron-dense elements means low scattered intensities. Even if this was not a trivial issue in the past, the availability of more powerful beamlines and more sensible detectors have drastically improved the state of the art in data collection and processing.

As described by elastic Rayleigh scattering, waves are re-radiated by interference with the electron cloud, maintaining same wavelength and frequency. The crystal diffraction phenomenon is described by Bragg's law (Fig. I-5), which takes place in the conditions for constructive interference of two individual waves colliding two objects on different parallel planes of the same series (corresponding to a triad of Miller indices $h k l$) at distance d having same θ angle of reflection.

As each series of parallel planes deviated the incident beam like a light with respect to a mirror, the diffracted X-rays are commonly named reflections. Each reflection corresponds to a spot on the detector and is assigned to a $h k l$ triad.

At the synchrotron facility, a discrete, single wavelength (typically around 1 Å) is selected through a monochromator and a highly focused, collimated beam, with sizes between 5 and 100 microns, is used to irradiate the crystal. The crystal, laid on the loop used for fishing and transportation, is mounted on the goniometer head and kept under a constant stream of nitrogen during the diffraction experiment in order to minimize the damage caused by the

irradiation. Thanks to the goniometer on which the loop is mounted, the crystal can be freely rotated around an axis (φ) perpendicular to the incident X-ray beam. As multiple images must be collected, the crystal is rotated around this axis in a continuous way, while data are collected at discrete intervals of angles. The detectors used in current experiments are CCD or pixel devices, capable of acquiring images with high turnover. The level of technology behind state-of-the-art detectors is so advanced that nowadays the bottleneck of data acquisition is the computational and storage step and not the acquisition itself. The most external spots (outer shell) collected by the detector will correspond to the high-resolution data.

Bragg's law for diffraction: $n\lambda = 2d \sin\theta$

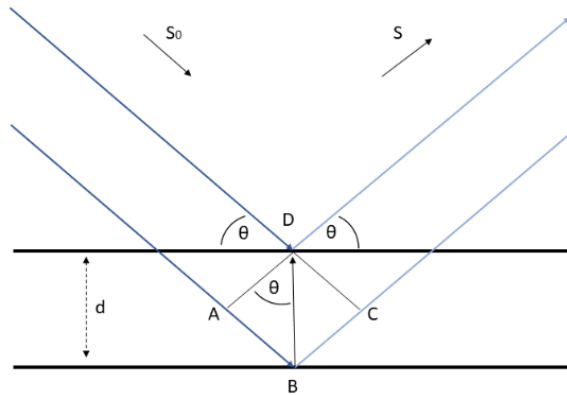


Figure I-5: diffraction phenomena according to Bragg's law. S_0 are incident waves while S are the scattered ones. The diffraction or scattering angle θ is defined by the plane on which D lays and the vector of the wave. The distance d between the two planes and the prosecution of S (AD or CD) also defines the same θ angle.

Thanks to the first preliminary orthogonal images, which are typically two or four, the space group and the unit cell parameters can be determined (with possible errors in correct enantiomorphs assignment). This information can be used by an operator to decide the best strategy for data acquisition, based on consideration about the symmetry of the crystal and the multiplicity of the information. Two parameters must be settled in the strategy, being the number of images and the oscillation angle. Their combination is required to have the highest possible completeness before the crystal is damaged by the radiations

accumulated in the observation process. Depending on the crystal, the radiation damage can be important in lowering data quality. Current state knowledge of crystal parameters is advanced enough that software can determine the best theoretical strategy to be performed on a crystal.

The collected data must be stored and then processed with computer programs like XDS [17] and iMosflm [18] to perform indexing and integration of the single diffraction images followed by merging and scaling procedures performed by CCP4 programs (Aimless [19]) or Phenix [20], which produce a unique data file containing the list of *hkl* reflections and the corresponding intensity values. The *mtz* file that is now obtained contains the data to be used for structure solution, which requires an interpretable electron density map to fit the polypeptide chain and obtain the atomic model of the target (Fig. I-2). This map is calculated by the Fourier transform of the structure factors for each reflection, containing both amplitude and phase components. Indeed, the Fourier transform is a mathematical tool that allows passing from the reciprocal space of the diffraction images back to the real space of the atoms that originated the scattered waves. However, at this stage we must deal with the most important limitation in crystallography, the phase problem. This issue is intrinsic to the X-ray diffraction method, as the phase information is lost and cannot be retrieved from the X-ray diffraction pattern reflections directly. Despite this, the phase problem can be solved in different ways. One of these is the molecular replacement method that I personally used for my PhD research projects. This method necessarily can be applied when the structure of a protein homologous to the target one is already available (i.e. present in the PDB database or already solved in the laboratory and not published yet). This is defined as the search model and is used to extrapolate a set of phases (though not complete and not accurate), which can be nevertheless combined to the experimental intensities measured on the target crystal. In order to be used, the search model must have a degree of homology with the target protein; this also means that typically the search model is a member of the same protein family of the target, or that it shares at least some folding features. By applying the Patterson function, which can be regarded as a reverse Fourier transform that not considers phases, a map is generated. This Patterson map is not in a real space and must not be confused with the reciprocal space of the intensities or with the real space of atoms. Once the Patterson map is generated, rotation and translation functions are applied to

correlate the target to the search model. Applying roto-translation functions to minimize differences between Patterson functions of search and target, allows usage of the phases from the search model. The space of all possible rigid body transformations of a molecule has 6 dimensions, with 3 angles for rotation and 3D translation vectors. Molecular replacement is basically the method to find these angles and vectors. After this, the phases of the structure taken as model are combined with the amplitudes of the new protein and new structure factors are calculated. This step is routinely performed by using the CCP4:Phaser program [19], and the result is an electron density coupled to a first raw model fitted inside it. Depending on the identity level with the search model, the target will be fitted in the electron density with a certain level of confidence and some portions may not be fitted at all, to avoid introduction of model bias. By using the Coot program [21], it is the operator that must take care of proper fitting of the remaining parts of the protein, as well as to assess the quality of the automatically fitted parts. It is important in this stage to evaluate the chemical feasibility of interactions such as H bonds in α helices and β strands, and to make sure that the rotamers of amino acids and the Ramachandran plot are consistent with statistical distributions. At the same time, the data themselves must not be ignored in favour of pure statistics. Non-peptidic molecules such as cofactors and other ligands are introduced following the indications of electron density. This more accurate model of the structure undergoes cyclic refinement and validation procedures to assess data quality to eventually obtain the final coordinate file used for structural analysis.

The structural determination is a rigorous procedure with various statistical validation checks at each step. The most important parameters to be considered for data are the R_{work} , the R_{free} , and their ratio. These values indicate the measure of agreement between observed and calculated data. While R_{work} (see the equation below) is representing the standard for assessing such agreement, R_{free} consists in a cross validation, calculated on a subset of reflections not used in the refinement step. Since the two values are intrinsically entangled, being part of the same dataset, a too large discrepancy in their values is not admitted, and so their ratio should be close to 1. Typically, as these values are expressed in percentage, a gap of 5% or lower is a good indicator. At the same time, the R_{work} and R_{free} values must be low (around 20-25%).

$$R = \frac{\sum hkl | F_{obs}(hkl) - F_{calc}(hkl) |}{\sum F_{obs}(hkl)}$$

Another common parameter is the R_{merge} , calculated on independent measurements of the intensity of reflections. Since this value is affected by multiplicity of the collected datasets and must be adjusted accordingly, and also suffers from other case-dependent issues – the description of which escapes the aim of this thesis – another measure of quality was introduced, the CC1/2 [22]. The CC1/2 is calculated by splitting the dataset in two halves, each one having random measurements of each unique reflection. For each half set the correlation coefficient is calculated and the two values, when compared, must have a ratio close to 1 at low resolution and close to 0.1 at high resolution. All these values are used to evaluate that the processed data are consistent when the procedure is done on different batches.

I.7 References

1. Underhaug J, Aubi O, Martinez A. Phenylalanine hydroxylase misfolding and pharmacological chaperones. *Curr Top Med Chem.* 2012;12(22):2534-45.
2. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. The structural basis of ribosome activity in peptide bond synthesis. *Science.* 2000 Aug 11;289(5481):920-30.
3. Khatter H, Vorländer MK, Müller CW. RNA polymerase I and III: similar yet unique. *Curr Opin Struct Biol.* 2017 Jul 22; 47:88-94.
4. Zhao S, Kumar R, Sakai A, Vetting MW, McKay Wood B, Brown S, Bonanno JB, Hillerich BS, Seidel DR, Babbitt PC, Almo SC, Sweedler JV, Gerlt JA, Cronan JE, Jacobson MP. Discovery of new enzymes and metabolic pathways by using structure and genome context. *Nature.* 2013 Oct 31;502(7473):698-702.
5. Fiorentini F, Geier M, Binda C, Winkler M, Faber K, Hall M, Mattevi A. Biocatalytic Characterization of Human FMO5: Unearthing Baeyer-Villiger Reactions in Humans. *ACS Chem Biol.* 2016 Apr 15;11(4):1039-48.

6. Yu H, Huang H. Engineering proteins for thermostability through rigidifying flexible sites. *Biotechnol Adv.* 2014 Mar-Apr;32(2):308-15.
7. Wijma HJ, Floor RJ, Janssen DB. Structure- and sequence-analysis inspired engineering of proteins for enhanced thermostability. *Curr Opin Struct Biol.* 2013 Aug;23(4):588-94.
8. Monsef Shokri M, Ahmadian S, Akbari N, Khajeh K. Hydrophobic substitution of surface residues affects lipase stability in organic solvents. *Mol Biotechnol.* 2014 Apr;56(4):360-8.
9. Brondani PB, Dudek HM, Martinoli C, Mattevi A, Fraaije MW. Finding the switch: turning a baeyer-villiger monooxygenase into a NADPH oxidase. *J Am Chem Soc.* 2014 Dec 10;136(49):16966-9.
10. Dudek HM, Fink MJ, Shivange AV, Dennig A, Mihovilovic MD, Schwaneberg U, Fraaije MW. Extending the substrate scope of a Baeyer-Villiger monooxygenase by multiple-site mutagenesis. *Appl Microbiol Biotechnol.* 2014 May;98(9):4009-20.
11. Voet D, Voet JG, *Biochemistry*, 4th Edition, John Wiley & Sons
12. Zhou D, Zhang GG, Law D, Grant DJ, Schmitt EA. Thermodynamics, molecular mobility and crystallization kinetics of amorphous griseofulvin. *Mol Pharm.* 2008 Nov-Dec;5(6):927-36.
13. Sheehan D. *Physical Biochemistry: Principles and Applications*. John Wiley & Sons
14. Bard J, Ercolani K, Svenson K, Olland A, Somers W. Automated systems for protein crystallization. *Methods.* 2004 Nov;34(3):329-47.
15. Oswald C, Smits SHJ, Bremer E, and Schmitt L. Microseeding - a powerful tool for crystallizing proteins complexed with hydrolyzable substrates. *Int J Mol Sci.* 2008 Jun;9(7):1131-41.
16. Rupp B. *Biomolecular Crystallography Principles, Practice, and Application to Structural Biology*. Garland science
17. Kabsch W. XDS. *Acta Crystallogr D Biol Crystallogr.* 2010 Feb;66(Pt 2):125-32.

18. Powell HR, Johnson O, Leslie AG. Autoindexing diffraction images with iMosflm. *Acta Crystallogr D Biol Crystallogr*. 2013 Jul;69(Pt 7):1195-203.
19. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr*. 2011 Apr;67(Pt 4):235-42.
20. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. 2010 Feb;66(Pt 2):213-21.
21. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. 2004 Dec;60(Pt 12 Pt 1):2126-32.
22. Karplus PA, Diederichs K. Linking crystallographic model and data quality. *Science*. 2012 May 25;336(6084):1030-3.

Chapter II:**The crystal structure of UDP-apiose/UDP-xylose synthase (AXS) enlightens the mechanism for alternative product formation**

During my PhD training, I focused most of my experimental work on a project concerning the cytosolic plant enzyme UDP-apiose/UDP-xylose synthase (AXS). This protein started to be studied in the '70s and the interest for its peculiar reactivity was recently revived by our Austrian collaborators in Graz (Austria), at the Austrian Centre for Industrial Biotechnology (ACIB) and Technical University Graz (TU Graz). The aim of my work on AXS has been to purify, crystallize and determine the three-dimensional structure of this enzyme, in order to disclose the peculiar features that enable it to perform a rather unique reaction in the context of plant biology. Along with the structural studies, the biochemical characterization of AXS was carried out to disclose the fine molecular mechanisms occurring in the enzyme active site. Altogether, these results gave insights into the formation of a major component in plant cell wall, with important consequences in basic research, but also on activated sugars production and applied plant biology.

II.1 Introduction

Despite the interest and effort put by many in characterizing AXS in the last 40 years, the nature of the ring contraction leading to the formation of the product UDP-Api is still a huge gap in the description of the reaction mechanism. We focused on rationalizing this obscure point by means of structural biology and biochemical characterization, upon site-directed mutagenesis of active site mutants.

II.1.1 The AXS enzyme

Since its first discovery in the 1970s [1], AXS was considered an interesting catalyst, given its capability to synthesise two nucleotide-activated sugars,

namely UDP-Xylose and UDP-Apiose (Fig. II-1 C and D). The enzyme was indeed named after such a dual reactivity, as it was first characterized in *Lemna minor* (common duckweed). In this respect, the enzyme was immediately compared to the homologous protein UXS [2], present in plants but also in animals and men. This enzyme retains the capability to synthesise UDP-Xyl, but is not active in UDP-Api production.

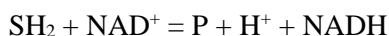
The interest for AXS never faded in the plant biochemistry community, and in 2003 an article was published on the characterization of AXS enzymes coded by *AXS1* and *AXS2* genes in *Arabidopsis thaliana* [3]. The two encoded soluble proteins share 98% sequence identity and so the studies were focused on the study of one variant only, being *AtAXS1*. The 397 aminoacids long monomer has a molecular weight of 44.7 kDa and is active in a dimeric form of 90 kDa. From bioinformatic predictions, AXS is known to be part of the NAD(P)-dependent epimerase/dehydrogenase family, featuring a Rossmann fold of β -sheets surrounded by α -helices, in the middle of which the cofactor is accommodated. In this enzymatic family, the cofactor, in proximity of the consensus sequence GxGxxG [4], can be either retained during the entire reaction cycle, as in AXS, or released upon reduction to NAD(P)H, as in the case of ArnA [5].

II.1.2 AXS ligands and physiological role

The enzyme utilizes NAD as cofactor (Fig. II-1 A), first as oxidizing agent for a reaction intermediate prone to be spontaneously decarboxylated, and then to reduce the same intermediate to either the pyranosidic (UDP-Xyl) or the furanosidic (UDP-Api) products. NAD is a well-known enzymatic cofactor utilized by many proteins to catalyse a vast array of reactions. It is composed of an adenosine moiety bound to a nicotinamide adenine via a pyrophosphate, where both phosphate groups are linked to the ribose rings at 4' position. It shares structure with NADP, with respect to which it lacks an extra phosphate group bound to the ribose of adenosine in 2' position. Such an extra phosphate is used by enzymes as a recognition element, to pick the specific cofactor, and often important residue substitutions are present among homologs utilizing one or the other molecule [6]. Shifting of cofactor specificity is an important target in biotechnological modification of proteins, as NADP is a more expensive molecule than NAD [7]. NAD is active thanks

Chapter II

to the nicotinic ring, where the electronic delocalization offered by the aromatic ring and the presence of the carboxy-amide in 5' position, promote the role of carbon in 4' position as a hydride ion acceptor. The enzymatic reactions in which NAD exerts its role occur with removal of a hydride ion and of a proton from the substrate; while the ion is bound to C4 of the nicotinic ring, which goes from trigonal to tetrahedral geometry, the proton is released in solution, as explicated in the following reaction:



where S is the substrate, P is the product and the NAD passes from the oxidized (NAD^+) to the reduced form (NADH).

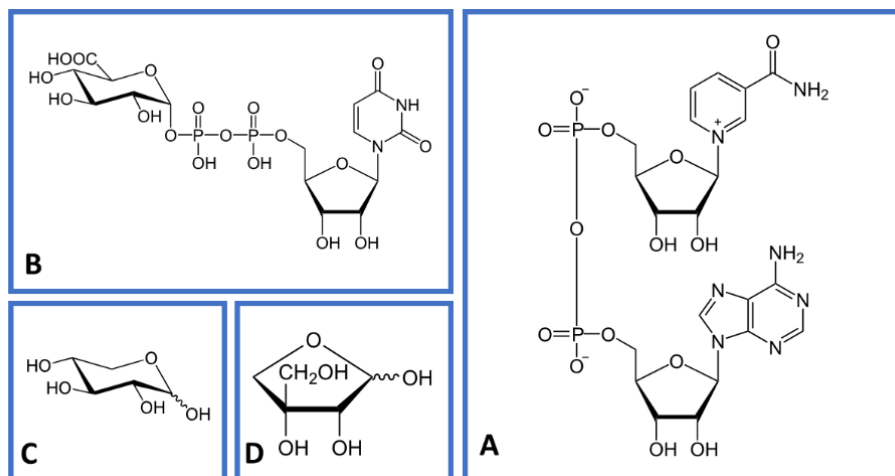


Figure II-1: ligands of AXS. NAD (A), UDP-GlcUA (B), xylose (C) and apiose (D).

In many cases, like in AXS, the reduced NADH is used in the second half of the reaction, and oxidized back to NAD^+ . Otherwise it is released, and, in cells, it can be used by another enzyme to perform the reduction of a substrate, regenerating again the oxidized NAD^+ . The substrate UDP-GlcUA (Fig. II-1 B) is a nucleotide-activated sugar, composed of uridine diphosphate and a pyranoside, the glucuronic acid, which is a six-membered ring (five carbons plus one oxygen) with a carboxylic group bound to the carbon in 5' position.

The two products are the UDP-activated sugars xylose and apiose. Xylose (Fig. II-1 C) is also a pyranoside, but with respect to glucuronic acid it lacks the carboxylic group. Apiose (Fig. II-1 D) is a furanoside, so it has a five-membered ring composed of four carbon atoms and one oxygen atom. In position 3' a branch, bound to the carbon atom, is present, consisting of a CH₂OH, that is extending out of the ring plain. UDP-GlcUA and UDP-Xyl are widely represented activated sugars, which are used in many different metabolic pathways by all eukaryotic organisms. UDP-GlcUA is used in phase II metabolism by glucuronyltransferases to solubilize xenobiotics by conjugating them to the sugar moiety [8]. Xylose is particularly important since it is the first element attached to the polypeptide chain for the addition of sugar chains to proteoglycans [9]. This is exemplified by the presence of the aforementioned UXS [10] also in *Homo sapiens*, where an endogenous biosynthesis of xylose is required for this purpose. In plants, it constitutes the starting material for the synthesis of xylan and xyloglucan in the primary cell wall [11], while in bacteria it is a constituent of the LPS and a putative target to address some spreading plant diseases [12].

The reactivity of AXS (Fig. II-2) has been already extensively described in literature [13]. It consists in the oxidation of UDP-GlcUA in position 4' of the sugar moiety, upon which the carboxylic group bound in position 5' is hydrolysed, leaving the catalytic environment in form of bicarbonate. The 4-keto-intermediate that is formed can then undergo two different reactions, both involving the reducing activity exerted by the newly formed NADH. A first pathway, which is shared with the homolog UXS (Fig. II-3), leads to the formation of xylose, with the reduction of the ketone in position 4'. A second one, which is typical of AXS only, consists in the retro-aldol cleavage of the C-C bond between positions 2' and 3'. Upon relocation of the C3 out of the plain of the sugar ring, a new bond is formed between C2 and C4, leaving the hydroxymethyl as a branch, bound to the C4 (then renumbered as C3 in the furanoside).

As discussed later, the crystal structure of wild type AXS allowed to identify two more ligands bound in the active site, namely a UDP molecule, which is a known inhibitor of AXS [2], and a phosphate group, derived from the crystallization mother liquor. Despite their non-physiological role in the

enzyme's reactivity, they provided important clues to define the enzyme reactivity.

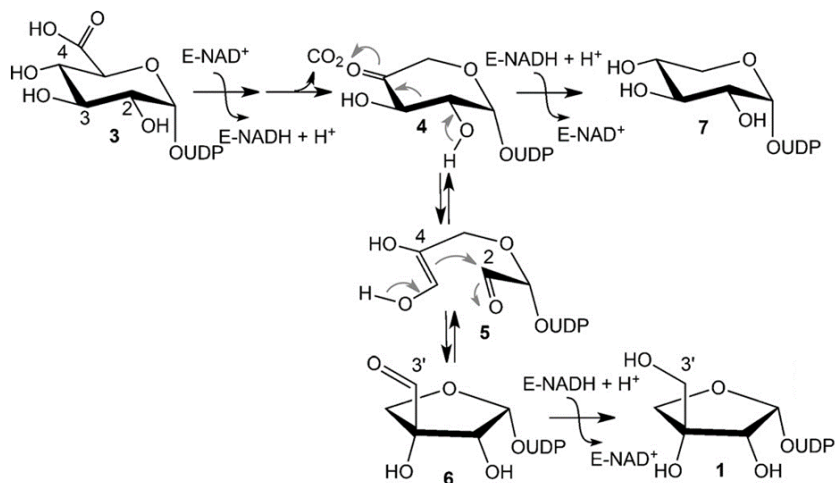


Figure II-2: general reaction scheme of AXS as already proposed in previous literature. The formation of UDP-Api (1) from UDP-GlcUA (3) proceeds through rearrangement of the 4-keto-intermediate (4) – which is also the precursor of the other product UDP-Xyl (7) – to the opened structure (5) that allows ring contraction to the 3-keto-intermediate (6). Image from [28].

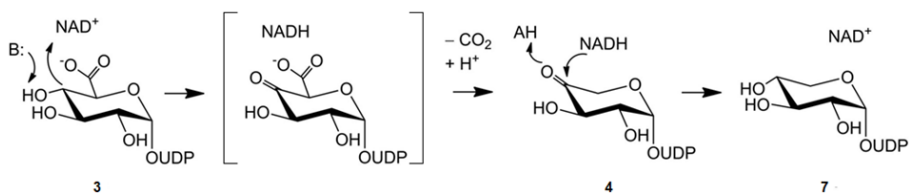


Figure II-3: reaction catalysed by UXS. UDP-Xyl only is produced. Image from [10].

II.1.3 Role of the enzyme in plant cell wall organization

The presence of a cell wall is a rather unique feature of plant cells, and it is meant to provide some degree of rigidity and to protect the plasma membrane.

According to the replication phase in which a plant cell is, the wall can be described in two different ways, as primary or secondary one [14]. Reflecting the necessities of a growing cell, the primary cell wall is elastic, while the secondary cell wall, which grows later, is rich in lignin, providing mechanical resistance [15]. One of the main components in primary cell wall is RG-II, a polysaccharide made of a galacturonic acid skeleton with four sidechains (A, B, C, D) of different, poorly represented sugars. These include apiose, which is doubly involved in RG-II organization; first it is linking the side chain B to the galacturonic acid backbone and second it is linking side chain A to the same. Most importantly, the apiose residue linking chain A to the backbone is also involved in the formation of a borate cross-link with the same sugar of another rhamnogalacturonan molecule. By doing so, rhamnogalacturonan reaches its dimeric form present *in muro* (in the wall structure), and forms the intricate network of the cell wall [16].

II.1.4 Interest in studies on AXS

AXS is being scientifically investigated to understand the peculiar reactivity of the enzyme and to identify the functionally relevant residues, which may lead to the rational engineering of the enzyme mutants specifically active in UDP-Api synthesis. Such a molecule, being a precursor of apiose, is interesting for commercial use. In particular, finding a convenient way to produce apiose and UDP-Api would be beneficial for many research groups, which cannot easily use this molecule in compound screening at the moment. Activated sugars production and modification is also relevant for solubilization of conjugated compounds, as it is possible to link them to an aglycone of interest, like an antioxidant molecule.

The pure biological interest in AXS goes to the understanding of its role in plants during evolution. As the cell wall is one of the critical elements conferring rigidity and the option of vertical growth in plants, to elucidate when and how the synthesis of an essential sugar for RG-II composition started is a major goal in explaining the choice of plants for a sedentary life-style. In this respect, also the comparison with UXS will turn useful to explain the homology level and the partial functional superposition of the two enzymes. Disclosing the reactivity mechanism of AXS is also an extremely

interesting topic from the pure biochemical perspective, since it will work as an example on how ring contraction is operated by biological catalysts.

II.2 Materials and methods

Chemicals for the preparation of buffers and crystallization solutions were bought from Sigma-Aldrich. Affinity and size exclusion chromatography columns were purchased from GE Healthcare, as well as the ÄKTA purifiers. Concentrators were from Merck. Bio-rad instruments were used to run PCR and for SDS-PAGE and acrylamide gel imaging. Master mix components for PCR were from ThermoFisher Scientific and New England Biolabs. Microscopes for crystals imaging were from Olympus, while the crystallization robot was a Douglas Instruments Oryx-8. Crystallization plates were from Hampton Research. Loops for crystals collection came from Hampton Research and Molecular Dimensions. Screening kits for crystallization conditions were purchased from Quiagen, Jena Bioscience and Molecular Dimensions. SAXS experiments were conducted at the ESRF on the BM-29 beamline. On beamlines ID-23, ID-29 and ID-30 at ESRF and on beamlines PX and PXIII at SLS the X-ray diffraction data were collected. The programs XDS, iMosflm, CCP4, Phenix and Coot were used for data processing, model building and refinement. Structural images were prepared using UCSF Chimera. Homology models were calculated by Swiss-model and enthalpy predicted by SERp. For what concerns the data collected in Graz, HPLC runs were performed on a Shimadzu system with a Kinetex C18 column and NMR measurements were done on a Variant Unity Inova 500MHz.

II.2.1 Protein expression

The AXS1 gene from *Arabidopsis thaliana* was cloned in a pET26b (+) vector to build a plasmid encoding for the recombinant N-ter, 6xHis-tagged AXS variant. 50 μ l of BL21(DE3) competent cells were transformed with 1 μ l of plasmid. After 30 minutes of incubation in ice, heat shock was performed at 42°C for 45 seconds and the cells were placed in ice for 3 minutes, before being provided with 250 μ l of LB medium for recovery. After 1 hour at 37°C

and 200 rpm shaking, cells were plated on kanamycin-containing LB-agar medium, since the plasmid provides resistance to this antibiotic for selection. The Petri dish was moved to 37°C overnight. The day after, two colonies were picked with a pipette tip and transferred into two falcon tubes with 5 mL of kanamycin-containing LB medium, thus the mini-cultures were grown at 37°C and 200 rpm shaking overnight. These mini-cultures were used to prepare glycerol stocks, which can be stored at -80°C and can be used to start big cultures growth. Glycerol stocks were composed of 300 µl cells and 700 µl LB-glycerol 1:1. One glycerol stock was used to prepare a pre-inoculum of 120 mL in kanamycin-containing LB medium, grown overnight at 37°C with 200 rpm shaking. The day after, six 5L flasks, filled with 1L of LB medium each, were inoculated with 20 mL of pre-inoculum. The cell cultures were grown at 37°C with 200 rpm shaking till an optical density at 600 nm of 0.6-0.8 was reached. At this phase no antibiotic was added, whereas anti-foam was dispensed in few drops. The induction of protein expression was performed by adding 250 µM IPTG. Cells were left in incubation overnight at 18-20°C with 130 rpm shaking. About 16 hours after induction, the cells were separated from the medium by low-speed centrifugation at 3600 RCF, collected with a spatula and transferred in sterile falcon tubes, which were flash-frozen in liquid nitrogen. Cells were kept in freezer at -20°C. A 6 L culture yields around 18 grams of cells in total. The same protocol for expression was used on both the wild type AXS and the C100A mutant enzyme and this also applies to the following steps of purification. The further discussed mutants, which were produced in Graz and there tested for biochemical characterization, underwent identical expression and purification procedures, though possibly with different columns and machines. The mutant was designed upon sequence alignment with UXS, finding position 100 was in the active site and presented a possibly relevant substitution. After structural determination of AXS wild type, the mutant underwent crystallization trials to obtain the structure of the protein with the substrate bound in the active site.

II.2.2 Protein purification

Nine grams of harvested cells were dissolved in 50 mL of buffer A composed of 20 mM HEPES pH 7.5 and 50 mM NaCl. An excess of DNase and protease inhibitors (Leupeptin, Pepstatin and PMSF), together with lysozyme, was added to the cell suspension. For cell disruption, sonication was performed on the sample for 5 cycles in pulse mode (ultrasonic intervals of 7 seconds were alternated to static intervals of 9 seconds, leaving the sample in ice, at an amplitude between 70% and 80% for 1 minute and 35 seconds). The cell lysate was centrifugated for 45 minutes at 70k RCF and the resulting pellet was discarded, while cell-free extract was collected, pooled and filtered to remove impurities. Using a super-loop, the extract was then loaded onto a 5 ml Nickel His-Trap column (GE healthcare) for affinity purification equilibrated with three column volumes of buffer A. During loading, the flow-through, containing all the other proteins that did not bind to the column, was collected for precaution. After that, three column volumes of buffer A were flowed through the column to wash out all weakly and not specifically bound proteins. The AXS enzyme, whose C-terminal His-tag interacts with the matrix functionalized with nickel ions, was then eluted in a single peak using 60% buffer B, having the same composition of buffer A plus 250 mM imidazole. Since we had been using an ÄKTA system, which allowed us to monitor absorbance of the sample passing through the column at 280 nm (detecting the peptide bond), we could conveniently collect only the portion of elution corresponding to the absorbance peak (Fig. II-4). Peak fractions were nevertheless analysed by SDS-PAGE to verify the presence of the enzyme and check the purity (Fig. II-5).

The eluted protein sample was pooled and concentrated up to 35 mg/mL, then it was divided in aliquots of 500 μ L, which were flash-frozen in liquid nitrogen and stored at -80°C. An AXS sample was used to perform size-exclusion chromatography on a Superdex200 column equilibrated in buffer A (Fig. II-6), which allows to separate molecules depending on their size. The AXS protein was eluted in a monodispersed single peak consistent with a dimeric form of the protein. The sample was concentrated up to 92 mg/mL.

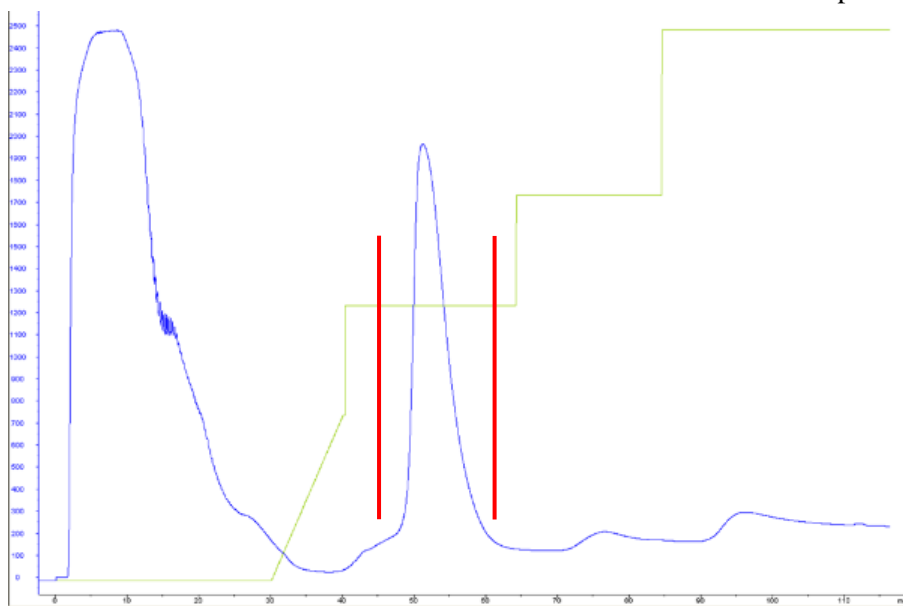


Figure II-4: affinity chromatography on AXS sample as described above. The green and blue traces show buffer B concentration and absorbance value at 280 nm, respectively. Between the two red lines is the peak collected and pooled for further purification.

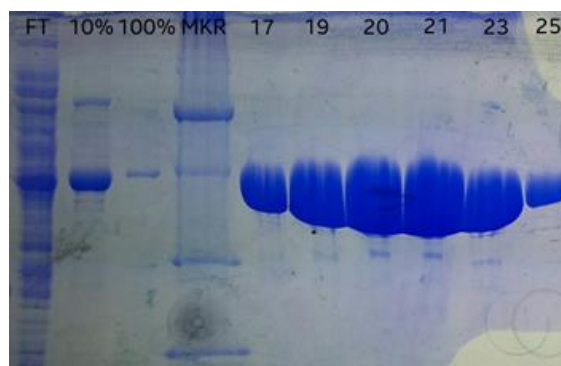


Figure II-5: SDS-PAGE analysis of the affinity chromatography fractions during AXS purification. From left to right the flow through (FT), the wash unbound (10%), the 100% buffer B wash after protein elution, molecular weight marker proteins (MKR) and the peak fractions eluted with 60% B. The AXS band shows a molecular weight around 45 kDa (same as the marker band), consistent with the calculated 44.7 kDa. AXS was so over-expressed to be also present in wash unbound and buffer B fractions.

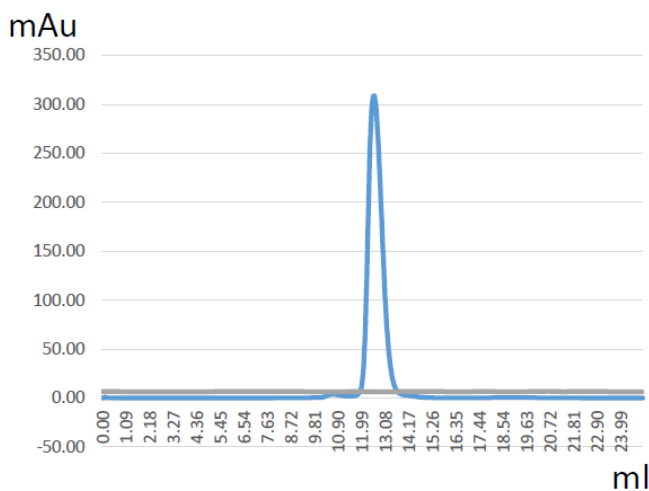


Figure II-6: peak of size-exclusion chromatography performed on an AXS affinity-purified sample. The Superdex 200-HR column used allowed elution of a sharp peak consistent with a protein dimer. The sample is very monodispersed and the contaminant present in very limited amount elutes earlier, so it is efficiently separated.

II.2.3 Crystallization experiments

The first parameter to be assessed for crystallization experiments is protein concentration. To determine the proper concentration to be used, a pre-crystallization test (PCT) can be performed [17]. A PCT kit from Qiagen consists of four reagents, the first two of which have low precipitants content and are used to prepare the preliminary crystallization drops. Protein concentration at this step should range at least from 5 to 20 mg/mL. If a light precipitate is observed, then it is worth trying the tested concentration range for many conditions, while if heavy precipitate is found, it means that, even with the low precipitant concentration, the protein must be diluted. In case no precipitation is observed at all, this means that either the protein or the precipitants concentration is not high enough. When this last case occurs, the protein can be mixed with the other two reagents, having higher precipitants concentrations. Formation of a light precipitate is interpreted as a positive result and it is a suggestion to proceed with the crystallization screening using the tested concentration. In the other cases, the kit indicates a series of actions

to be taken, such as diluting the sample or perform a native gel electrophoresis to assess its homogeneity.

Due to the high solubility of AXS, the PCT was positive for different concentration values, up to 100 mg/ml. Many concentrations were then tested, from 35 to 90 mg/ml. The screening was done with a Douglas Oryx 8 robot (Fig. II-7 right) to use a minimal amount of sample and test as many conditions as possible. Later, commercial kits were used for manual screening in vapour diffusion, both in sitting and hanging drop configurations. Also, many manual screenings were performed with in-house prepared solutions, to test a vast range of concentrations for ammonium sulphate, low and high molecular weight polyethylene glycol (PEG) and 2-methyl-2,4-pentanediol. Crystallization plates were always prepared in double and stored at 20°C and 4°C to maximize output (Fig. II-7 left). For each condition, a drop was also prepared with the buffer in which the protein was dissolved, as a control in case of salt crystals formation. The drops were checked the following days after their deposition for the first week and then once per week. Depending on the results, the most promising conditions were optimized by manual crystallization in vapour diffusion.

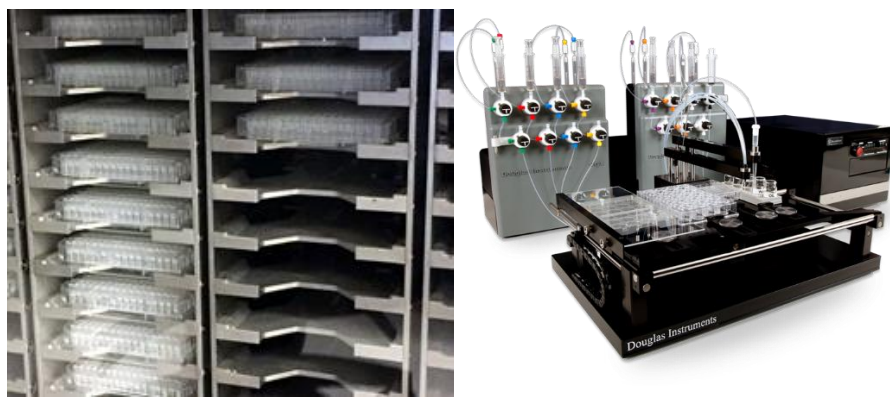


Figure II-7: on the left, crystallization plates stored for crystal growth at controlled temperature. On the right, the first model of crystallization robot that was put on the market by Douglas Instruments, an Oryx 8.

Since AXS was particularly reluctant to crystallize, even dialysis was tried. Even if the method is known to work well, its use has diminished over the years, mostly because of the large volumes required to set up each condition and the impossibility to automate the process. An AXS sample of volume 22 μ l was put in glass dialysis buttons and sealed with a single layer of dialysis membrane, cut to proper dimensions. The sealed button was then soaked into a glass vial containing 10 ml of reservoir solution. Buttons were periodically extracted from the solutions with tweezers and checked under the microscope in the first hours after soaking and on the following days. Massive precipitation could only be observed, and the method was dismissed.

The commercial kits used on AXS samples in vapour diffusion were eleven, each one composed of 96 conditions. Plates were prepared with apoprotein or in cocrystallization with different combinations of the inhibitors UDP and UDP-glucose, alone or with the cofactor in reduced or oxidized state. Also, the substrate UDP-glucuronate was used in cocrystallization, after preliminary results were obtained. Considering also the hand-made experiments, the two temperature values tested and the many different protein concentrations, we can easily count a number over 8,000 single conditions has been tried.

Since the poly-Histidine tag used for purification is known to impair crystallization in some – quite rare – instances [18], I produced a cleavable version of the AXS-6xHis construct. For this task, the PCR protocol for Phusion DNA polymerase from ThermoFisher was used to insert the TEV cleavage site between the coding sequence of the protein and the tag. The same crystallization experiments performed on the original construct were replicated on this variant, but no positive hits could be found.

Limited proteolysis was performed in sitting drop configuration; it consists in adding a small – catalytic – amount of the protease trypsin to the sample. In this way, the most flexible parts of the protein structure should be cleaved, leaving only the more structured central fold to crystallize by generating tighter packing interactions. This strategy was tested on the *wt*AXS construct in many handmade sitting drop conditions as well as in commercial solutions dispensed in high-throughput by the robot, but no positive hits were found.

After different approaches were tried, passing through lysine methylation, site-directed mutagenesis and SAXS characterization – which will all be described in the following sections – AXS crystals were eventually obtained. The crystals of *wt*AXS grew in a mix containing 0.8 M sodium phosphate and 0.8 M potassium phosphate as precipitants/salt additives, and 0.1 M HEPES sodium salt pH 7.5 as buffer. UDP and NAD⁺ were added in cocrystallization at 2 mM concentration for both ligands. To safely freeze crystals, they were treated with a solution containing 20% v/v cryoprotectant. After trials with PEG400 and glycerol it was found that the latter was better serving the purpose, enhancing the diffraction images quality. The ESRF (Grenoble, France) and the SLS (Villigen, Switzerland) synchrotron beamlines were both used to irradiate crystals, and both provided similar datasets for what concerns resolution and completeness (see next paragraph). The crystals of the AXS mutant C100A, which is further discussed, have been grown in same conditions of the *wt*AXS, but adding 2 mM UDP-glucuronate and 2 mM NADH as ligands. Glycerol was used as cryoprotectant in 20% v/v concentration.

II.2.4 Crystallography

X-ray diffraction data on crystals of *wt*AXS were collected on Dectris detectors of the Pilatus generation at the ESRF (Grenoble, France) and the SLS (Villigen, Switzerland) synchrotron beamlines. After preliminary orthogonal characterization, the data collection strategy suggested by ISPyB at ESRF was used to obtain best completeness. The automatic MASSIF data collection at ESRF was also exploited to collect data. This service is fully automated, from sample loading to crystal centring, preliminary orthogonal characterization, collection strategy, indexing, scaling and, if sequence is provided, it can even try to solve the structure. For the best diffracting crystal of *wt*AXS, 1500 images were collected at a *phi* angle of 0.1°, for a total of 150°, and exposure time of 0.05 seconds. Detector distance was set to have a maximum resolution of 2.5 Å, in order to avoid data loss, despite final resolution was found to be 3 Å (Table II-1). Beam wavelength was set at 1 Å and transmission of the beam was 20% (when working on ID29 beamline at ESRF). Because of the severe radiation damage observed over time in the dataset, only the first 850 images were further processed in indexing and integration by using XDS, which provided an *hkl* file. The *hkl* file was input

to *aimless* in the CCP4 package, which operated scaling and intensities merging.

The output *mtz* file of the processed data was the first to be used for structure solution and following refinement steps. The program Phaser was used to operate molecular replacement using the coordinates of the 3SLG PDB entry, which is referred to the protein PbgP3, from *Burkholderia pseudomallei*. The 37% sequence identity between PbgP3 and AXS made some automated tracing necessary to make the model building process faster, even if Chainsaw was used to prepare the sequence. To do so, I used programs from the Phenix package, namely *FEM* to enhance the electron density map and then *Autobuild* to trace secondary elements. The output model underwent substantial manual tracing to remove errors of out-of-register residues placement, and to fill the electron density of loops. The structure was solved in space group P3₂21 (number 154) at 3.0 Å resolution, with two molecules in the asymmetric unit. All residues were traced except for the seven N-terminal residues and for a flexible loop, where residues 61-68 could not be traced. Multiple cycles of manual refinement in Coot were alternated to Refmac5 runs. The quality of data was assessed constantly by checking Ramachandran plots, rotamers and chiral volumes distributions and by submitting the model to MolProbity.

For the mutant C100A, data were collected using similar procedures followed for the wt AXS, except that, already knowing the cell parameters and the issue of radiation damage, I limited data collection to 900 images. After the usual steps in *XDS* and *aimless*, the wtAXS model was used as reference to solve the structure of the C100A mutant at 3.5 Å resolution (Table II-1).

II.2.5 Chemical methylation

Chemical methylation of the surface lysine residues was tried in order to make AXS less soluble. This approach is aimed at reducing protein solubility by masking charged residues with methyl groups. To do so, the pure protein sample can be diluted to a concentration ranging between 1 and 10 mg/ml and then processed with a methylation kit. I used the Jena Bioscience Methylation kit, which is provided with all the reagents and instructions for sequential steps and timing of reactions. I followed the protocol performing the

operations in a cold room at a constant temperature of 10°C. The tested concentrations in different tubes were 0.5, 1, 3, 5 and 7 mg/ml. During the overnight reaction step, the protein was precipitating in a concentration-dependent way, showing much less precipitate at concentrations lower than 7 mg/ml. Even when no precipitation was observed, the small amount of processed material did not display the usual behaviour in size exclusion chromatography. This result is discussed here because of its concise nature. After multiple trials, the methylation strategy was abandoned, since it was decreasing protein's quality and usable quantity.

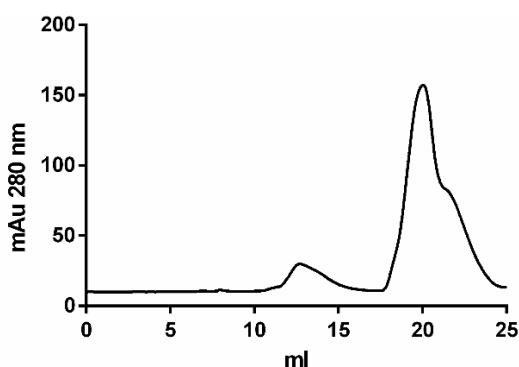


Figure II-8: elution profile of size exclusion chromatography performed on sample treated with methylation kit. The correct elution volume presents a low, broadened peak, while most of the absorbance is registered for low molecular weight molecules, being reagents in the kit.

II.2.6 Site-directed mutagenesis

As the methylation approach failed, the solubility issue was addressed by site-directed mutagenesis of surface residues. These were targeted as they can form charged patches, able to promote repulsive forces. To determine which residues to mutate, the SERp server was used. This tool, provided by the University of California, generates a three-dimensional model of the protein, based on homology modelling, with emphasis on secondary structures, and assigns entropy values to the different portions composing it. By doing so, three charged exposed clusters of residues were found. These sites were targeted by mutagenesis, operated on the original construct of AXS, by applying the PCR protocol for Phusion DNA polymerase from ThermoFisher.

Three mutants were made, all having mutation to Alanine for the residues K167-E168, E179-K180-Q181 and K328-E329, respectively. The surface mutants were expressed and purified as the original construct and crystallization experiments were set up in the same way. In this case the high concentrations of the wild type could not be reached, as precipitation was observed in drops when passing 60 mg/ml. Even in this case, no positive hits could be found.

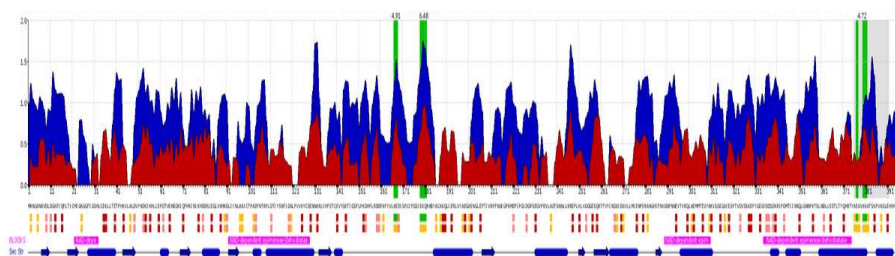


Figure II-9: entropy profile calculated for the *a priori* AXS model generated by SERP over the entire sequence. The parts highlighted in green are the charged patches exposed on the surface, which the program suggests mutating in non-charged residues.

II.2.7 Small angle X-ray scattering

While the theoretical description of this methodology escapes the aim of this thesis, a brief introduction is due. This low-resolution method is used on biological samples to determine their shape and oligomeric state. Because of its limitations in resolution, it is particularly informative in the study of complexes, to understand the reciprocal arrangements of single components, the structure of which is determined by finer methodologies. Even if limited in resolution, SAXS has the non-trivial advantage of being a technique that works on samples in solution. As the name suggests, it is also based on X-ray diffraction but at small angles of deviation from the primary beam, which results in poor intensity and resolution. A SAXS profile is generated through a Fourier Transform of the scattering information, and it is plotted in Guinier derivatization (Fig. II-10 left) as the natural logarithm of the scattering intensity $\text{Log}(I)$ in function of the scattering vector $S=4\pi\sin\theta/\lambda$. From Guinier analysis the radius of gyration of the protein is calculated and the size of the

molecule – so also its molecular weight, especially for globular objects – can be calculated [20]. Another convenient way to visualize SAXS data is the Kratky plot (Fig. II-10 right), which is obtained multiplying the $\text{Log}(I)$ values by S^2 . This mode of representing data is suited to evaluate the overall folding of the molecule and its aggregation state. It is important to subtract the solvent signal before a SAXS experiment, as the technique is extremely sensitive and background noise would impair data reliability. To do so, few microliters of the same buffer in which the protein was maintained during size-exclusion chromatography must be used; same buffer means the one left after the chromatography run. SAXS experiments can be performed in batch mode, with a small volume of sample which is taken to the synchrotron (although the sample can undergo some kind of degradation during the transportation) or alternatively a sample can be eluted directly on site through an HPLC connected to the beamline (SEC-SAXS).

For crystallization purposes, SAXS can be informative. The profile of a SAXS curve is telling us about the aggregation state of a protein sample, its raw conformation (globular or not) and its solubility at a precise concentration. This last information is only useful when different concentrations are tested (and in general in SAXS all data must always be related to measurements taken on samples at different concentration, at least three points). As multiple concentrations of protein are measured to understand the changes in the profile with respect to the crowding of molecules, one concentration value can be found for which the sample is as much concentrated as possible without signs of unspecific aggregation. In the Figure II-10 instance, the same sample of AXS, measured in batch mode, is plotted in both the described ways.

The protein samples were passed through Superdex 200 and a concentrated stock was prepared, to make dilutions at 10, 20, 40 and 80 mg/ml for batch analysis and another 10 mg/ml sample for SEC-SAXS. The samples were sent to the ESRF where some of my colleagues performed the diffraction experiment, taking back the datasets, which were then evaluated by the program ScÅtter. As expected from literature and from sample handling in the lab, the Guinier analysis confirmed a molecular weight of about 88 kDa, consistent with a dimer of the 44.7 kDa protein. This information comes from evaluation of the signal for small q values. In the same way, the Kratky plot

confirmed the presence in solution of a globular protein, as shown by the bell-shaped peak, and the monodispersity of the sample was assessed as well, as it can be seen from the signal at high q values.

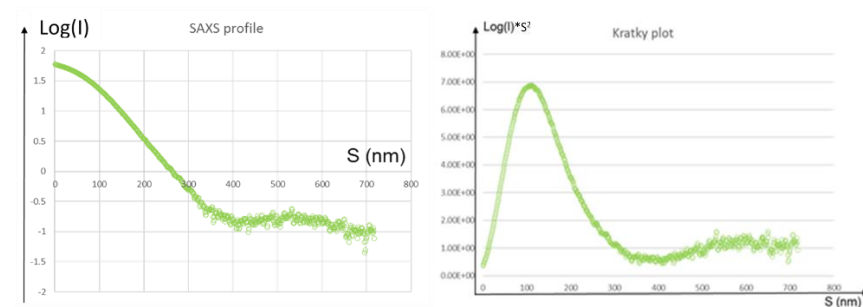


Figure II-10: on the left, Guinier analysis of a 10 mg/ml sample of AXS eluted from size exclusion chromatography just before measurement. On the right, Kratky plot obtained from the same data.

While these data came from the SEC-SAXS experiment, and were shown here as they do not constitute a real result, the ones coming from analysis of different concentrations of AXS will be presented in the next section, as they make an interesting point, which resulted in helping with protein crystallization. Even if SAXS can be used as a pure structural technique, in the case of my work on AXS it was only applied as a fine analytical tool to better understand the protein. In this respect no models were generated on the base of the collected curves, as working on an enzyme would have most likely given a spherical object with poor structural features.

II.3 Results and discussion

The SAXS experiments performed on AXS were very informative to understand the biophysical properties of the protein, above all its solubility at different concentrations. In this respect, the experiments described in materials and methods section revealed a repulsive behaviour of the protein when its concentration was increased. Such a phenomenon has been already

described [19] and the theory behind it is that the interparticle repulsion, mostly caused by superficial charges, is responsible for the extremely high solubility of the protein. As we can see in Figure II-11, when samples of proteins with this behaviour are tested in SAXS at different concentrations, the Guinier analysis is not showing sign of aggregation. On the contrary, the apparent mass of the object decreases when increasing the concentration. This is consistent with higher Brownian motions in solution, which are due to repulsion, and result in faster objects, the mass of which seems smaller than the one of the same objects at lower concentrations.

The SAXS analysis provided a rationale for the failure of the preliminary crystallization experiments. Indeed, the absence of precipitate in AXS crystallization conditions was not due to too low sample concentration but instead to the propensity of the protein to stay in solution because of the intraparticle repulsion.

Following this observation, massive screening of AXS crystallization conditions was performed, lowering the concentration of the protein to 25 mg/ml. Crystals of AXS obtained in the mentioned conditions were regularly shaped solids, elongated along a hexahedral base, as it can be seen in Figure II-12. While some presented multiplicity in their growth, most of them were single and quite robust. Upon cracking test with a needle, they tended to break in a regular fashion instead of dissolving as it generally happens with protein crystals. However, the biological nature of the crystalline objects was clear from their physical response to light polarization. In the same conditions AXS was co-crystallized with different ligands and using the C100A mutant form.

The structure of *wt*AXS and AXS C100A presents a dimer in the asymmetric unit (Fig. II-13), as expected from the size-exclusion chromatography profile (Fig. II-6). The root mean square deviation (rmsd) value between the two structure is 0.2149 Å. The structure of AXS was solved using XDS on the diffraction images and then the programs of the CCP4 suite for scaling, phasing and refinement. As search model for molecular replacement the chain A of the PDB entry 3SLG was used. For C100A mutant the phases were taken from the wild type structure. The interface between the two molecules is represented by two α -helices, a β -strand and two loops. The α -helices in particular provide quite an extensive contact surface by establishing

hydrophobic interactions. In Table II-1 the crystallographic parameters of the two datasets are reported.

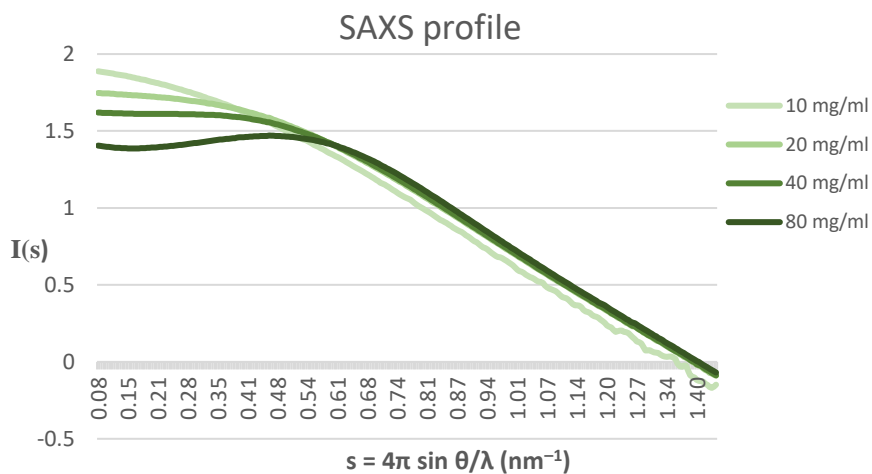


Figure II-11: Guinier analysis of superimposed SAXS data from AXS samples measured in batch method at different concentrations. While the rest of the plot is matching, for small q values (left side) the signal increases with the sample concentration.

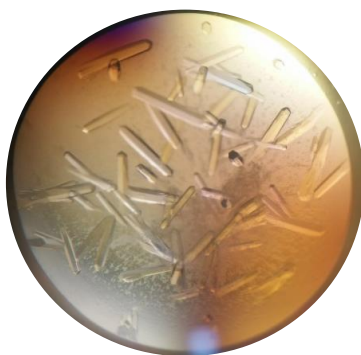


Figure II-12: crystals of AXS reproduced by hand in the original conditions. Light polarization can be seen from the different colours of the objects depending on their orientation. The objects were indeed transparent. As size reference, the drop in the picture has a total volume of 4 μl .

	AXS wt	AXS C100A
Resolution range	49.1 - 3.0 (3.2 - 3.0)	48.4 - 3.5 (3.8 - 3.5)
Space group	P ₃ 2 ₁	P ₃ 2 ₁
Unit cell (Å), (°)	146.0 146.0 132.8 90 90 120	144.6 144.6 130.5 90 90 120
Total reflections	143316 (20101)	173716 (38422)
Unique reflections	32264 (4488)	20833 (4867)
Multiplicity	4.4 (4.5)	8.3 (7.9)
Completeness (%)	98.9 (95.4)	99.8 (7.9)
Mean I/sigma (I)	9.2 (1.0)	13.1 (1.9)
R-merge (%)	0.146 (1.291)	0.151 (1.138)
CC1/2	0.992 (0.378)	0.998 (0.844)
R-work (%)	0.207 (0.369)	0.200 (0.360)
R-free	0.253 (0.377)	0.243 (0.317)
Number of non-hydrogen atoms	6009	5870
Protein	5848	5723
Ligands	138	147
RMS (bonds) (Å)	0.013	0.013
RMS (angles) (°)	1.82	1.91
Ramachandran favoured (%)	94.44	91.85
Ramachandran allowed (%)	4.75	7.34
Ramachandran outliers (%)	0.81	0.82
Average B-factor	77.85	125.51
Protein	80.35	131.0
NAD	74.30	125.85
UDP	70.00	-
UDP-GlcUA	-	124.15

Table II-1: crystallographic table of AXS wt and AX C100A. In parentheses the high-resolution shell values.

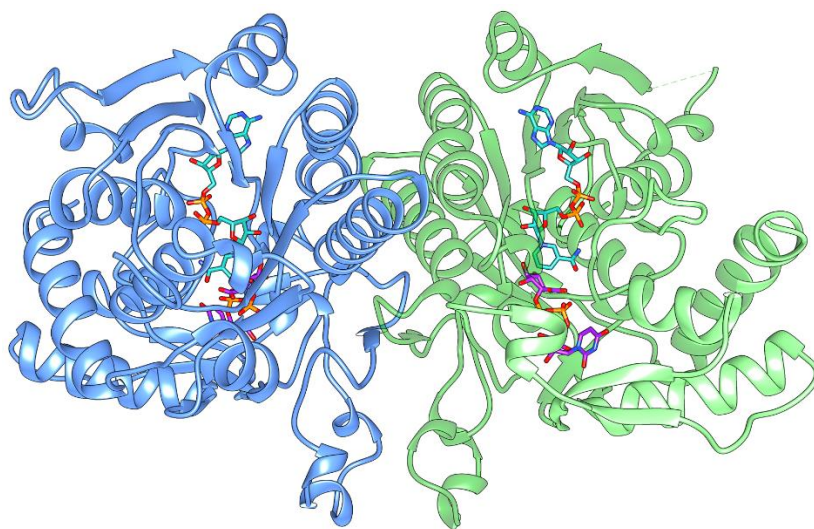


Figure II-13: AXS C100A dimer with chain A in blue and chain B in transparent green. NAD is represented in cyan carbons and UDP-GlcUA in purple carbons; oxygen atoms are red, nitrogen atoms blue and phosphorous atoms orange. Hydrophobic interactions are established between the two long α -helices of each monomer and between the elongated unstructured loop in the lower part of the image.

In *wt*AXS, the inhibitor UDP, which is the aglycone of the UPD-GlcUA substrate, takes the very same position of the latter, as observed in the structure of the mutant. The structure of the C100A mutant, which had already been designed based on sequence alignment with UXS, was indeed obtained with the purpose of observing the ligand inside the less active enzyme (see further biochemical data in tables II-2 to II-5). While no sugar is present in the wild type structure, where UDP is bound, a phosphate is holding the equivalent position, quite remarkably, given the close distance with the pyrophosphate group of UDP. The phosphate molecule is clearly coming from the crystallization condition, where 0.8 M sodium phosphate and 0.8 M potassium phosphate are present. The phosphates of UDP and UDP-GlcUA establish interactions with the three arginine residues 182, 260 and 341. In this respect, the presence of the extra phosphate from reservoir solution in UDP-bound structure finds an explanation, as stabilized by local charges. No differences in folding are observed between the two structures and the divergence is limited even when comparing AXS to the homolog UXS (Fig.

II-16). In this case, the core of the Rossmann fold is conserved, with loops having different length in the two proteins. The rmsd between AXS (chain A) and UXS monomer is 1.630 Å.

Before solving the structure of the C100A mutant, the 3.0 Å resolution structure of the wt AXS was used to model UDP-GlcUA in the catalytic pocket. To better fit the substrate, we took advantage of the published structure of ArnA (PDB 1Z7E), which shares the same residues of UXS in the active site and was solved with UDP-GlcUA bound [21, 22]. The model we developed for substrate binding was confirmed when the mutant structure was obtained at 3.5 Å resolution, finding the sugar conformation to be the same as that present in the ArnA active site.

The NAD cofactor is hosted by the Rossmann fold, that constitutes the core of AXS. Here, the sidechain of Arg235 and the main chain nitrogen atoms of residues 27 to 30 are involved in cofactor binding. The adenine of NAD is not involved in major interactions with the environment surrounding it. Concerning cofactor binding, it should be noticed that, while the molecule was added in excess for crystallization experiments – so that it was not a limiting factor – it was actually present in complex with the protein from the purification step. Since during purification I monitored protein concentration with a Nanodrop spectrophotometer, I could observe a slight shift of the typical peak at 280 nm toward the direction of 260 nm. This shift was consistent with the variation in protein concentration and reproducible amongst different purifications. While such a phenomenon was first considered a nucleic acid contamination, upon purification with Heparin column this possibility was rejected, and the shift interpreted as retention of the NAD cofactor bound to the enzyme. Indeed, following activity studies confirmed slight activity of the purified protein, even in absence of an excess of cofactor. Also, when a purified protein sample was first measured and then boiled, the soluble fraction – upon elimination of the denatured protein pellet – was still containing NAD and NADH (Fig. II-14) released by the protein.

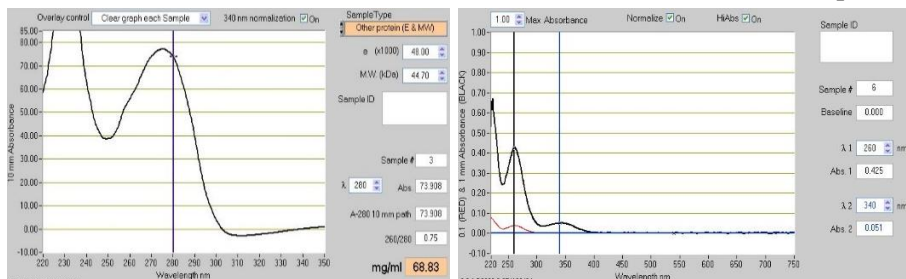


Figure II-14: absorption spectrum of AXS sample measured on a Nanodrop before boiling (left) and spectrum of the same sample measured upon boiling and protein removal by centrifugation (right).

The most noticeable differences in the active site of AXS, with respect to hUXS, are the cysteine residues 100 and 140 in place of Ala79 and Ser119 (Fig. II-15). While the rest of the residues involved in coordination of the ligands binding is consistent and even the supposed active residues are conserved [10], these two residues attracted my interest because of the position they occupy with respect to the sugar moiety of UDP-GlcUA. Based on such observation I developed a model, which will be better explained in the conclusions of this chapter. In order to prove this model, some mutations were proposed to our collaborators in Graz. The mutant C100A, which was later characterized also from the structural point of view, was meant to re-establish the minor hindrance observed in the UXS active site, and to remove the nucleophilic character. In the same way, the mutant C140S was designed to resemble the UXS environment. The combinatorial mutant C100A-C140S was supposed to impair UDP-Api formation, since an environment identical to UXS is recreated in this mutant. Mutants design, expression and purification were done in Graz. Our collaborators performed also substrate conversion studies by chromatography and NMR characterization of products. In tables II-2 to II-5, the conversion experiments run on *wt*AXS and mutants, and measured after 24 hours, are summarized.

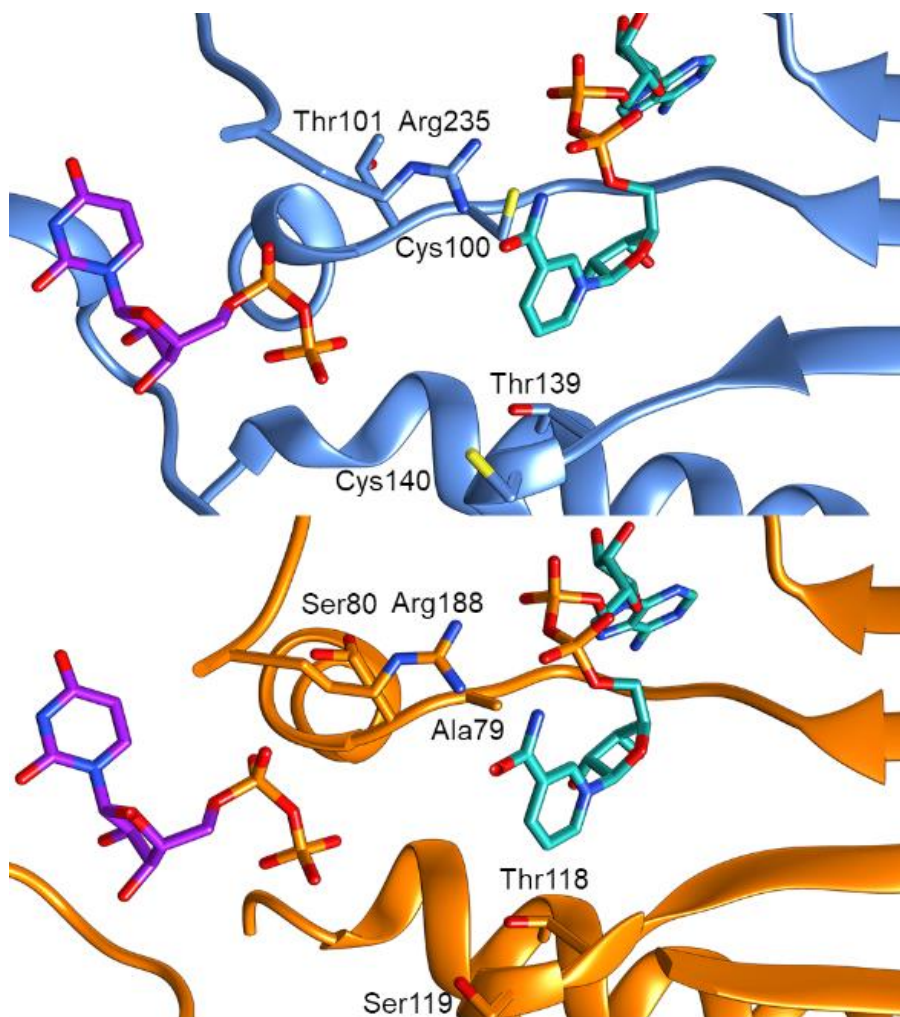


Figure II-15: comparison of the active sites of AXS (blue) and UXS (orange) showing the conserved folding around the ligands (UDP in purple carbons and NAD in cyan carbons, for both models) and their position in the structure. Most of the residues, not shown for sake of clarity, are conserved and occupy the same positions, like Arg235/188 and Thr139/118. However, some catalytically relevant aminoacids of AXS, like Thr101, Cys100 and Cys140 are substituted in UXS for Ser80 (shown in double conformation), Ala79 and Ser119.

AXS has a quite high selectivity for UDP-Api formation, around 60%, with a side selectivity of 25% for UDP-Xyl and minor formation of the reaction intermediate UDP-4-keto-Xyl (Tab II-2). Since UDP-Api is an unstable compound, it is hydrolysed to UMP and Apiose-cyclic phosphate. For this reason, cyclic phosphate is measured as well, in small amount. Note that the measurement of this compound is actually based on the equimolar release of UMP, which may even derive from hydrolysis of the UDP-GlcUA and UDP-Xyl. The first mutant C100A (Tab II-3) strongly reduces accumulation of the main product UDP-Api, while the secondary product UDP-Xyl increases in concentration. At the same time the 4-ketointermediate, which is generated upon decarboxylation of position 5' of UDP-GlcUA and oxidation in position 4', is also more represented than in the sample after conversion with wild type enzyme. This reversion of products ratio is consistent with a reduced capability of the enzyme to perform the ring contraction reaction. Also, the increased intermediate presence is explained by a reduced activity of the enzyme which releases the compound before catalysing the second step of the reaction. The behaviour of mutant C140S (Tab II-4) shows an even stronger effect of the substitution, as UDP-Api production is reduced in a similar way to the C100A mutant, but it also accumulates more intermediate, meaning that even the reduction of position 4' which leads to UDP-Xyl release is strongly affected. For the double mutant C100A-C140S (Tab II-5), the effect is a half-way, with 4-keto-intermediate formation between the values of the previous mutants and a UDP-Api formation which is also between the two values.

Even if a complete knock out of the UDP-Api formation was not achieved, the mutant enzymes prove the essential role of the two cysteines in performing the reaction in the proper ratio. The influence of second sphere actors, meaning other residues neighbouring the active site, must be more important than supposed and deserves further investigation. It is likely that a subtle balance in pK_a values of the residues composing the active site is governing the reaction ratios. Despite the impossibility to impair UDP-Api formation, the effect of the single and double mutations helped developing the following model. The comparison with the structural model of C100A AXS, where the substrate is bound, allowed assigning to each residue a network of interactions with the atoms of the sugar moiety and the cofactor (Fig. II-16).

Compound	$\mu\text{moles } (\mu\text{mol})$	Selectivity (%)
Cyclic phosphate	106	6,02
UDP-Apiose	1069	60,84
4- Ketointermediate	138	7,83
UDP-Xylose	444	25,30
Total	1757	100,00

Table II-2: quantification of products corresponding to the NMR signals recorded after 24 hours conversion experiments with *wf*AXS.

Compound	$\mu\text{moles } (\mu\text{mol})$	Selectivity (%)
Cyclic phosphate	107	5,33
UDP-Apiose	381	19,05
4- Ketointermediate	365	18,29
UDP-Xylose	1147	57,33
Total	1757	100,00

Table II-3: quantification of products corresponding to the NMR signals recorded after 24 hours conversion experiments with AXS C100A mutant.

Compound	$\mu\text{moles } (\mu\text{mol})$	Selectivity (%)
Cyclic phosphate	14	0,72
UDP-Apiose	479	23,92
4- Ketointermediate	1086	54,31
UDP-Xylose	421	21,05
Total	2000	100,00

Table II-4: quantification of products corresponding to the NMR signals recorded after 24 hours conversion experiments with AXS C140S mutant.

Compound	$\mu\text{moles } (\mu\text{mol})$	Selectivity (%)
Cyclic phosphate	0	0
UDP-Apiose	319	23,09
4- Ketointermediate	562	40,65
UDP-Xylose	501	36,26
Total	1383	100,00

Table II-5: quantification of products corresponding to the NMR signals recorded after 24 hours conversion experiments with AXS C100A-C140S mutant.

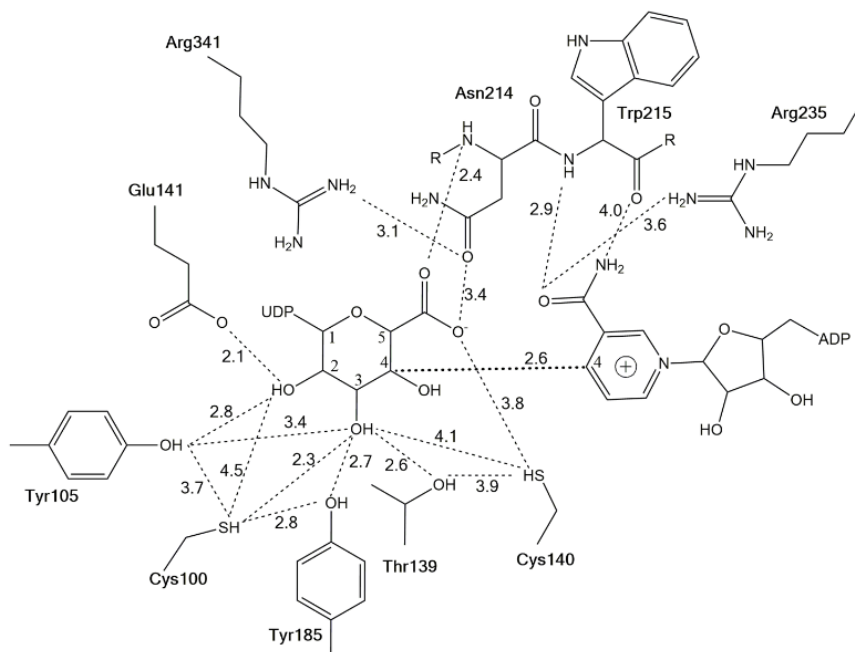


Figure II-16: two-dimensional representation of the H-bonding distances in AXS active site. The complex network of distances is expressed in Angstrom. Numbering of the carbon positions is reported for the glucuronic acid moiety of UDP-GlcUA. Carbon 4 of the nicotinamide is also labelled.

II.4 Conclusions and perspectives

Altogether, the results allowed us to conclude that Cys100 and Cys140 are essential to efficiently produce UDP-Api. To explain this, a model was developed and described below, which partially overlaps the one already theorised in literature. The promiscuity of AXS in product formation is also explained in accordance to this model.

The first step in AXS reaction is the oxidation of UDP-GlcUA to the UDP-4-keto-Xylose intermediate, which occurs upon decarboxylation in position 5' and oxidation of the hydroxyl group in position 4' (Fig. II-2). In mutant C140S and C100A-C140S the accumulation of this intermediate can be explained by a lack of strong interactions which are established by the more

nucleophilic sulphur atom of cysteines. In this scenario, the contribution of the ligand-binding residues seems to be more relevant than the active ones, so that the ligand leaves the pocket before the reaction is completed and so before oxygen in position 4' is reduced, producing the intermediate. This phenomenon also occurs with the wild type enzyme but to a much lesser extent; the reaction is pictured in Figure II-17. At the same time, less UDP-Api is formed, as Cys140 is the final reducer of the oxygen in position 3' after the branched chain sugar has been formed. It must be noticed that this balance of different driving forces is also actual in the wild type enzyme, as UDP-Xyl and the intermediate are also synthesised by that, meaning that AXS has no way to make the action of the two cysteines prevail on other effectors.

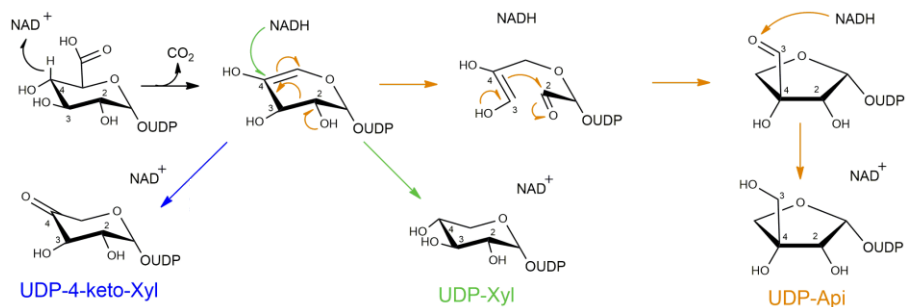


Figure II-17: reactions catalysed by AXS. Formation of UDP-4-keto-Xyl intermediate (blue), UDP-Xyl (green) and UDP-Api (orange).

The results on mutant C100A highlight that the presence of Cys100 is important – but not fundamental – for the ring opening to occur efficiently. This aminoacid is likely to facilitate a shuttling of the proton bound to the oxygen in position 3' and in this way to make the ring prone to condensation. The mutation C100A could be buffered by the close-by tyrosines 105 and 185, which may act as well as shuttles. I speculate that Cys100 coordinates the oxygen in position 3' of the sugar, before the furanosidic form is established, while this task is later performed by the already discussed Cys140.

For what concerns the double mutant, mutation C100A makes the ring condensation less likely to start and mutation C140S reduces the consolidation of the furanose structure. In the double mutant, the ratio UDP-Xyl/UDP-4-keto-Xyl is shifted toward the formation of the first, as the mutations are partially regaining the UXS-specific configuration.

The incapability to avoid UDP-Xyl synthesis is likely due to the presence of Tyr185, which is conserved in UXS (Tyr147). Such residue has already been described as an interactor of the oxygen in position 4' [23]. By manually modelling the UDP-Api product into the active site, it is possible to see the fair amount of interactions conserved by the molecule with the surrounding residues. Upon displacement of C3 from the plane of the ring, the oxygen in position 3' reaches a position closer to the C4 of NAD, still able to interact with Cys100. Also, tyrosine residues 105 and 185 maintain their interactions respectively with oxygen 2' and the displaced oxygen 3', while Thr139 keeps facing oxygen 4' at H bonding distance. The close distance of Tyr105 and Tyr185 to Cys100 is also compatible with the activation of such a residue to its thiolate form, making it ready to interact with oxygen 3'. Even lacking a mechanistic demonstration, on the basis of surface and cavity analysis we can speculate that the CO₂/HCO₃ molecules can easily leave the active site.

The structure of AXS reveals the unique features of this enzyme, and how it can perform a promiscuous reaction yielding two different products. The availability of structural data has been essential in guiding the rational design of mutants to assess the biochemical properties of AXS. The results of this work are of a dual nature. On one side, the basic research in biological field is enriched by understanding such a singular mechanism. The unearthing of the pathways leading to sugars biosynthesis in plants is important to understand plant growth and to develop methodologies in strengthening cultivations. These mechanisms are also target of many plant pathogens, which are becoming a bigger and bigger threat to large portions of the market and to food availability around the world. For example, the very well-known pathogen *Xylella fastidiosa* is targeting the UXS gene [12], impairing xylem's integrity. Also from the pure chemical perspective, the ring contraction catalysed by enzymes is still a poorly investigated aspect of biocatalysis, and the elegant way used by AXS to perform it is one of the few examples available. On the other hand, the possible applications of AXS in production

of rare activated sugars is an interesting perspective. In particular, the coupling of this enzyme to others working in upstream reactions [24], starting from activated glucose, and the coupling to Leloir glycosyltransferases [25]. In this perspective, we have already turned our attention to some candidates to be characterized. The combination of different enzymes in the right sequence may lead to the design of a pathway that goes from the recycling of industrial waste to high value activated sugars, used in pharmaceuticals and in research, and currently not available at affordable prices.

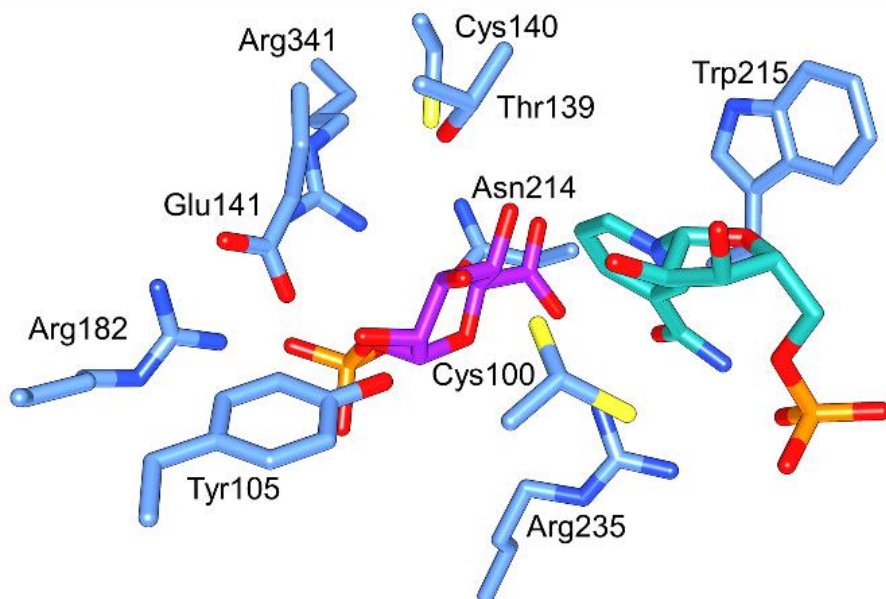


Figure II-19: three-dimensional model of AXS active site. The image combines ligands from two datasets. The glucuronic acid of UDP-GlcUA is represented in purple carbons and comes from the C100A mutant structure while the nicotinamide riboside of NAD is in cyan carbons and derives from the wild type AXS structure. Active site residues are represented as light blue carbons. Cys100 is represented in two conformations, a distal one – from the crystallographic structure – and a proximal one – introduced in the model to explain the reactivity. Despite being close to the active site, Tyr185 is not shown for the sake of clarity. The image is representative of all the residues which have been or will be mutated to assess their role in biocatalysis.

In future perspective, it will be important to assess the role of other residues facing the active site, as described in Figure II-19, and this is currently an undergoing process. To prove the retro-aldol cleavage mechanism occurring in ring-contraction [26] the usage of a fluoro-analogue [27], which behaves as an inhibitor of the enzyme and gets stuck in the active site is scheduled and the compound currently under preparation – the chemical is not available on the market and must be prepared by multienzyme reaction – in Graz. In this way, it will be possible to further prove the proposed mechanism and shed light on this singular reaction. In the overall framework of a project on this sugar-modifying class of enzyme, our current interest has also been attracted by homologs which work on different activated sugars. By studying these proteins we foresee to expand our knowledge on their fine reactivity and to better understand what we have already observed with AXS.

II.5 References

1. Grisebach H, Schmid R. Chemistry and biochemistry of branched-chain sugars. *Angew Chem Int Ed Engl.* 1972 Mar;11(3):159-73.
2. Matern U, Grisebach H. UDP-apiose/UDP-xylose synthase. Subunit composition and binding studies. *Eur J Biochem.* 1977 Apr 1;74(2):303-12.
3. Møllhøj M, Verma R, Reiter WD. The biosynthesis of the branched-chain sugar d-apiose in plants: functional cloning and characterization of a UDP-d-apiose/UDP-d-xylose synthase from *Arabidopsis*. *Plant J.* 2003 Sep;35(6):693-703.
4. Yin S, Kong JQ. Transcriptome-guided gene isolation and functional characterization of UDP-xylose synthase and UDP-D-apiose/UDP-D-xylose synthase families from *Ornithogalum caudatum* Ait. *Plant Cell Rep.* 2016 Nov;35(11):2403-2421.
5. Gatzeva-Topalova PZ, May AP, Sousa MC. Structure and mechanism of ArnA: conformational change implies ordered dehydrogenase mechanism in key enzyme for polymyxin resistance. *Structure.* 2005 Jun;13(6):929-42.

6. Savino S, Ferrandi EE, Forneris F, Rovida S, Riva S, Monti D, Mattevi A. Structural and biochemical insights into 7 β -hydroxysteroid dehydrogenase stereoselectivity. *Proteins*. 2016 Jun;84(6):859-65.
7. Perozich J, Kuo I, Wang BC, Boesch JS, Lindahl R, Hempel J. Shifting the NAD/NADP preference in class 3 aldehyde dehydrogenase. *Eur J Biochem*. 2000 Oct;267(20):6197-203.
8. Hu Z, Wells PG. Human interindividual variation in lymphocyte UDP-glucuronosyltransferases as a determinant of in vitro benzo[a]pyrene covalent binding and cytotoxicity. *Toxicol Sci*. 2004 Mar;78(1):32-40.
9. Götting C, Kuhn J, Kleesiek K. Human xylosyltransferases in health and disease. *Cell Mol Life Sci*. 2007 Jun;64(12):1498-517.
10. Eixelsberger T, Sykora S, Egger S, Brunsteiner M, Kavanagh KL, Oppermann U, Brecker L, Nidetzky B. Structure and mechanism of human UDP-xylose synthase: evidence for a promoting role of sugar ring distortion in a three-step catalytic conversion of UDP-glucuronic acid. *J Biol Chem*. 2012 Sep 7;287(37):31349-58.
11. Zhong R, Teng Q, Haghghat M, Yuan Y, Furey ST, Dasher RL, Ye ZH. Cytosol-Localized UDP-Xylose Synthases Provide the Major Source of UDP-Xylose for the Biosynthesis of Xylan and Xyloglucan. *Plant Cell Physiol*. 2017 Jan 1;58(1):156-174.
12. Alencar VC, Jabes DL, Menegidio FB, Sasaki GL, de Souza LR, Puzer L, Meneghetti MC, Lima MA, Tersariol IL, de Oliveira RC, Nunes LR. Functional and Evolutionary Characterization of a UDP-Xylose Synthase Gene from the Plant Pathogen *Xylella fastidiosa*, Involved in the Synthesis of Bacterial Lipopolysaccharide. *Biochemistry*. 2017 Feb 7;56(5):779-792.
13. Gebb C, Baron D, Grisebach H. Spectroscopic evidence for the formation of a 4-keto intermediate in the UDP-apiose/UDP-xylose synthase reaction. *Eur J Biochem*. 1975 Jun;54(2):493-8.
14. Houston K, Tucker MR, Chowdhury J, Shirley N, Little A. The Plant Cell Wall: A Complex and Dynamic Structure As Revealed by the Responses of Genes under Stress Conditions. *Front Plant Sci*. 2016 Aug 10;7:984.
15. Keegstra K. Plant cell walls. *Plant Physiol*. 2010 Oct;154(2):483-6.

16. O'Neill MA, Ishii T, Albersheim P, Darvill AG. Rhamnogalacturonan II: structure and function of a borate cross-linked cell wall pectic polysaccharide. *Annu Rev Plant Biol.* 2004; 55:109-39.
17. Watson AA, O'Callaghan CA. Crystallization and X-ray diffraction analysis of human CLEC-2. *Acta Crystallogr Sect F Struct Biol Cryst Commun.* 2005 Dec 1;61(Pt 12):1094-6.
18. Derewenda ZS. The use of recombinant methods and molecular engineering in protein crystallization. *Methods.* 2004 Nov;34(3):354-63.
19. Putnam CD, Hammel M, Hura GL, Tainer JA. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys.* 2007 Aug;40(3):191-285.
20. Koch MH, Vachette P, Svergun DI. Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q Rev Biophys.* 2003 May;36(2):147-227.
21. Gatzeva-Topalova PZ, May AP, Sousa MC. Crystal structure and mechanism of the *Escherichia coli* ArnA (PmrI) transformylase domain. An enzyme for lipid A modification with 4-amino-4-deoxy-L-arabinose and polymyxin resistance. *Biochemistry.* 2005 Apr 12;44(14):5328-38.
22. Gatzeva-Topalova PZ, May AP, Sousa MC. Crystal structure of *Escherichia coli* ArnA (PmrI) decarboxylase domain. A key enzyme for lipid A modification with 4-amino-4-deoxy-L-arabinose and polymyxin resistance. *Biochemistry.* 2004 Oct 26;43(42):13370-9.
23. Harper AD, Bar-Peled M. Biosynthesis of UDP-xylose. Cloning and characterization of a novel *Arabidopsis* gene family, UXS, encoding soluble and putative membrane-bound UDP-glucuronic acid decarboxylase isoforms. *Plant Physiol.* 2002 Dec;130(4):2188-98.
24. Liu Y, Thoden JB, Kim J, Berger E, Gulick AM, Ruzicka FJ, Holden HM, Frey PA. Mechanistic roles of tyrosine 149 and serine 124 in UDP-galactose 4-epimerase from *Escherichia coli*. *Biochemistry.* 1997 Sep 2;36(35):10675-84.

25. Leloir LF. The enzymatic transformation of uridine diphosphate glucose into a galactose derivative. Arch Biochem Biophys. 1951 Sep;33(2):186-90.

26. Choi SH, Mansoorabadi SO, Liu YN, Chien TC, Liu HW. Analysis of UDP-D-apiose/UDP-D-xylose synthase-catalyzed conversion of UDP-D-apiose phosphonate to UDP-D-xylose phosphonate: implications for a retroaldol-aldol mechanism. J Am Chem Soc. 2012 Aug 29;134(34):13946-9.

27. Choi SH, Ruszczycky MW, Zhang H, Liu HW. A fluoro analogue of UDP- α -D-glucuronic acid is an inhibitor of UDP- α -D-apiose/UDP- α -D-xylose synthase. Chem Commun (Camb). 2011 Sep 28;47(36):10130-2.

28. Eixelsberger T, Horvat D, Gutmann A, Weber H, & Nidetzky B. Isotope Probing of the UDP-Apiose/UDP-Xylose Synthase Reaction: Evidence of a Mechanism via a Coupled Oxidation and Aldol Cleavage. Angewandte Chemie (International Ed. in English), 56(9), 2503–2507. <http://doi.org/10.1002/anie.201609288>

- A review on apiose: Pičmanová M, Møller BL. Apiose: one of nature's witty games. Glycobiology. 2016 May;26(5):430-42.

- RG-II in different plants: Matsunaga T, Ishii T, Matsumoto S, Higuchi M, Darvill A, Albersheim P, O'Neill MA. Occurrence of the primary cell wall polysaccharide rhamnogalacturonan II in pteridophytes, lycophytes, and bryophytes. Implications for the evolution of vascular plants. Plant Physiol. 2004 Jan;134(1):339-51. Epub 2003 Dec 11.

- AXS depletion and RG-II deficiency: Ahn JW, Verma R, Kim M, Lee JY, Kim YK, Bang JW, Reiter WD, Pai HS. Depletion of UDP-D-apiose/UDP-D-xylose synthases results in rhamnogalacturonan-II deficiency, cell wall thickening, and cell death in higher plants. J Biol Chem. 2006 May 12;281(19):13708-16

- AXS in *Ornithogalum caudatum*: Yin S, Kong JQ. Transcriptome-guided gene isolation and functional characterization of UDP-xylose synthase and UDP-D-apiose/UDP-D-xylose synthase families from *Ornithogalum caudatum* Ait. Plant Cell Rep. 2016 Nov;35(11):2403-242

Chapter III:**PockeMO: small evolutionary differences in protein folding drastically change substrate acceptance**

The aim of this work was to provide a structural description of the enzyme PockeMO, found in a fungus and predicted to be a Baeyer-Villiger monooxygenase (BVMO). Structural determination and analysis of the active site was coupled to biochemical characterization of the enzyme. The work presented in this chapter was started when I was still a Master student and was carried on during my PhD. It consisted of crystallization trials, structure elucidation, and of modelling substrates into the enzyme PockeMO. This project was carried out in collaboration with the group of Prof. Marco Fraaije, Rijksuniversiteit Groningen, where the biochemical characterization of the enzyme was performed.

PockeMO is a NADPH-dependent monomeric flavoenzyme of 75 kDa. Its sequence comes from the fungus *Thermothelomyces thermophila*. The enzyme was discovered by Dr. Hanna M. Dudek by spotting the consensus sequence FXGXXXHXXXW(P/D), typical of BVMOs [1], in the fungal genome. The Baeyer-Villiger reaction is well-known in chemistry, consisting in the conversion, via oxidation, of ketones and cyclic ketones to esters and lactones, respectively. The reaction can be performed by means of traditional chemistry, but the process involves usage of peracids in acidic environment and requires high temperatures [2], making it potentially dangerous and definitely not environment-friendly. Several organisms were found capable of performing the Baeyer-Villiger reaction without using peroxyacids and generally with high enantio-selectivity. For this reason, the option of biocatalysis is quite convenient and much effort has been put in characterizing different members of the BVMO class [3], [4]. Of course, the main goal in this research field is to acquire enough knowledge to enhance the physicochemical features of the enzymes, generating better performing biocatalysts. This is usually done by making mutants, which display higher resistance to organic solvents, or that tolerate higher temperatures than the wild type enzyme. Numerous studies have been published on practical

implementation, upscale and yield augmentation for productive processes based on BVMOs [5], [6].

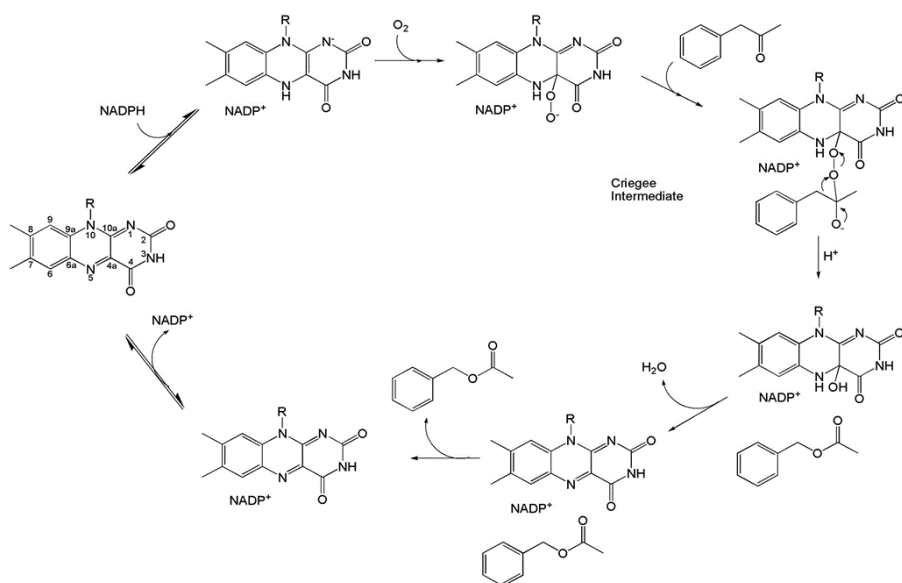


Figure III-1: reaction cycle of a typical Baeyer-Villiger monooxygenase. Starting from left, the intervention of NADPH is an essential step to reduce FAD, making it ready to delocalize electrons and accept molecular oxygen. After formation of the peroxyflavin intermediate, the substrate forms a complex with FAD, known as Criegee intermediate. Thanks to nucleophilic substitution, the final product of the Baeyer-Villiger oxidation is formed and released, while the newly formed hydroxyflavin is decomposed into oxidized flavin and H₂O. Please note that the NADP⁺ coenzyme remains bound during the entire reaction cycle being only released during the final step. Image from [8].

Both chemical and enzymatic reactions pass through the formation of an instable Criegee intermediate [7]. This step is necessary for the introduction of an oxygen atom onto which an alkyl group is transferred; the result is a molecule where an oxygen atom has been introduced between two carbons (Fig. III-1). What makes BVMOs different from other more common monooxygenases is the two intermediates step from the peroxyflavin to the Criegee intermediate. During the BVMO-mediated conversion of ketones, while the flavin is retained in the enzyme, NADPH must be constantly

provided, as it is consumed and released in its oxidized form. Molecular oxygen is also consumed while H₂O is generated.

The article presents a complete structural and biochemical description of PockeMO, which is demonstrated to act on a large pattern of substrates by letting them into the catalytic pocket with a different mechanism with respect to other previously described monooxygenases. Moreover, PockeMO is a very stable enzyme, which may serve as a starting point for possible industrial implementations.

Below the text of the article we published in the Journal of American Chemistry Society regarding the work on PockeMO, together with the supporting material for sake of completeness.

References

1. Fraaije MW¹, Kamerbeek NM, van Berkel WJ, Janssen DB. Identification of a Baeyer-Villiger monooxygenase sequence motif. *FEBS Lett.* 2002 May 8;518(1-3):43-7.
2. Yang N, Su Z, Feng X, Hu C. Theoretical studies on the asymmetric Baeyer-Villiger oxidation reaction of 4-phenylcyclohexanone with m-chloroperoxybenzoic acid catalyzed by chiral scandium(III)-N,N'-dioxide complexes. *Chemistry.* 2015 May 4;21(19):7264-77.
3. Fraaije MW, Wu J, Heuts DP, van Hellemond EW, Spelberg JH, Janssen DB. Discovery of a thermostable Baeyer-Villiger monooxygenase by genome mining. *Appl Microbiol Biotechnol.* 2005 Jan;66(4):393-400.
4. Franceschini S, van Beek HL, Pennetta A, Martinoli C, Fraaije MW, Mattevi A. Exploring the structural basis of substrate preferences in Baeyer-Villiger monooxygenases: insight from steroid monooxygenase. *J Biol Chem.* 2012 Jun 29;287(27):22626-34.
5. Milker S, Fink MJ, Rudroff F, Mihovilovic MD. Non-hazardous biocatalytic oxidation in Nylon-9 monomer synthesis on a 40 g scale with efficient downstream processing. *Biotechnol Bioeng.* 2017 Aug;114(8):1670-1678.

6. Scherkus C, Schmidt S, Bornscheuer UT, Gröger H, Kara S, Liese A. Kinetic insights into ϵ -caprolactone synthesis: Improvement of an enzymatic cascade reaction. *Biotechnol Bioeng*. 2017 Jun;114(6):1215-1221.
7. Yachnin BJ, Sprules T, McEvoy MB, Lau PC, Berghuis AM. The substrate-bound crystal structure of a Baeyer-Villiger monooxygenase exhibits a Criegee-like conformation. *J Am Chem Soc*. 2012 May 9;134(18):7788-95.
8. Malito E, Alfieri A, Fraaije MW, Mattevi A. Crystal structure of a Baeyer-Villiger monooxygenase. *PNAS USA*. 2004 Sep 7;101(36):13157-62.

Polycyclic Ketone Monooxygenase from the Thermophilic Fungus *Thermothelomyces thermophila*: A Structurally Distinct Biocatalyst for Bulky Substrates

Maximilian J. L. J. Fürst,^{||,†} Simone Savino,^{||,‡} Hanna M. Dudek,[†] J. Rúben Gómez Castellanos,[‡] Cora Gutiérrez de Souza,[§] Stefano Rovida,[‡] Marco W. Fraaije,^{*,†} and Andrea Mattevi^{*,‡}

[†]Molecular Enzymology Group, University of Groningen, Nijenborgh 4, 9747AG, Groningen, The Netherlands

[‡]Department of Biology and Biotechnology, University of Pavia, Via Ferrata 1, 27100, Pavia, Italy

[§]Stratingh Institute for Chemistry, University of Groningen, Nijenborgh 4, 9747AG, Groningen, The Netherlands

Supporting Information

ABSTRACT: Regio- and stereoselective Baeyer–Villiger oxidations are difficult to achieve by classical chemical means, particularly when large, functionalized molecules are to be converted. Biocatalysis using flavin-containing Baeyer–Villiger monooxygenases (BVMOs) is a well-established tool to address these challenges, but known BVMOs have shortcomings in either stability or substrate selectivity. We characterized a novel BVMO from the thermophilic fungus *Thermothelomyces thermophila*, determined its three-dimensional structure, and demonstrated its use as a promising biocatalyst. This fungal enzyme displays excellent enantioselectivity, acts on various ketones, and is particularly active on polycyclic molecules. Most notably we observed that the enzyme can perform oxidations on both the A and D ring when converting steroids. These functional properties can be linked to unique structural features, which identify enzymes acting on bulky substrates as a distinct subgroup of the BVMO class.

The Baeyer–Villiger oxidation is a widely used reaction in organic synthesis to break C–C bonds through the insertion of a single oxygen atom adjacent to a carbonyl moiety, yielding esters or lactones from ketones. Since the inserted oxygen is usually derived from a peroxide or peracid, industrial scale processes raise environmental and safety concerns.¹ Baeyer–Villiger monooxygenases (BVMOs) represent an attractive biocatalytic alternative, which uses molecular oxygen as the oxidant and often displays superior chemo-, regio-, and enantioselectivity.² The prototype BVMO, cyclohexanone monooxygenase from *Acinetobacter calcoaceticus* NCIMB 9871 (CHMO), was shown to oxidize hundreds of small aromatic, linear, and cyclic ketones with high enantioselectivity.³ Furthermore, one major hurdle for industrial application of BVMOs, the requirement of stoichiometric amounts of NADPH as reducing cofactor, has been successfully addressed.⁴ What remains problematic, however, is the poor solvent tolerance and thermostability of most BVMOs. Even though close to a hundred BVMOs have been recombinantly produced and characterized so far, only two enzymes stand out by their stability: phenylacetone monooxygenase from *Thermobifida fusca*⁵ (PAMO) and the

recently discovered CHMO from *Thermocristum municipale*.⁶ Yet, these biocatalysts have a restricted substrate scope, being mainly active on small aromatic or cyclic aliphatic ketones. From an application point of view, it would be attractive to have access to a robust BVMO acting on structurally demanding compounds, as it is the case for many pharmaceuticals.

In an effort to exploit the genome of thermophiles for the discovery of new enzymes with interesting catalytic properties, we considered *Thermothelomyces thermophila* ATCC 42464. This genome-sequenced, thermophilic fungus efficiently degrades cellulose and other compounds derived from plant biomasses and is becoming an attractive organism for large-scale protein production.⁷ We identified a gene encoding for a protein containing the typical BVMO fingerprint.⁸ The gene is located in a cluster of hypothetical genes with unknown functions. A BLAST search in the UniProt database revealed ten characterized proteins with significant sequence identities (37–44%) to the *T. thermophila* protein. Four are monooxygenases involved in the biosynthesis of meroterpenoids in *Aspergillus*.^{9–11} Three are BVMOs from *Streptomyces* taking part in the processing of the antibiotic pentalenolactone.^{12,13} One is a BVMO forming the carbonate functionality in the mycotoxin cytochalasin.¹⁴ The remaining two are the biocatalytically characterized BVMOs from *Dietzia sp.*¹⁵ and *S. coelicolor*.¹⁶ A correlation between sequence similarity and substrate scope has been proposed for BVMOs,¹⁷ and indeed, a feature shared by these enzymes is the activity on bulky, polycyclic ketones. Consistently, a phylogenetic analysis assigns the *T. thermophila* protein to a cluster containing versatile enzymes active on large substrates, such as cyclododecanone monooxygenase from *Rhodococcus ruber*¹⁸ (CDMO) and cyclopentadecanone monooxygenase from *Pseudomonas sp. HI-70*¹⁹ (CPDMO) (Figure S1). Collectively, these data led us to hypothesize that the *T. thermophila* protein is a BVMO that combines two attractive properties: (i) thermostability and (ii) activity on relatively large molecules.

An *Escherichia coli* codon-optimized copy of the gene was cloned in two translational fusion variants: a SUMO tagged construct and a cofactor-recycling phosphite dehydrogenase (PTDH) fusion.²⁰ Both constructs yielded very high expression

Received: November 28, 2016

Published: December 23, 2016

levels in *E. coli*, and the proteins were easily obtained in high purity and yield using standard metal affinity chromatography by exploiting the N-terminal 6xHis tag (Figure S2A). The purified yellow protein exhibited a typical flavoprotein spectrum with absorbance peaks at 371 and 454 nm (Figure S2B). SDS treatment revealed noncovalently bound FAD as a cofactor, and the extinction coefficient ϵ_{454} of the holoenzyme was estimated to be $10.2 \text{ mM}^{-1} \text{ cm}^{-1}$. Upon mixing the enzyme with $100 \mu\text{M}$ NADPH and the BVMO model substrate bicyclo[3.2.0]hept-2-en-6-one (13), rapid consumption of NADPH was detected (Figure S3). By monitoring NADPH consumption in the absence of substrate, we determined an uncoupling rate of only 0.02 s^{-1} . We also compared the catalytic behavior of the two recombinant protein variants and determined the activity of the enzyme fused to PTDH and without any tag (obtained by cleaving SUMO). As previously observed, the effect of the dehydrogenase fusion was very modest and, in fact, resulted in a slightly higher activity (Figure S3) and unchanged uncoupling rate. Since the resulting self-sufficient biocatalyst can be used in reactions without constant addition of NADPH, as long as phosphite is added to drive the coenzyme regeneration reaction, all subsequent experiments were performed with the PTDH-fused enzyme. The preliminary analysis was completed by determining basic enzyme characteristics for optimal catalytic conditions. The pH profile highlighted good stability in a range of 6.5–9 with a slight preference for moderately alkaline conditions (Figure 1A).

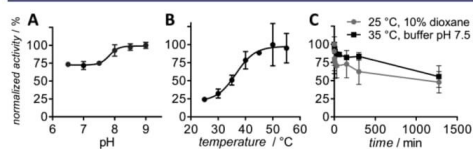


Figure 1. Effects of pH, temperature, and organic solvents on the activity of *T. thermophila* BVMO (PockeMO). $2 \mu\text{M}$ of PTDH-PockeMO and 0.1 mM NADPH were mixed with 13.

We next sought to probe the enzyme's thermostability. We found the activity to be highest at $50 \text{ }^\circ\text{C}$, where it was four times faster than at room temperature (Figure 1B). Using the ThermoFAD method,²¹ we further observed that both native as well as a PTDH-fused enzyme have an apparent melting temperature (T_m) of $47 \text{ }^\circ\text{C}$, the second highest reported for a BVMO after PAMO. This T_m is sustained in a pH range of 7–9 and is increased a further $2 \text{ }^\circ\text{C}$ upon exposing the protein to $100 \mu\text{M}$ NADP⁺, consistent with a tight binding of this dinucleotide cofactor (Figure S4). In parallel, we evaluated the effect of organic solvents which was probed by adding DMSO, 1,4-dioxane, ethanol, acetone, or 1-propanol. The resulting reductions in T_m ranged from $1\text{--}5 \text{ }^\circ\text{C}$ to $2\text{--}8 \text{ }^\circ\text{C}$ with 5% and 10% solvent, respectively (Figure S4). The enzyme seemed to exhibit a better tolerance toward polar solvents, with DMSO having the least and 1-propanol the highest impact on T_m . Because DMSO can act as a substrate, and the effect of the remaining solvents was comparable, we chose dioxane as the preferred cosolvent. We monitored the enzymatic activity as a function of time and observed a satisfactory half-life of $\sim 24 \text{ h}$ when incubating the enzyme in buffer at $35 \text{ }^\circ\text{C}$, or at rt with 10% dioxane (Figure 1C). These experiments indicated that the BVMO features a considerable degree of thermostability and only moderate sensitivity to organic solvents.

We next investigated the enzyme's substrate scope and selectivity (fully listed in Table S1). The closely related CPDMO and CDMO are highly active on large cyclic ketones (Figure S1). Consistently, we found conversion of 6 and 7 (Table 1 and Figure S5). The 15-membered cyclopentadecanone (8)

Table 1. Substrate Scope of PockeMO

Category	Structure
Linear ketones	$1: \text{R}^1 = \text{CH}_3(\text{CH}_2)_3, \text{R}^2 = \text{CH}_3$ $2: \text{R}^1 = \text{Cyclohexane}, \text{R}^2 = \text{H}$ $3: \text{R}^1 = \text{CH}_2\text{-Phe}, \text{R}^2 = \text{H}$
cyclic ketones	$4: n = 1$ $5: n = 3$ $6: n = 6$
Substituted cyclic ketones	$9: \text{R}^1 = \text{Phe}, \text{R}^{2,3,4,5} = \text{H}$ $10: \text{R}^2 = \text{Phe}, \text{R}^{1,2,4,5} = \text{H}$ $11: \text{R}^{2,4,5} = \text{CH}_3, \text{R}^{1,3} = \text{H}$
Bicyclic ketones	$12: (\text{CH}_2)_5\text{CH}_3$ 13 14
Steroids	15 16 17 18

was fully converted, and kinetic parameters were determined: $k_{\text{cat}} = 0.1 \text{ s}^{-1}$ and $K_M = 144 \mu\text{M}$ (Figure S6). Of special interest, we observed efficient conversions of structurally demanding compounds, including several steroids. In particular, we tested stanolone (17), which has only one carbonyl group positioned at C3 of the A-ring and observed complete conversion after 24 h (Figure S5). As a reference, we performed the conversion with CPDMO, for which the 3-keto-4-oxa-4a-homo isomer was reported as the only product.²² The MS spectra of the resulting products with the expected mass shift of +16 Da were the same for both enzymes (Figure S5), and we confirmed the structure by NMR (Figure S7). With pregnenolone (18), a steroid harboring a carbonyl group only at the C17 side chain, we found close to full conversion after 24 h (Figures 2 and S5). We again observed the

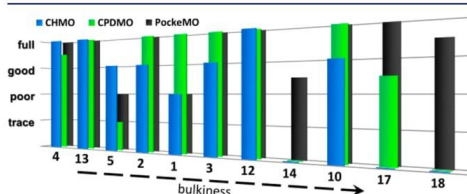


Figure 2. Conversion efficiencies of PockeMO, CPDMO, and CHMO (all PTDH fused) on a mix of substrates. Efficiencies are scaled semiquantitatively according to analysis of GC peaks of substrates and products into full, good or poor conversion, or trace amounts detectable.

+16 Da mass shift, and the product was determined by NMR to be the corresponding acetate (Figure S8). When the enzyme was presented with androstenedione or androstadienedione (15 and 16), steroids with keto groups in both the A and D rings, the enzyme selectively oxidized the D-ring to yield the pharmaceutically relevant testo(lo)lactone (Figure S5). Again, complete conversion was observed after 24 h.

To our knowledge, this is the first report of an isolated BVMO that is able to oxidize the keto functionalities at these different positions. Steroid monooxygenase of *Rhodococcus rhodochrous* (STMO) can only convert steroidal carbonyls positioned on the C17 side chain and not in the rings.²³ A homologue from *Cylindrocarpus radicolica* in addition is able to lactonize the D ring.²⁴ CPDMO accepts some steroids with the keto group on the 3 and 17 position, but is inactive on open chain ketones.²² The *T. thermophila* BVMO represents a biocatalyst combining these abilities (Table 1, red arrows), and we therefore named it *polycyclic ketone monooxygenase* (PockeMO).

We further explored the enzyme's substrate profile using several substrates frequently tested with other BVMOs (Tables 1 and S1). For this purpose, we used a highly efficient substrate screening method, based on the conversion of a mixture of compounds and subsequent separation and analysis via GC-MS. We tested a mixture of 16 linear, aromatic, cyclic, and polycyclic ketones (Figure S5) and could immediately identify 13 conversion products. Of the unsubstituted cyclic ketones, PockeMO converted cyclohexanone to 100%, but the conversion of cyclooctanone was poor in comparison. The monosubstituted cyclic ketones 9, 10, and 12 were converted with yields close to 100%. Approximately half of the disubstituted cyclic ketone 11 was converted, whereas the unsaturated variant of this molecule (isophorone) was not accepted. The aromatic ketone 3 was fully converted, and also the bridged, bicyclic 14 was converted with moderate yield (Table S1, Figure 2). Notably, although being only a poorly accepted substrate, the linear 4-octanone (1) was exclusively converted to the abnormal product propyl pentanoate (Figure S9). These experiments elegantly demonstrated the activity on an array of substrates of varying degree of bulkiness. At the same time, none of the compounds seemed to exhibit inhibitory effects on the enzyme. To put this further into perspective, we performed the conversion of a mix of 15 ketones with purified CHMO, CPDMO, and PockeMO. Figure 2 shows the semiquantitative comparison of the conversion efficiency of the three enzymes and clearly shows PockeMO's excellent all-round performance as well as its particular superiority with larger compounds.

For BVMOs, enantio- and regioselectivity have frequently been studied by analyzing the conversion of racemic 13 into the respective lactones. A recent extensive study on BVMO activity in various fungi revealed diverse selectivities and only a moderate phylogenetic correlation.²⁵ We found that PockeMO (4 μ M enzyme, 2 h) completely converted 10 mM racemic 13 enantiodivergently, yielding two regioisomeric lactones, with *ee* values of 100% and 97%, respectively (Figure S10). Thus, the enzyme exhibited both high enantio- and regioselectivity as well as high activity toward this substrate. Kinetic analysis confirmed this notion as a k_{cat} of 3.3 s⁻¹ was measured ($K_m = 0.4$ mM).

Given these valuable stability and catalytic properties as well as the fact that PockeMO belongs to the structurally uncharacterized CPDMO/CDMO subfamily of BVMOs (Figure S1), it was of interest to determine the enzyme's three-dimensional structure. After removal of the SUMO tag, the protein could be crystallized as a complex with FAD and NADP⁺ and its structure solved at a resolution of 2.0 Å (Table S2). Most of the residues could be modeled, apart from loop 607–614. PockeMO exhibits the typical BVMO domain organization with an FAD-domain (residues 1–229 and 479–655), an NADP-domain (residues 230–314 and 417–478), and a helical domain (residues 315–416) (Figure 3). The closest structural homologues are STMO²⁶ and PAMO (rmsd of 1.91 and 1.75 Å for S23 and S33 equiv of *C α*

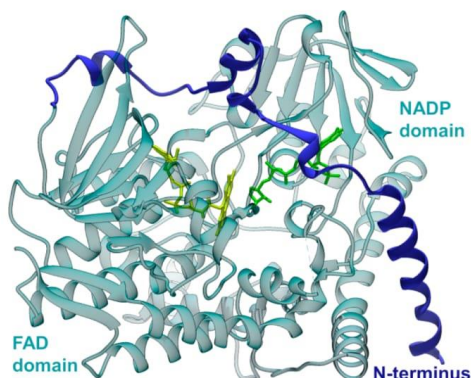


Figure 3. Overall structure of PockeMO (PDB entry 5MQ6). The N-terminal extension (residues 1–73; dark blue) is specific to this BVMO enzyme subclass (Figure S1). FAD is yellow and NADP⁺ green (nicotinamide ring is disordered).

atoms, respectively, with 30% sequence identity to both). The location of the FAD and NADP cofactors is consistent with that of other BVMOs, and the flavin shows no significant deviation from planarity (Figure S11).

Despite this clear sequence and structural homology with other BVMOs and the conservation of the overall fold, PockeMO features some striking structural peculiarities. First, a 70-residue N-terminal extension forms a long α -helix followed by an elongated stretch that wraps around the enzyme in the vicinity of the FAD-domain (Figure 3). The interactions between the 44 N-terminal residues and the rest of the protein are extensive and establish 24 H-bonds, 5 salt bridges, and many hydrophobic contacts. Notably, this elongated N-terminus is a feature of all BVMOs in the phylogenetic group of enzymes acting on complex molecules (Figures S1 and S12A). Second, there is a structural element ranging from residue 316 to 388 whose characteristic conformation reshapes the active site. This region has low sequence similarity to the canonical BVMOs (Figure S12B), in which this portion is structurally conserved (as exemplified by PAMO, Figure 4A). In PockeMO, this stretch forms a characteristic conformation of alternating loops and α -helices. As a result, the substrate-binding site of PockeMO is a long and wide tunnel-like cavity, which starts from the loop 587–624 and ends at the *re*-face of FAD (Figure 4B–C). The flexible loop delimited by residues 606 and 615 serves as a gate-like mechanism, constituting a longer element when compared to the same topological loop in PAMO (residues 495–515). Modeling shows that bulky molecules can readily be accommodated in the PockeMO's substrate tunnel (Figure S13). Thus, the enzyme exhibits characteristic structural features, which are fully consistent with the notion that it belongs to a distinct subgroup of BVMO enzymes (Figure S1).

In summary, we identified a novel, robust, and versatile enzyme performing Baeyer–Villiger oxidations. PockeMO combines thermostability and organic solvent tolerance with a broad substrate profile. Specifically, the enzyme accepts bulky and complex substrates and converts them with high efficiency, making it a promising candidate for application as an industrial biocatalyst. In addition, the elucidated crystal structure not only lays the basis for enzyme engineering but also can be regarded as

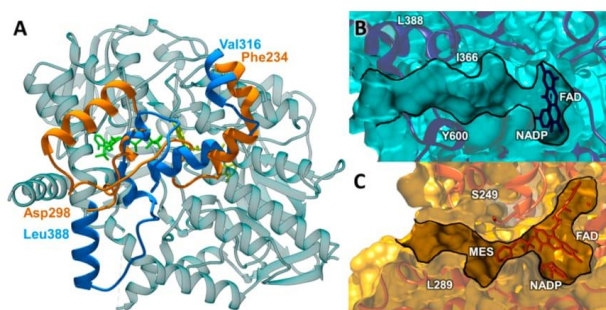


Figure 4. Characteristic structural features of PockeMO. (A) Residues 316–388 (bright blue) cover the active site and adopt a conformation not observed in other BVMO structures as exemplified by the superimposed equivalent residues of PAMO (PDB 2YLT, 234–298; orange; Figure S11). (B) PockeMO structure cut along the substrate entrance to outline the wide tunnel forming the active site. (C) PAMO active site features a narrower channel, ranging from the solvent exposed surface to the active site with MES bound.

a prototype of an evolutionary and structurally distinct group of BVMOs. Careful further analysis should contribute to a better understanding of the still largely unknown mechanism by which substrate selectivity in BVMO is tuned. The structural and sequence features of PockeMO can also be exploited to identify new and diverse BVMOs evolved to process relatively large substrates.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b12246.

Experimental procedures; Figures S1–S13; Tables S1, S2 (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*m.w.fraaije@rug.nl (M.W.F.)

*mattevi@ipvgen.unipv.it (A.M.)

ORCID

Marco W. Fraaije: 0000-0001-6346-5014

Andrea Mattevi: 0000-0002-9523-7128

Author Contributions

^{||}M.J.L.J.F. and S.S. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The research for this work has received funding from the European Union (EU) Project ROBOX (Grant Agreement No. 635734) under EU's Horizon 2020 Programme Research and Innovation actions H2020-LEIT BIO-2014-1.

■ REFERENCES

- (1) Leisch, H.; Morley, K.; Lau, P. C. *Chem. Rev.* **2011**, *111*, 4165.
- (2) Kamerbeek, N. M.; Janssen, D. B.; van Berkel, W. J.; Fraaije, M. W. *Adv. Synth. Catal.* **2003**, *345*, 667.
- (3) Mihovilovic, M. D.; Müller, B.; Stanetty, P. *Eur. J. Org. Chem.* **2002**, *2002*, 3711.
- (4) Torres Pazmino, D. E.; Winkler, M.; Glieder, A.; Fraaije, M. W. *J. Biotechnol.* **2010**, *146*, 9.

- (5) Fraaije, M. W.; Wu, J.; Heuts, D. P.; van Hellemond, E. W.; Spelberg, J.; Janssen, D. B. *Appl. Microbiol. Biotechnol.* **2005**, *66*, 393.
- (6) Romero, E.; Castellanos, J. R. G.; Mattevi, A.; Fraaije, M. W. *Angew. Chem., Int. Ed.* **2016**, *55*, 15852.
- (7) Berka, R. M.; Grigoriev, I. V.; Ottilar, R.; Salamov, A.; Grimwood, J.; Reid, L.; Ishmael, N.; John, T.; Darmond, C.; Moisan, M. C.; et al. *Nat. Biotechnol.* **2011**, *29*, 922.
- (8) Riebel, A.; de Gonzalo, G.; Fraaije, M. W. *J. Mol. Catal. B: Enzym.* **2013**, *88*, 20.
- (9) Matsuda, Y.; Iwabuchi, T.; Wakimoto, T.; Awakawa, T.; Abe, I. *J. Am. Chem. Soc.* **2015**, *137*, 3393.
- (10) Matsuda, Y.; Wakimoto, T.; Mori, T.; Awakawa, T.; Abe, I. *J. Am. Chem. Soc.* **2014**, *136*, 15326.
- (11) Matsuda, Y.; Awakawa, T.; Wakimoto, T.; Abe, I. *J. Am. Chem. Soc.* **2013**, *135*, 10962.
- (12) Seo, M. J.; Zhu, D.; Endo, S.; Ikeda, H.; Cane, D. E. *Biochemistry* **2011**, *50*, 1739.
- (13) Jiang, J.; Tetzlaff, C. N.; Takamatsu, S.; Iwatsuki, M.; Komatsu, M.; Ikeda, H.; Cane, D. E. *Biochemistry* **2009**, *48*, 6431.
- (14) Hu, Y.; Dietrich, D.; Xu, W.; Patel, A.; Thuss, J. A.; Wang, J.; Yin, W. B.; Qiao, K.; Houk, K. N.; Vederas, J. C.; Tang, Y. *Nat. Chem. Biol.* **2014**, *10*, 552.
- (15) Bisagni, S.; Hatti-Kaul, R.; Mamo, G. *AMB Express* **2014**, *4*, 23.
- (16) Park, J.; Kim, D.; Kim, S.; Kim, J.; Bae, K.; Lee, C. *J. Microbiol. Biotechnol.* **2007**, *17*, 1083.
- (17) Szolkowy, C.; Eltis, L. D.; Bruce, N. C.; Grogan, G. *ChemBioChem* **2009**, *10*, 1208.
- (18) Kostichka, K.; Thomas, S. M.; Gibson, K. J.; Nagarajan, V.; Cheng, Q. *J. Bacteriol.* **2001**, *183*, 6478.
- (19) Iwaki, H.; Wang, S.; Grosse, S.; Bergeron, H.; Nagahashi, A.; Lertvorachon, J.; Yang, J.; Konishi, Y.; Hasegawa, Y.; Lau, P. C. *Appl. Environ. Microbiol.* **2006**, *72*, 2707.
- (20) Torres Pazmino, D. E.; Riebel, A.; de Lange, J.; Rudroff, F.; Mihovilovic, M. D.; Fraaije, M. W. *ChemBioChem* **2009**, *10*, 2595.
- (21) Forneris, F.; Orru, R.; Bonivento, D.; Chiarelli, L. R.; Mattevi, A. *FEBS J.* **2009**, *276*, 2833.
- (22) Beneventi, E.; Ottolina, G.; Carrea, G.; Panzeri, W.; Fronza, G.; Lau, P. C. *J. Mol. Catal. B: Enzym.* **2009**, *58*, 164.
- (23) Leipold, F.; Rudroff, F.; Mihovilovic, M. D.; Bornscheuer, U. T. *Tetrahedron: Asymmetry* **2013**, *24*, 1620.
- (24) Itagaki, E. *J. Biochem.* **1986**, *99*, 825.
- (25) Butinar, L.; Mohorčić, M.; Deyris, V.; Duquesne, K.; Iacazio, G.; Claeys-Bruno, M.; Friedrich, J.; Alphand, V. *Phytochemistry* **2015**, *117*, 144.
- (26) Morii, S.; Sawamoto, S.; Yamauchi, Y.; Miyamoto, M.; Iwami, M.; Itagaki, E. *J. Biochem.* **1999**, *126*, 624.

1 SUPPORTING FIGURES

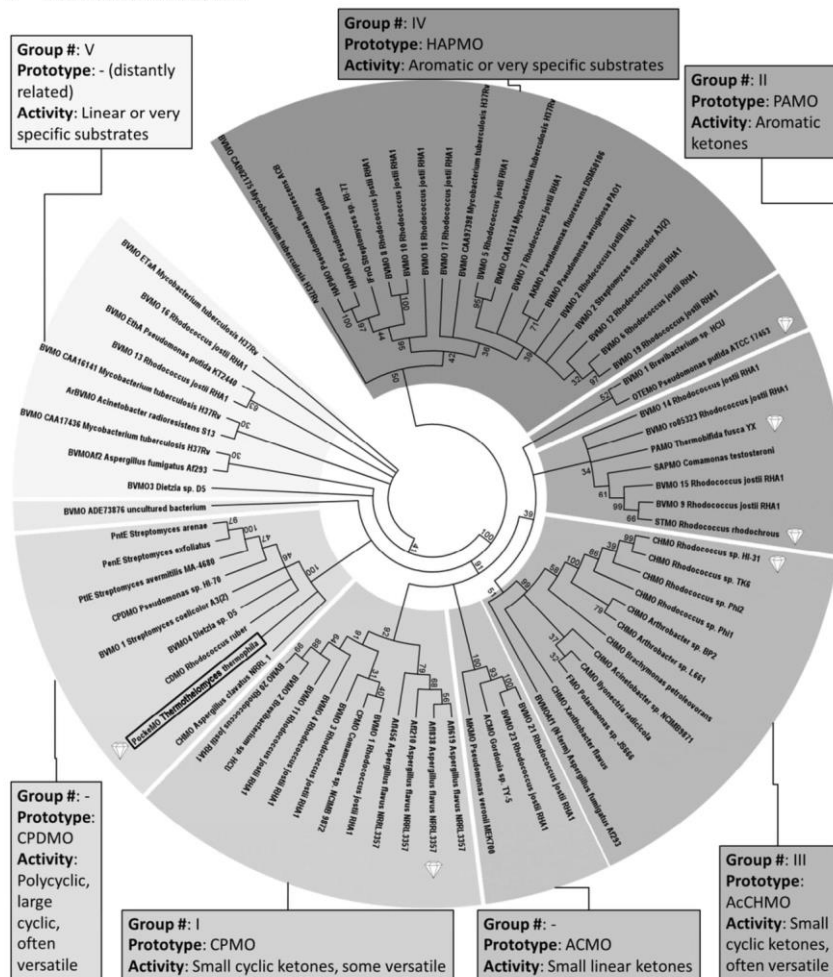


Figure S1. Phylogenetic tree with recombinantly expressed BVMOs. A clustalW-calculated protein sequence alignment of 79 type I BVMOs was used to generate the tree using the Maximum-Likelihood method. Robustness was tested with 500 bootstraps (values at the nodes) and the cut-off was 30%. Phylogenetic groups are annotated from literature and in accordance with previous classifications. Group numbers are according to the “Grogan classification”^[1] and prototypes represent the best characterized representative of a group. Crystal structures are indicated by a crystal symbol. The accession numbers of all sequences are as follows:

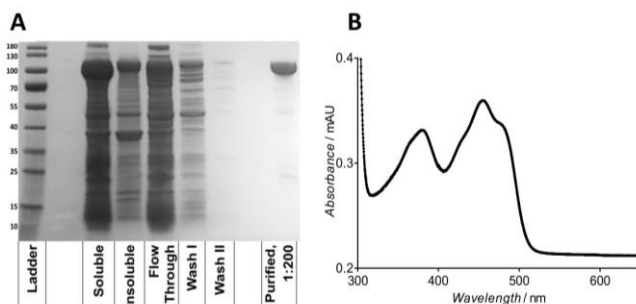


Figure S2. SDS-PAGE of enzyme purification (A) and spectrum (B) of purified *Thermotheomyces thermophila* BVMO (PockeMO) fused to PTDH. The purification procedure is described in more detail in section 3.4. The soluble fraction is obtained after separation of cell debris (insoluble fraction) from the soluble fraction after disruption of cells via sonication. This cell-free extract was subjected to a Nickel column which was washed with buffer (wash I) and buffer containing 5 mM imidazole (wash II). The eluted protein was desalted to yield pure protein. Yields ranged from 120-180 mg pure protein per liter of culture, depending on the prepared batch.

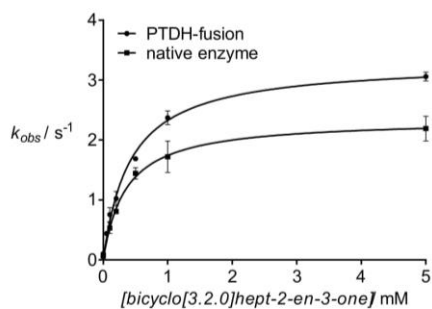


Figure S3. Kinetic profile for *Thermotheomyces thermophila* BVMO (PockeMO) fused to phosphite dehydrogenase or after cleavage of the SUMO tag. NADPH consumption rates were measured in air-saturated 50 mM Tris-HCl buffer at pH 7.5 and 25 °C (enzyme 2 μ M; NADPH, 100 μ M). Plots were fit to the Michaelis-Menten equation.

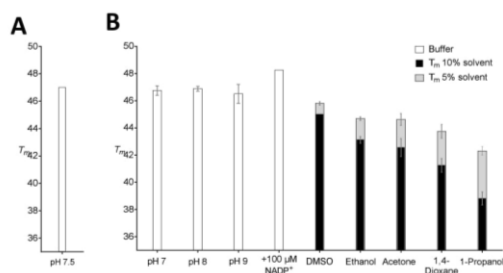
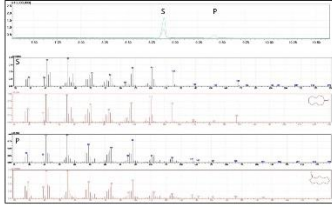
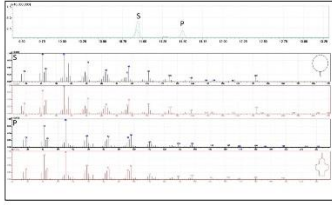


Figure S4. Apparent melting temperature (T_m) of *Thermotheomyces thermophila* BVMO (PockeMO) native enzyme (A) and the PTDH-fusion enzyme (B) in different solvents and varying pH values.

Conversion of cycloandrostane:

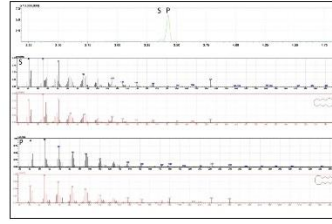


Conversion of cycloandrostane:

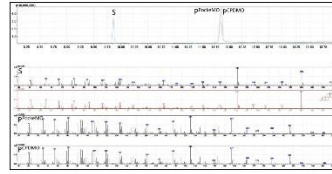


56

Conversion of cycloandrostane:

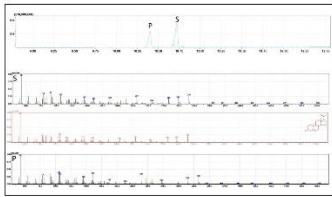


Conversion of androstane:

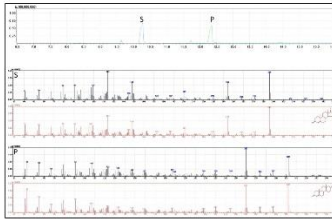


57

Conversion of pregnenolone:

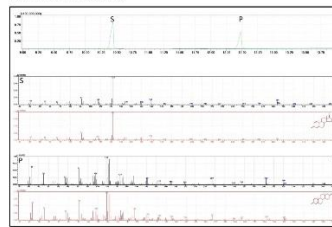


Conversion of androstenedione:



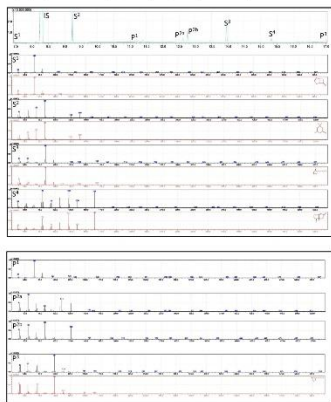
58

Conversion of androstenedione:



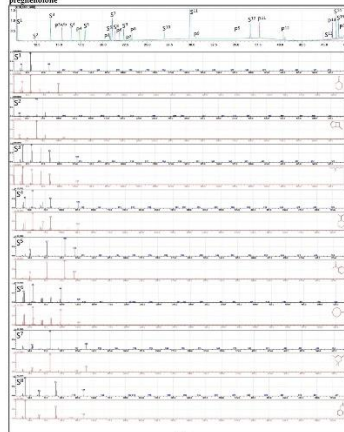
59

Conversion of mixed bicyclo[2.2.0]hept-2-en-6-one, 3,3-dimethylcyclohexanone, bicyclopentanon, (S)-(+)-2,5,7-trimethylheptan-2-one, 4-methyl-2-cyclohexenone

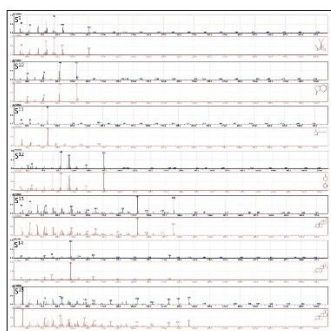


514

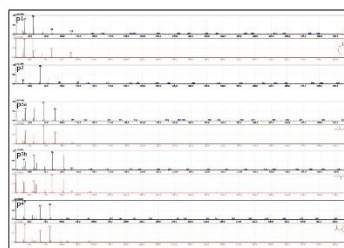
Conversion of mixed cyclohexanone, bicyclo[2.2.0]hept-2-en-6-one, 4-octanone, cyclohexylmethylketone, acetophenone, cyclooctanone, isophorone, phenylacetone, (1S)-camphor, 2-indanone, 2-bicyclopentanon, 4-phenylcyclohexanone, stamolon, androstenedione, progesterone



515



516



517

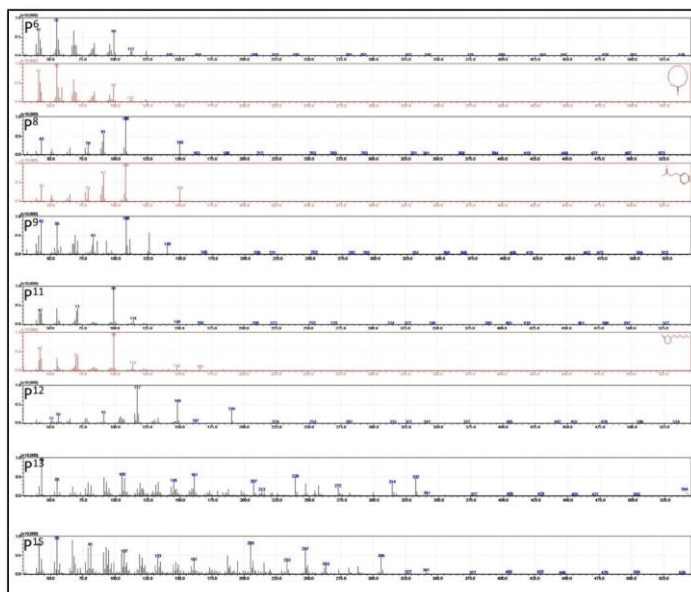


Figure S5. GC-MS results of conversions with PockeMO. First shown is the GC-chromatogram as an overlay of the sample containing the substrate (S)/the substrate mix (S'-S'') incubated with enzyme (green line), and the same sample without enzyme (control, blue line). In the enzyme containing sample, the resulting product peak is abbreviated P or P'-Pⁿ for the conversion mix. The MS fragmentation spectra are shown below: in black the actual measurement, and in red the spectrum in the library (NIST11) that lead to identification, together with the chemical structure. The similarity of all spectra shown with that of the library is > 90% in all cases. Some compounds were not present in the library (closest hit similarity < 80%), which is indicated with a gap after the measurement spectrum. Compounds were then identified as the Baeyer-Villiger reaction product of the substrate due to the GC peak shift and the appearance of an MS parent peak with a 16 Da higher mass, corresponding to an oxygen insertion. For the last mix of 15 substrates, PockeMO (green line), CHMO (pink line) and CPDMO (brown line) were compared. With the exception of peak P³ (4-octanone), all enzymes converted the substrates to the same regioisomer (when applicable), in these cases, only one representative MS spectrum is shown. Bicyclo[3.2.0]hept-2-en-6-one was included in each mix as known control substrate to indicate potential inhibiting effects.

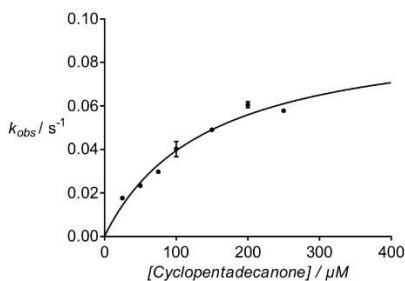


Figure S6. Kinetic profile of PockeMO with cyclopentadecanone. NADPH consumption rates were measured in air-saturated 50 mM Tris-HCl at pH 7.5 and 25 °C (enzyme 2 μM; NADPH, 100 μM). Plots were fit to the Michaelis-Menten equation. Substrate was solubilized using 5% dioxane. Under these conditions, 250 μM was the solubility limit of this compound, therefore no higher substrate concentrations could be measured

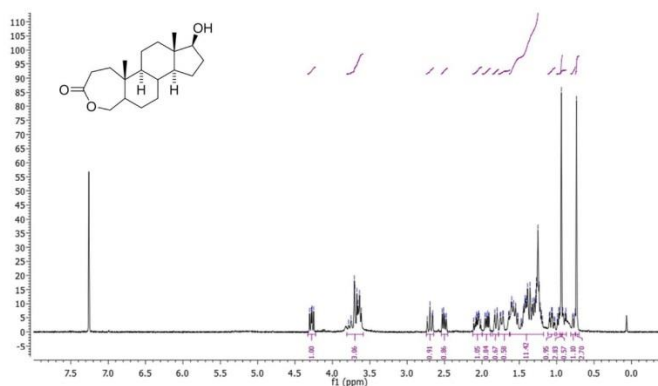


Figure S7. ^1H NMR of stanolone conversion product with PockeMO. ^1H NMR (400 MHz, Chloroform- d) δ 4.28 (dd, $J = 13.1, 8.8$ Hz, 1H), 3.81 – 3.59 (m, 3H), 2.70 (t, $J = 14.1$ Hz, 1H), 2.50 (dd, $J = 14.6, 7.3$ Hz, 1H), 2.12 – 2.00 (m, 1H), 1.93 (dd, $J = 14.4, 7.5$ Hz, 1H), 1.81 (d, $J = 12.6$ Hz, 1H), 1.74 (dd, $J = 13.0, 3.3$ Hz, 1H), 1.63 – 1.18 (m, 1H), 1.06 (td, $J = 12.9, 4.2$ Hz, 1H), 0.94 (s, 3H), 1.00 – 0.87 (m, 1H), 0.81 – 0.75 (m, 1H), 0.74 (s, 3H). The NMR spectrum is in agreement with published data for the depicted 3-keto-4-oxa-4a-homo lactone.^[5]

S19

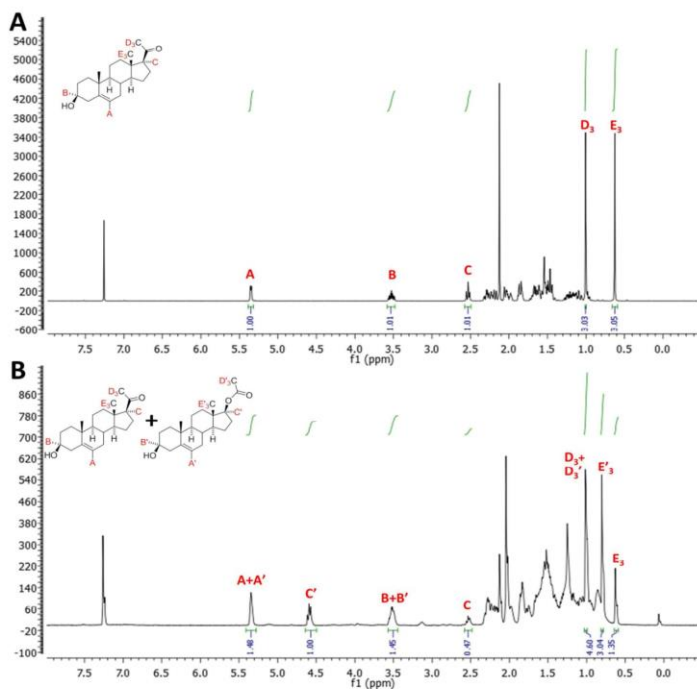


Figure S8. ¹H NMR of pregnenolone and the conversion mix with PockeMO.

(A) ¹H NMR of pregnenolone pure substrate as reference. ¹H NMR (400 MHz, Chloroform-*d*) δ 5.35 (m, *J* = 5.4 Hz, 1H), 3.53 (m, *J* = 11.1 Hz, 1H), 2.53 (t, *J* = 8.9 Hz, 1H), 2.36 – 1.93 (m, 8H), 1.91 – 1.80 (m, 2H), 1.74 – 1.39 (m, 8H), 1.30 – 1.02 (m, 3H), 1.00 (s, 3H), 0.63 (s, 3H).

(B) ¹H-NMR of the reaction crude, showing a mix of the reaction product and substrate (67.6% conversion according to the peak integration). A significant shift of the peak corresponding to proton C of the substrate (2.53 ppm, 1H) to C' (4.59 ppm, 1H) is observed due to the addition of the vicinal oxygen. Taking the integration of peak C' as reference for one proton of the product and C for one proton of the substrate, the integration of the peaks at 5.35 ppm and at 3.52 ppm correspond to the sum of the contribution of product and substrate to these peaks. These two protons A (5.35 ppm, 1H) and B (3.52 ppm, 1H) are too distant from the new oxygen to show a detectable change in shielding, so that the protons corresponding to product and substrate overlap. Moreover, a shift of proton E (0.63 ppm, 3H) to E' (0.80 ppm, 3H) is observed.

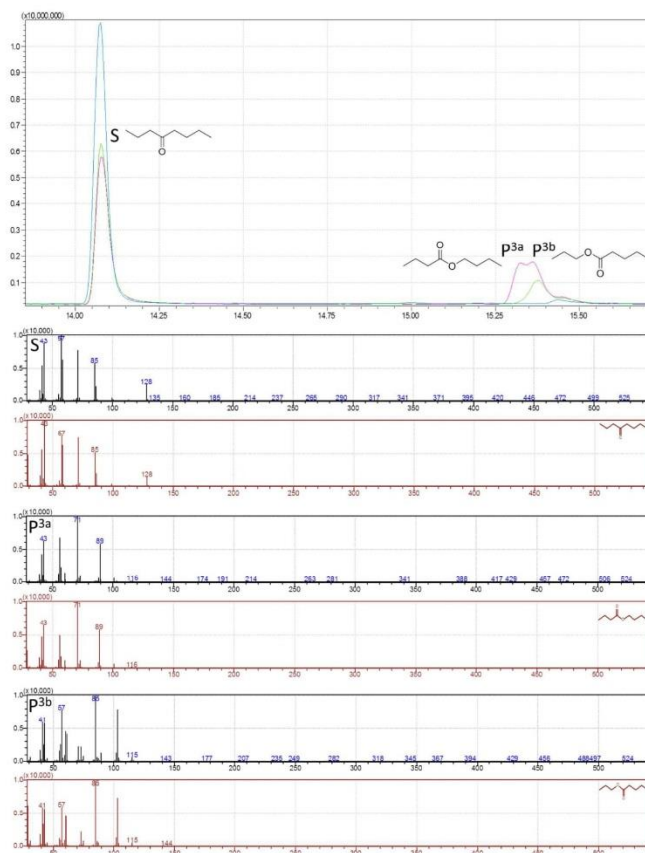


Figure 59. GC-MS results of conversion of 4-octanone with PockeMO. Top graph: GC-chromatogram as an overlay of the sample containing the substrate (S)/the substrate without enzyme (control, blue line). The samples containing enzyme correspond to PockeMO (green) and AcCHMO (pink) and the resulting product peaks are abbreviated P. The sample incubated with AcCHMO yields two peaks that partially overlay (P^{3a} and P^{3b}), while the PockeMO sample contains only one peak, product P^{3b}. The MS fragmentation spectra are shown below: in black the actual measurement, and in red the spectrum in the library (NIST11) that lead to identification, together with the chemical structure. The similarity of all spectra shown with that of the library is > 95% and clearly can distinguish between the normal product P^{3a} and the abnormal product P^{3b}, with the two of them leading to entirely different mass fragmentation spectra.

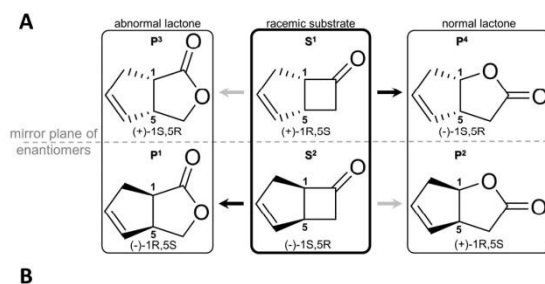


Figure S10. Stereoselectivity of bicyclo[3.2.0]hept-2-en-6-one (BCH) conversion by *Thermothelomyces thermophila* BVMO (PockeMO). A: Oxidation of racemic BCH yields either all the four possible isomers (grey arrows) or two regioisomers (black arrows). B: GC-Chromatogram of chiral GC of racemic BCH incubated with PAMO or PockeMO or without enzyme. The retention times of the two isomeric ketones are 9.2 min (S^1 ; 1S,5R isomer) and 9.4 min (S^2 ; 1R,5S isomer). The order for the lactones is: abnormal lactone 1R,5S: 17.1 min (P^1); normal lactone 1R,5S: 17.6 min (P^2); abnormal lactone 1S,5R: 18.0 min (P^3); normal lactone 1S,5R: 18.2 min (P^4). PAMO yields the four possible lactones, with a preference for the 1R, 5S ketone, while PockeMO converts both ketones fast to the abnormal lactone 1R, 5S and the normal lactone 1S, 5R with enantiomeric excesses of 97% and 100%, respectively.

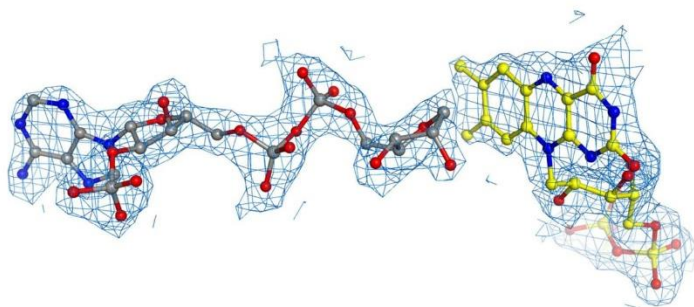


Figure S11. Quality of the crystallographic data. Electron density of NADP⁺ (grey carbons; the nicotinamide ring is disordered and not included in the model) and FAD (yellow carbons) to exemplify the quality of the X-ray data (2.0 Å resolution). The weighted $2F_o - F_c$ map is contoured at 1.4 σ level.

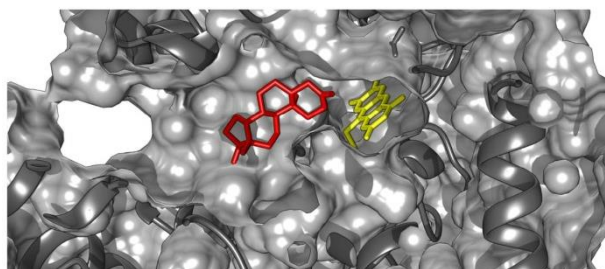




















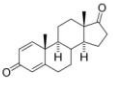
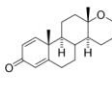
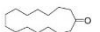
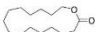
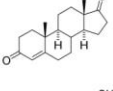
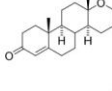

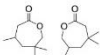
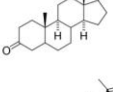
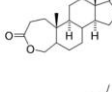

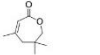
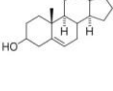
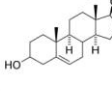

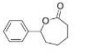





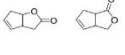


Figure S13. Modeling of stanolone (one of the bulkiest PockeMO substrates; 17 in Table 1 of the main text) bound to the PockeMO active site. The protein surface is in grey, the FAD in yellow, and the modeled ligand in red. Modeling was performed with the program Vina (<http://vina.scripps.edu/>).

2 SUPPORTING TABLES

Table S1. Compounds analyzed for conversion with PockeMO

Structure	Product(s) ^[a]	Conversion ^[b]			
		+			
		+++			++ ^[c]
		+++			-
		+			+++
		++			-
		++			+++
		+++			+++
		++ ^[c]			+++ ^[e]
		-			++ ^[c]
		+++ ^[c]			
		+++ ^[c]			
		(+++)			
		+++ ^[d]			

[a] Conversion products are shown as determined for PockeMO. If no conversion was observed, the theoretical product(s) are shown.

[b] Degree of conversion (Conv.) was determined semi-quantitatively by analysis of the GC peaks for the substrate and the product and categorized as 100%, +++, >50%, ++, 1-50%, +, or 0%, -. Parentheses indicate that no product peak was identifiable and conversion was only inferred from the decrement of the substrate peak.

[c] The product of this compound was not in the MS spectrum library. The MS spectra showed the parent peak with the +16 Da mass shift. The structure depicts the expected BV product.

[d] The identity of the products was confirmed via chiral GC and by comparison with the reference catalyst PAMO.

[e] The identity of the product was confirmed by comparison with the reference catalyst CPDMO and NMR.

3 SUPPORTING MATERIAL AND METHODS

3.1 General methods

All chemical reagents were purchased from Sigma-Aldrich or TCI Europe, unless otherwise stated. Oligonucleotide primers were synthesized by Sigma-Aldrich or Eurofins. DNA sequencing was performed by GATC (Konstanz, Germany).

3.2 Phylogenetic analysis

Phylogenetic analysis with 79 BVMO sequences was performed using MEGA 7.0^[3] to generate the alignment using the ClustalW algorithm with the BLOSUM protein weight matrix. Based on this alignment, a phylogenetic tree was generated using the Maximum-Likelihood (ML) method using these parameters: model: Jones-Taylor-Thornton; uniform rates, heuristic model: nearest-neighbor-interchange. The robustness was then tested with 500 bootstraps and the cut-off for displaying a common node in the final tree was 30%.

3.3 Cloning

A synthetic gene (GenScript - New Jersey, USA) was cloned in a pET-SUMO vector. For cloning into the pCRE vector, the gene was amplified via PCR with 0.2 μ M forward primer GACTCGAGATCTGCTGCTGGTATGGCACCGTCTG and reverse primer GCGTTTCACTTCTGAGTTCG using the PfuUltra II Hotstart PCR Master Mix (Agilent Technologies) with 50 ng pET-SUMO-pockeMO vector DNA. The thermocycler conditions were 92 °C - 2' / [92 °C - 30" / 55 °C - 30" / 72 °C - 4'] x 30 / 72 °C - 10'. The PCR product and the vector were digested with XhoI and HindIII (New England Biolabs) overnight. The vector was then digested with alkaline phosphatase (Roche) overnight. The PCR product was purified and the backbone of the digested vector was cut from an agarose gel and also purified. The two fragments were ligated with T4 ligase (Promega) for 6 h and 5 μ l of the ligation were used to transform 100 μ l of competent *E. coli* NEB 10 β cells. Plasmid was prepared using a QIAprep Spin Miniprep Kit. In-frame insertion and sequence integrity was confirmed by sequencing (GATC-Biotech, Germany). Unless otherwise indicated, all enzymes were used according to manufacturer's instructions. In-frame insertion and sequence integrity was confirmed by sequencing (GATC-Biotech, Germany).

3.4 Protein expression and purification

PockeMO

A baffled Erlenmeyer flask (200 ml - 5 L) was filled up to 10% of the volume with TB medium and inoculated with 5 ml of an overnight culture of *E. coli*. For expression of the SUMO fusion, BL21 (DE3) transformed with pET-His₆-SUMO-pockeMO was used and the medium was supplemented with 50 μ g/ml kanamycin. Cells were grown at 37 °C until an OD₆₀₀ of ~1, and then expression was induced by addition of isopropyl β -D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. Cells were shifted to 30 °C and allowed to grow 6-12 hours before harvest. For expression of the phosphite dehydrogenase fusion, chemocompetent *E. coli* NEB 10 β cells transformed with pBAD-His₆-PTDH-Linker-pockeMO (pCRE-pockeMO) were used to inoculate the main culture. 0.02% L-arabinose and 50 μ g/ml ampicillin were added and the flask was incubated at 24 °C for 36 h with shaking.

Cells were harvested by centrifugation (6,000xg for 15 min at 4°C, JA-10.5 rotor, Beckman Coulter) and frozen at -20 °C for later use. Cells were then disrupted by sonication and subsequently subject to centrifugation (15,000xg for 45 min at 4°C, JA-17 rotor, Beckman Coulter). The cell free extract was filtered and loaded on 2 mL of Ni²⁺ Sepharose HP (GEHealthcare) pre-equilibrated with buffer and incubated for 1 h at 4 °C with rotating. Then, the column was washed with three column volumes of 50 mM Tris/HCl pH 7.5, followed by three column volumes of 50 mM Tris-HCl pH 7.5 containing 5 mM imidazole. The protein was

eluted using 50 mM Tris-HCl pH 7.5 with 500 mM imidazole. Fractions containing yellow protein were applied on a pre-equilibrated Econo-Pac 10DG desalting column (Bio-Rad). The desalted sample was used directly or flash frozen in liquid nitrogen after optional addition of 10% glycerol (final concentration) and stored at -80 °C. Purity and integrity of the purified enzyme batch was confirmed by SDS-Gel analysis and recording of a UV-visible spectrum.

CPDMO and CHMO

Cyclopentadecanone monooxygenase from *Pseudomonas sp.* HI 70 (CPDMO) and cyclohexanone monooxygenase from *Acinetobacter calcoaceticus* NCIMB 9871 (CHMO) were expressed from an identical to above pBAD-His₆-PTDH-Linker-BVMO plasmid, using the same conditions with 0.02% L-arabinose, 50 µg/ml ampicillin and incubation at 24 °C for 36 h with shaking. Purification was also performed exactly as described above for PockeMO.

3.5 Determination of extinction coefficient

Purified enzyme was diluted to a concentration of 5 - 10 µM and distributed to two reaction tubes. To one of the samples, sodium dodecyl sulfate was added to a final concentration of 0.1% (w/v). After 10 minutes the UV-visible spectrum of both samples was recorded. The concentration of holoenzyme originally present was then defined as equal to the concentration of free FAD determined in the SDS sample, by using an extinction coefficient of 11,300 M⁻¹ cm⁻¹. The extinction coefficient of the enzyme was then calculated for the wavelength of highest absorption using the Beer-Lambert law.

3.6 NADPH consumption assay and kinetic parameters

To determine enzymatic activity and kinetic parameters, reactions were followed spectrophotometrically by measuring the decrease in absorption at 340 nm. The reaction mixture (100 - 200 µl) typically contained 50 mM Tris-HCl pH 7.5, 2 µM isolated enzyme, 100 µM NADPH, and varying amounts of substrate, solubilized in 2.5 or 10% co-solvent, if required. All components except for the substrate were kept on ice until immediately before use and the reaction was started by adding NADPH. The mix was then quickly transferred to a 0.1 ml quartz cuvette and the absorbance at 340 nm was measured for at least 40 seconds. Kinetic parameters were obtained by fitting the obtained data to the Michaelis-Menten equation using the software GraphPad Prism 6. In every experiment, control reactions to monitor uncoupling rates were carried out by measuring NADPH consumption in the absence of substrate in buffer (without or with the co-solvent, depending on the experiment).

3.7 pH profile

Activities were measured by first preparing samples of 1 mM bicyclo[3.2.0]hept-2-en-6-one in 100 mM Tris-HCl buffer of the desired pH. Then, 1 µM enzyme (final) and 100 µM NADPH (final) buffered in 10 mM Tris-HCl pH 7.5 were added from concentrated stock solutions in a volume ratio substrate:enzyme:NADPH 8:1:1. Following mixing, NADPH consumption rate was followed spectrophotometrically at 340 nm. The reactions were carried out at 25 °C.

3.8 Temperature profile

Activity at different temperatures was measured with samples of 1 mM bicyclo[3.2.0]hept-2-en-6-one in 50 mM Tris-HCl pH 7.5, mixed with 1 µM of enzyme (final) and 100 µM NADPH (final). Except for the substrate, before mixing, solutions were kept on ice and then the mix was transferred into a spectrophotometer cuvette. The spectrophotometer cell temperature was beforehand set to the desired value and the cuvette was allowed to heat up for 10 seconds, before starting the measurement at 340 nm to follow NADPH consumption.

3.9 T_m determination via ThermoFAD

To determine the apparent melting temperature (T_m) of PockeMO in different conditions, duplicate samples of 25 μ l were prepared in a 96-well thin wall PCR plate. A 50 mM Tris buffer with HCl-adjusted pH containing the desired additives was used. The samples contained 1 mg/ml purified enzyme. The plate was heated from 20 °C to 90 °C, increasing temperature by 0.5 °C every 10 seconds, using an RT-PCR machine (CFX96-Touch, Bio-Rad Laboratories) that measured fluorescence using a 450–490 excitation filter and a 515–530 nm emission filter. The melting point was defined as the temperature when the first derivative of the observed fluorescent signal showed a maximum.

3.10 Temperature and solvent stability

Activity measurements were performed in 50 mM Tris/HCl using 1 mM bicyclo[3.2.0]hept-2-en-6-one as the substrate. NADPH consumption was followed spectrophotometrically at 340 nm for 1 min. Purified enzyme was kept on ice until the first activity measurement and subsequently incubated without shaking at 25 or 35 °C in a thermomixer (Eppendorf). Several measurements were performed for up to 24 h after incubation start.

3.11 Bioconversions

A typical conversion mix contained 50 mM Tris/HCl pH 9.0 buffer, 10 μ M purified phosphite dehydrogenase enzyme, 10 mM sodium phosphite and 100 μ M NADPH. In a closed 20 ml glass vial, 1 ml of that mix was shaken at 30 °C for 24 h before analysis. For the large cyclic ketones and steroids the substrate stock solution concentration was 8 mM solubilized in 100% dioxane. Conversion of ketone mixes were prepared with a substrate mix stock solution in 100% dioxane that contained each of the ketones in a concentration of 4 mM (mix of 4-7 substrates) or 2 mM (mix of 15 substrates). The stock was then diluted 1:10 in the final reaction mix.

3.12 Analysis of isomeric bicyclo[3.2.0]hept-2-en-6-one lactones by chiral GC

Bioconversion was performed as described above, but for only two hours, with 4 μ M enzyme and 10 mM racemic bicyclo[3.2.0]hept-2-en-6-one. 500 μ l of the conversion mixes were extracted three times with 500 μ l of tert-butyl methyl ether supplemented with 0.1% (v/v) mesitylene as an internal standard, dried over anhydrous sodium sulfate, and analyzed by chiral gas chromatography (GC) using a 7890A GC System (Agilent Technologies) and a CP Chiralsil Dex CB column (25 m x 0.25 mm x 0.25 μ m, Agilent) chiral column. The temperature program of the column oven was from 40 °C to 130 °C at 10 °C/min, then 130 °C for 15 min; and finally from 130 °C to 40 °C at 10 °C/min. Configurations of the products were assigned by comparison of the GC retention times order with published data.^[4]

3.13 Analysis of products by GC-MS

600 μ l of the conversion mixes were extracted three times with equal amounts of ethyl acetate, dried over anhydrous sodium sulfate and the solvent evaporated using a speed vac concentrator (Savant). Compounds were then resuspended in 300 μ l ethyl acetate and analyzed via a GC-MS QP2010 ultra (Shimadzu) with electron ionization and quadrupole separation. The column used was a HP-1 (Agilent, 30 m x 0.32 mm x 0.25 μ m). The programs for the GC can be inferred from the following table:

Compound/Mix of compounds	Program
Cyclododecanone, Cycloundecanone, Cyclopentadecanone	Injection temp.: 300 °C Oven program: 100 °C; 5 °C/min until 200 °C; 200 °C, 2 mins.
Steroids	Injection temp.: 250 °C Oven program: 190 °C, 2 mins; 12 °C/min until 325 °C; 325 °C, 3 mins.

Mix of ketones without steroids	Injection temp.: 250 °C Oven program: 30 °C, 3 mins; 7.5 °C/min until 200 °C; 200 °C, 3 mins; 15 °C/min until 325 °C; 325 °C, 3 mins.
Mix of ketones with steroids	Injection temp.: 260 °C Oven program: 30 °C, 5 mins; 5 °C/min until 70 °C; 70 °C, 5 mins; 5 °C/min until 130 °C; 130 °C, 5 mins; 15 °C/min until 325 °C; 325 °C, 1 min.

In all cases, 1 μ l was injected into the GC and the split ratio was 5:0. The software to analyze chromatograms, MS spectra and to generate the figures was GCMSsolution Postrun Analysis 4.11 (Shimadzu). The library for the MS spectra was NIST1.

3.14 NMR analysis

To produce sufficient amount for ¹H NMR, bioconversion was performed on a multi milliliter scale, with reaction mixes of 15 ml and 40 ml for stanolone and pregnenolone respectively. For stanolone, the standard concentration of 800 μ M, and for the highly insoluble pregnenolone, a concentration of 500 μ M was used in the reaction. Extraction was performed three times with ethyl acetate in a separation funnel, and the solvent was removed in a rotary evaporator. Both steroids afforded a white/light brown pellet which was dried over N₂ overnight and then resuspended in 1 ml CDCl₃. NMR analysis was performed using a Varian Unity Plus (¹H: 400 MHz) spectrometer. Chemical shifts are denoted in δ -units (ppm) relative to the residual solvent peak (CDCl₃: ¹H δ =7.26). The splitting parameters are named as follows: s = singlet, d = doublet, t = triplet, m = multiplet, dd = doublet of doublets.

3.15 Crystallography

Purified PockeMO used for crystallization experiments was concentrated to 12 mg ml⁻¹ of enzyme in 20 mM Tris-HCl buffer at pH 7.5. For initial crystallization screening, we tested the kits Classic 1 and 2 from Jena Bioscience, Midas 1 and 2 from Molecular Dimensions, and the pH clear and AmSO₄ suites from Qiagen. An Oryx 8 crystallization robot (Douglas Instruments, UK) was used for robotic screening. Crystallization hits were optimized manually using the sitting-drop setup. After optimization, the best diffracting crystals were found in a condition containing 2.7 M ammonium sulfate, 100 mM MES sodium salt pH 6.5, and 5% v/v PEG400. Crystals grew from drops prepared 1:1 pockeMO (9 mg ml⁻¹) and reservoir at 20 °C in 48 h as bright yellow solids, shaped as cut diamonds with hexagonal base. These crystals were extremely reproducible and co-crystallization with NADP⁺ was not impairing or improving their formation. Crystals were harvested from the mother liquor using nylon cryoloops (Hampton research, USA) and flash-cooled in liquid nitrogen after a short soak in a solution containing 20% v/v PEG400, 5 mM NADP⁺, 3 M ammonium sulfate, 100 mM MES sodium salt pH 6.5. X-ray diffraction data used for structure determination and refinement were collected at PX beamline of the Swiss Light Source in Villigen, Switzerland (SLS) and at the ID23-EH1 beamline of the European Synchrotron Radiation Facility in Grenoble, France (ESRF). Data were processed and scaled using the CCP4 programs *iMOSFLM* for indexing and integration and *AIMLESS* for space group assignment and scaling.^[5] The space group symmetry together with final data-collection and processing statistics are listed in Table S1. The structures were solved with *MOLREP*^[6], and the coordinates of PAMO (PDB: 2YLT)^[7] as search model. Manual building, addition of water molecules, and crystallographic refinement were performed with *COOT*,^[8] *REFMAC5*^[9] and other programs of the CCP4 suite. Structure quality and validation were assessed using the wwPDB validation server. The final model coordinates were deposited in the Protein Data Bank with accession code 5MQ6. Figures were created with Chimera.^[10]

Table S2. Data collection and refinement statistics.^[a]

	Oxidised enzyme PDB: 5MQ6
Wavelength (Å)	1
Resolution range	75.21 - 2.0 (2.071 - 2.0)
Space group	P ₂ , ₂ , ₂
Unit cell (Å), (°)	59.2 91.02 133.52 90 90 90
Total reflections	306952 (22686)
Unique reflections	49615 (3629)
Multiplicity	6.2 (6.2)
Completeness (%)	100.00 (100.00)
Mean I/sigma (I)	8.9 (1.6)
Wilson B-factor (Å ²)	25.50
R-merge (%)	0.162 (1.207)
CC1/2	0.98 (0.50)
Reflections used in refinement	48005 (3519)
Reflections used for R-free	1541 (103)
R-work (%)	0.153 (0.266)
R-free	0.231 (0.339)
Number of non-hydrogen atoms	5433
Macromolecule	4970
Ligands	98
Protein residues	636
RMS (bonds) (Å)	0.022
RMS (angles) (°)	2.13
Ramachandran favoured (%)	96
Ramachandran allowed (%)	3.9
Ramachandran outliers (%)	0.16
Rotamer outliers (%)	5
Clashscore	5.00
Average B-factor	32.84
Macromolecules	32.59
Ligands	33.74
Solvent	36.00

[a] Statistics for the highest-resolution shell are shown in parentheses.

REFERENCES

- [1] C. Szolkowy, L. D. Eltis, N. C. Bruce, G. Grogan, *ChemBioChem* **2009**, *10*, 1208-1217.
- [2] E. Beneventi, G. Ottolina, G. Carrea, W. Panzeri, G. Fronza, P. C. K. Lau, *Journal of Molecular Catalysis B-Enzymatic* **2009**, *58*, 164-168.
- [3] S. Kumar, G. Stecher, K. Tamura, *Molecular biology and evolution* **2016**, msw054.
- [4] N. M. Kamerbeek, A. J. Olsthoorn, M. W. Fraaije, D. B. Janssen, *Appl. Environ. Microbiol.* **2003**, *69*, 419-426.
- [5] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2011**, *67*, 235-242.
- [6] A. Vagin, A. Teplyakov, *J. Appl. Crystallogr.* **1997**, *30*, 1022-1025.
- [7] R. Ortu, H. M. Dudek, C. Martinoli, D. E. Torres Pazmino, A. Royant, M. Weik, M. W. Fraaije, A. Mattevi, *J. Biol. Chem.* **2011**, *286*, 29284-29291.
- [8] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2010**, *66*, 486-501.
- [9] G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2011**, *67*, 355-367.
- [10] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605-1612.

Chapter IV:**Hydroxysteroid dehydrogenase (HSDH) enzymes: similar active sites can host molecules in different orientations to modify specificity**

The aim of this project was to provide structural data to explain the different reactivities of 7 β -hydroxysteroid dehydrogenases (7 β -HSDH), enzymes that usually display a remarkable selectivity in producing ursodeoxycholic acid. In the framework of steroids exploitation for medicinal purposes, the molecule ursodeoxycholic acid is reckoned as a drug for treatment of cholesterol gallstones and is already approved by the FDA and used in clinical practice [1]. Such a molecule is particularly abundant in bear's liver and has been used in traditional medicine for a long time, but the extraction from natural sources is incompatible with the demand and environmentally unsustainable. The molecule is currently obtained by chemical route, using the Wolff–Kishner reduction of 12-keto-UDCA to UDCA. To replace the inefficient chemical process, numerous groups are trying to develop a biocatalytic way to convert related steroids in UDCA, taking advantage of bacterial enzymes of the oxidoreductase class, with high regio- and stereo-selectivity for the hydroxyl groups attached to the gonane scaffold.

7 β -HSDH has been found to selectively convert 7-keto-lithocholic acid into the molecule of interest by using NADPH as a cofactor [2]. In an attempt to rationalize the differences between the target protein and the homolog 7 α -hydroxysteroid dehydrogenase (7 α -HSDH) [3], which produces chenodeoxycholic acid using NADH and the same substrate 7-keto-lithocholic acid, we proceeded through expression, crystallization and determination of the three-dimensional structure of 7 β -HSDH. By providing a high-resolution model, future amelioration of the enzyme's catalysis by site directed mutagenesis may be performed, with possible application in drug production. We characterized 7 β -HSDH from *Collinsella aerofaciens*, a commensal bacterium present in human gut. The enzyme itself was already investigated for its biochemical properties [4] while the organism has been regarded as a source of oxidoreductases to perform reactions for added-value steroids production.

The high-resolution structure of 7β -HSDH compared to the one of 7α -HSDH provides an insightful example of reversed allocation of the ligand in the active site. Despite the 19% identity and the complete superposition of the overall folding, 7β -HSDH has a different arrangement of the charges in the active site and an elongated C-terminus. From the observation of these elements and the following modelling of the ligand – as it was not present in the structure we solved – we realized that, with respect to the homolog, the orientation is rotated 180° around the reduction site of the molecule. This binding mode is essential to allow correct stereo-specificity of the reaction, as the reduction site is placed in the right spot for the reaction to occur in β configuration, opposite to the α face where the hydroxyl points in the reaction performed by 7α -HSDH. The presence of the elongated C-terminus enfolding the aliphatic chain of the steroid assists the binding in the flipped orientation; in absence of this element, the steroid molecule would be too exposed to the solvent. This observation was confirmed by biochemistry experiments on trimmed mutants. A follow-up article was published by a Chinese group [5] one year after ours. In this work, the group managed to crystallize 7β -HSDH in complex with NADPH, confirming our binding model and the role of the two residues Arg40 and Arg41 in interacting with the additional phosphate of the cofactor.

The development of a one-pot multi-enzyme reaction for ursodeoxycholic acid production would be a convenient way to exploit the potentiality of the enzyme. This system is already established for some preliminary steps but not for the actual synthesis of UDCA [6], [7]. Thanks to the cofactor specificity 7α -HSDH can be used to convert the abundant chenodeoxycholic acid to 7-keto-lithocholic acid using NAD^+ and releasing NADH while 7β -HSDH can perform the final reaction using NADPH and releasing ursodeoxycholic acid and NADP^+ . It may even be possible to regenerate the cofactors in the same pot by using two other enzymes of the steroids modification pathway, where NADH is required to reduce a ketone in position 3 or 12 and NADP^+ can oxidize a hydroxyl group of the aliphatic chain. In this way, a recycling cycle can be established while using different raw material to synthesize the final molecule. The contribution of chemistry in production of ursodeoxycholic acid may be drastically reduced in the coming years by implementing the current knowledge on 7β -HSDH.

References

1. Hyun JJ, Lee HS, Kim CD, Dong SH, Lee SO, Ryu JK, Lee DH, Jeong S, Kim TN, Lee J, Koh DH, Park ET, Lee IS, Yoo BM, Kim JH Efficacy of Magnesium Trihydrate of Ursodeoxycholic Acid and Chenodeoxycholic Acid for Gallstone Dissolution: A Prospective Multicenter Trial. *Gut Liver*. 2015 Jul;9(4):547-55.
2. Shiotsuki H, Maeda Y, Chijiwa K. Purification and characterization of 7 β -hydroxysteroid dehydrogenase from rabbit liver microsomes. *J Steroid Biochem Mol Biol*. 2004 Jul;91(3):185-90.
3. Tanaka N, Nonaka T, Yoshimoto T, Tsuru D, Mitsui Y. Crystallization and preliminary X-ray crystallographic studies of 7 α -hydroxysteroid dehydrogenase from *Escherichia coli*. *Acta Crystallogr D Biol Crystallogr*. 1996 Jan 1;52(Pt 1):215-7.
4. Liu L, Aigner A, Schmid RD. Identification, cloning, heterologous expression, and characterization of a NADPH-dependent 7 β -hydroxysteroid dehydrogenase from *Collinsella aerofaciens*. *Appl Microbiol Biotechnol*. 2011 Apr;90(1):127-35.
5. Wang R, Wu J, Jin DK, Chen Y, Lv Z, Chen Q, Miao Q, Huo X, Wang F. Structure of NADP⁺-bound 7 β -hydroxysteroid dehydrogenase reveals two cofactor-binding modes. *Acta Crystallogr F Struct Biol Commun*. 2017 May 1;73(Pt 5):246-252.
6. Sun B, Kantzow C, Bresch S, Castiglione K, Weuster-Botz D. Multi-enzymatic one-pot reduction of dehydrocholic acid to 12-keto-ursodeoxycholic acid with whole-cell biocatalysts. *Biotechnol Bioeng*. 2013 Jan;110(1):68-77.
7. Sun B, Hartl F, Castiglione K, Weuster-Botz D. Dynamic mechanistic modeling of the multienzymatic one-pot reduction of dehydrocholic acid to 12-keto ursodeoxycholic acid with competing substrates and cofactors. *Biotechnol Prog*. 2015 Mar-Apr;31(2):375-86.



STRUCTURE NOTE

Structural and biochemical insights into 7 β -hydroxysteroid dehydrogenase stereoselectivity

Simone Savino,¹ Erica Elisa Ferrandi,² Federico Forneris,¹ Stefano Rovida,¹ Sergio Riva,² Daniela Monti,^{2*} and Andrea Mattevi^{1*}

¹ Department of Biology and Biotechnology, University of Pavia, via Ferrata 9, Pavia 27100, Italy

² Istituto di Chimica del Riconoscimento Molecolare, CNR, via Mario Bianco 9, Milano 20131, Italy

ABSTRACT

Hydroxysteroid dehydrogenases are of great interest as biocatalysts for transformations involving steroid substrates. They feature a high degree of stereo- and regio-selectivity, acting on a defined atom with a specific configuration of the steroid nucleus. The crystal structure of 7 β -hydroxysteroid dehydrogenase from *Collinsella aerofaciens* reveals a loop gating active-site accessibility, the bases of the specificity for NADP⁺, and the general architecture of the steroid binding site. Comparison with 7 α -hydroxysteroid dehydrogenase provides a rationale for the opposite stereoselectivity. The presence of a C-terminal extension reshapes the substrate site of the β -selective enzyme, possibly leading to an inverted orientation of the bound substrate.

Proteins 2016; 84:859–865.

© 2016 Wiley Periodicals, Inc.

Key words: biocatalysis; short-chain dehydrogenase; steroid; stereoselectivity; NADP.

INTRODUCTION

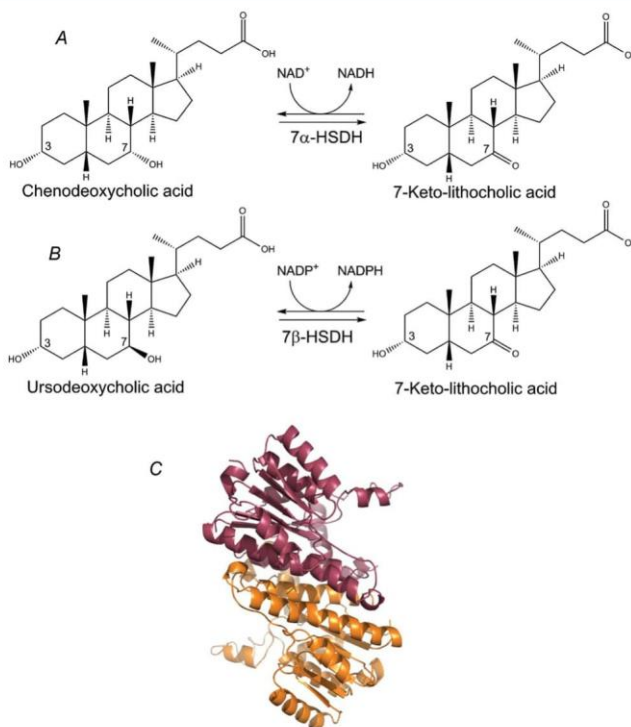
Hydroxysteroid dehydrogenases (HSDHs) are NAD(P)H-dependent enzymes belonging to the short-chain dehydrogenase/reductase family. They perform oxidation of the hydroxyl groups present on the core structure (gonane) of steroid molecules [Fig. 1(A,B)].¹ The gonane consists of 17 carbon atoms arranged in three cyclohexane and one cyclopentane rings, and HSDHs have been shown to display very high regioselectivity for the hydroxyl groups at different positions, for example, at C-3, C-7, and C-12 of bile acids.² Moreover, for each one of these positions, HSDHs usually show high stereoselectivity for either the hydroxyl group above (β configuration) or below (α configuration) the plane of the steroid molecule [with reference to standard abso-

lute configuration; see Fig. 1(A,B)]. In this framework, 7 β -hydroxysteroid dehydrogenase (7 β -HSDH; EC 1.1.1.201) catalyzes the NAD(P)⁺-dependent reversible oxidation of ursodeoxycholic acid to 7-keto-lithocholic acid, thus showing an exclusive specificity for the hydroxyl group in β configuration at the C-7 position

Additional Supporting Information may be found in the online version of this article.

Abbreviations: HSDH, hydroxysteroid dehydrogenase; 7 β -HSDH, 7 β -hydroxysteroid dehydrogenase; 7 α -HSDH, 7 α -hydroxysteroid dehydrogenase.

*Correspondence to: Andrea Mattevi; Department of Biology and Biotechnology, University of Pavia, via Ferrata 9, 27100 Pavia, Italy. E-mail: andrea.mattevi@unipv.it or Daniela Monti; Istituto di Chimica del Riconoscimento Molecolare, CNR, Via Mario Bianco 9, 20131 Milano, Italy. E-mail: daniela.monti@icrm.cnr.it
Received 26 January 2016; Revised 7 March 2016; Accepted 12 March 2016
Published online 23 March 2016 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.25036

**Figure 1**

HSDH substrates and three-dimensional structure of 7 β -HSDH. (A) Chenodeoxycholic acid (3 α ,7 α -dihydroxy-5 β -cholan-24-oic acid); and ursodeoxycholic acid (3 α ,7 β -dihydroxy-5 β -cholan-24-oic acid); (B) are the prototypical substrates of 7 α -HSDH and 7 β -HSDH enzymes, respectively. The two molecules differ for the configuration of the hydroxyl group at C-7 of the gonane steroid nucleus. (C) 7 β -HSDH is a dimer of identical subunits arranged with C2 symmetry. The two protein chains are colored differently. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

[Fig. 1(B)]. To date, only three 7 β -HSDHs have been cloned and characterized to some extent, specifically from *Collinsella aerofaciens*,³ *Clostridium absonum*,⁴ and *Ruminococcus gnavus* strains.⁵ Years ago, the substrate specificity of the *C. absonum* 7 β -HSDH was investigated⁶ and the enzyme was applied to the synthesis of a cholic acid derivative.⁷ More recently, the synthetic exploitation of 7 β -HSDHs for the preparative production of ursodeoxycholic acid, a FDA-approved drug for treatment of cholesterol gallstones, and current sole alternative to surgical intervention,⁸ has been widely investigated by different groups.^{9,10}

Intrigued by the question concerning the basis for the diversity in stereo- and regioselectivity among HSDH enzymes, we have solved the crystal structure of the 7 β -HSDH from *C. aerofaciens* and performed mutagenesis experiments. These data outline a conserved NADP(H) site, whereas the architecture of the substrate-binding pocket features a few characteristic elements that can be correlated with the distinct stereoselectivity of the enzyme. We foresee that this knowledge could pave the way for a rational engineering to enhance the enzymatic activity of this useful biocatalyst, further boosting its application in preparative synthesis of high value compounds.

MATERIALS AND METHODS

Reagents

All reagents were purchased from Sigma-Aldrich and ACROS organics. Screening kits for crystallization were obtained from Qiagen and Jena Bioscience. Crystallization plates were from Molecular Dimensions and Douglas Instruments for manual and robotic setups, respectively.

Protein expression and purification

E. coli BL21(DE3) cells were transformed with plasmid pET24b-7 β -HSDH, encoding for the N-terminally 6xHis-tagged 7 β -HSDH from *C. aerofaciens* ATCC 25986 (GenBank accession no: ZP 01773061).³ Pre-cultures starting from single bacterial colonies were grown overnight at 37°C and 220 rpm agitation in 200 mL flasks, containing 50 mL of LB supplemented with kanamycin as selection antibiotic. Preparative scale cultures were inoculated using 20 mL of the overnight pre-culture. The optical density of these cultures at 600 nm was monitored using a spectrophotometer (GE Healthcare). Once density reached the value of 0.85, flasks were transferred for one hour at 4°C before induction using 500 μ M isopropyl- β -D-thiogalactopyranoside. Cultures were gently shaken (140 rpm) overnight at 25°C. Cell harvesting was performed by low speed centrifugation (5000 rpm) to separate the liquid medium from cells. The cell pellet was re-suspended in buffer A (20 mM KPi, 500 mM NaCl, 20 mM imidazole, pH 8), supplemented with 4 μ g/mL DNase I and 0.5 mg/mL lysozyme, and was processed into an EmulsiFlex C3 homogenizer (Avestin). The soluble protein fraction was separated from cell debris by high speed centrifugation (70,000 rpm). The clarified supernatant was loaded on a NiSepharose FF 5 mL (GE Healthcare) column mounted on ÄKTA Prime system (GE Healthcare) at room temperature. The enzyme eluted with a 10% step buffer B (20 mM KPi, 500 mM NaCl, 1M imidazole, pH 8). The eluted sample was concentrated using Amicon Ultra Centrifugal filters (Merck Millipore), then loaded onto a Superdex 200 column (GE Healthcare) equilibrated in a buffer consisting of 20% v/v glycerol and 20 mM KPi, pH 7.0.

Protein crystallization

Purified 7 β -HSDH used for crystallization experiments was concentrated to 10–15 mg enzyme/mL in 20% v/v glycerol, 20 mM KPi pH 7.0 and supplemented with 5 mM NADP⁺. For initial crystallization screening, we tested the kits Classic 1 to 10 from Jena Bioscience, using a Douglas Oryx 8 crystallization robot (Douglas Instruments, UK). Crystallization hits were optimized manually using the sitting-drop setup. After optimization, the best

Table I

Data Collection and Refinement Statistics

PDB code	5FYD
Space group	P 2 ₁ 2 ₁ 2
<i>a</i> , <i>b</i> , <i>c</i> (Å) ^a	114.77, 83.04, 47.91
Resolution (Å)	1.6
<i>R</i> _{sym} (%) ^{b,c}	10.9 (90.2)
CC _{1/2} ^{d,e}	0.99 (0.436)
Completeness (%) ^c	100 (100)
Redundancy ^c	6.6 (6.4)
Unique reflections	61266
Average intensity/ σ ^f	12.1 (1.3)
Number of protein atoms	3918
Number of water atoms	345
Ramachandran outliers (%)	0.2 (1 amino acid)
Ramachandran allowed (%)	98
Average B, all atoms (Å ²)	23.0
<i>R</i> _{cryst} (%) ^g	17.6
<i>R</i> _{free} (%)	20.8
Rmsd bond length (Å)	0.009
Rmsd bond angle (°)	1.32

^aThe unit cell of the orthorhombic 7 β -HSDH crystals contains two enzyme subunits. Subunit A contains residues 1–259. Subunit B contains residues 1–203 and 212–259.

^b $R_{sym} = \sum_i |I_i - \langle I \rangle| / \sum_i I_i$, where I_i is the intensity of i^{th} observation and is the mean intensity of the reflection.

^cValues in parentheses are for reflections in the highest resolution shell.

^dA cutoff criterion for resolution limits was applied on the basis of the mean intensity correlation coefficient of half-subsets of each dataset (CC_{1/2}).

^e $R_{cryst} = \sum_i |F_{obs} - F_{calc}| / \sum_i |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively. R_{cryst} and R_{free} were calculated using the working and test sets, respectively.

condition producing crystals contained 1M sodium formate, 100 mM sodium acetate, pH 4.6. Crystals routinely grew at room temperature as well-shaped cubes after 3–4 days.

X-ray data collection and processing

Crystals were harvested from the mother liquor using nylon cryoloops (Hampton research, USA) and flash-cooled in liquid nitrogen after a short soak in a solution containing 25% v/v glycerol, 5 mM NADP⁺, 1.4M sodium formate, 100 mM sodium acetate, pH 4.6. X-ray diffraction data used for structure determination and refinement were collected from one single crystal at the beamline ID23-EH1 of the European Synchrotron Radiation Facility (ESRF, Grenoble, France). Data were processed and scaled using the CCP4 programs *iMOSFLM* for indexing and integration and *AIMLESS* for space group assignment and scaling.¹¹ The space group symmetry together with final data-collection and processing statistics are listed in Table I.

Structure determination and refinement

The crystal structure of 7 β -HSDH from *C. aerofaciens* was solved by molecular replacement using the program *PHASER*¹² and the apo-form structure of 7 α -HSDH

from *E. coli* as search model (PDB entry 1AHH).¹³ The electron density map calculated after molecular replacement was visually inspected and used for automated model building with the program *phenix.autobuild*.¹⁴ Manual building, addition of water molecules, and crystallographic refinement were performed with *COOT*, *REFMAC5* and other programs of the CCP4 suite.¹¹ Final refinement statistics are shown in Table I. Structure quality and validation were assessed using the wwPDB validation server. The final model coordinates were deposited in the Protein Data Bank with accession code 5FYD.

Mutagenesis and mutant characterization

Mutants carrying a C-terminal deletion of 14 (7 β -HSDH $_{\Delta 250-263}$) and 27 (7 β -HSDH $_{\Delta 237-263}$) amino acids were generated by PCR amplification as described in the Supporting Information. Both mutant proteins were expressed in *E. coli* BL21(DE3) cells (see Supporting Information). The mutant 7 β -HSDH $_{\Delta 250-263}$ was successfully purified using the same protocol as the wild-type protein to perform enzyme activity assays. *E. coli* 7 α -HSDH was produced and purified as described in Refs. 4,13. Enzymatic activities on wild-type 7 α -HSDH and 7 β -HSDH proteins were assayed spectrophotometrically measuring the reduction of NADP⁺ at 340 nm ($\epsilon = 6.22 \text{ mM}^{-1} \text{ cm}^{-1}$) and 25°C on a Jasco V-530 UV/VIS spectrophotometer (Easton, MD, USA). The assay mixture (1 mL) contained 2.5 mM substrate (cholic acid, chenodeoxycholic acid, ursocholic acid, or ursodeoxycholic acid) in 0.1M KPi pH 9.0 and 0.2 mM NADP⁺. One unit is defined as the enzyme activity that reduces 1 μmol of NADP⁺ per minute under these assay conditions.

RESULTS AND DISCUSSION

The overall structure of 7 β -hydroxysteroid dehydrogenase

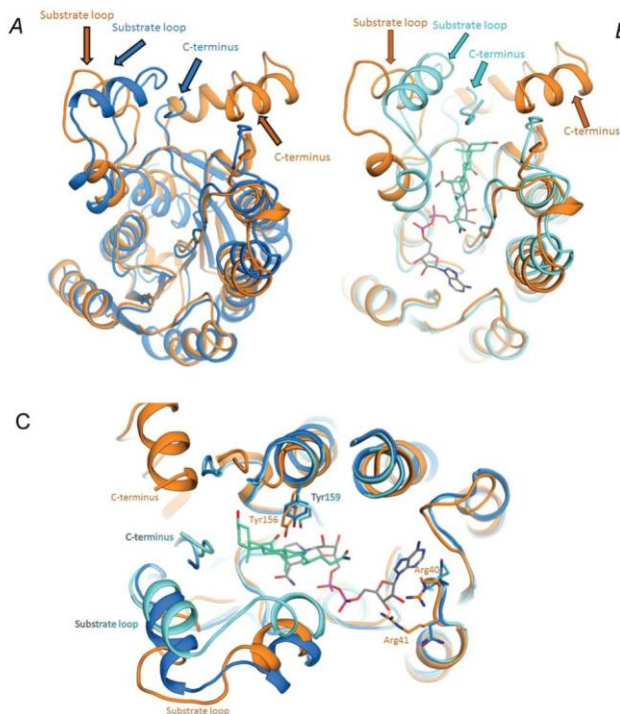
We have successfully expressed and purified in milligram amounts 7 β -HSDH from *C. aerofaciens* with a noncleavable N-terminal histidine tag. After extensive screening with commercially available kits, we obtained crystals of the apo-enzyme using the sitting drop-method and a reservoir consisting of 1M sodium formate and 100 mM sodium acetate, pH 4.6. Protein solutions used for crystallization contained up to 10 mM NADP⁺ or NADPH, which, however, turned out not to be bound in the crystalline enzyme (below). The crystals exhibited very good diffraction power and the crystal structure was solved and refined at 1.6 Å resolution as summarized in Table I.

7 β -HSDH is a tight dimer of identical subunits with extensive intermolecular contacts [Fig. 1(C)]. Typically,

members of the short-chain dehydrogenase/reductase family are tetrameric (i.e. a dimer of dimers; see Ref. 13). Inspection of the three-dimensional structure shows that 7 β -HSDH features a characteristic C-terminal α -helical extension [Fig. 2(A)], which apparently prevents formation of the tetrameric assembly by causing steric hindrance. Nevertheless, with the exception of this feature, the overall folding topology of 7 β -HSDH is very similar to that of the short-chain dehydrogenase family. It exhibits the classic $\alpha\beta$ doubly wound architecture in which the core of the Rossmann-fold domain forms the NAD(P)H-binding site. A PDB search performed with DALI¹⁵ indicates that 7 β -HSDH is structurally similar to several dehydrogenase structures, the closest one being NADPH-sorbitose reductase (entry 3AL2, Z-score 27.7). Importantly, 7 α -HSDH from *E. coli* is among the nearest homologs identified by DALI, with a Z-score of 25.5.¹³ DALI superposition of 7 α -HSDH and 7 β -HSDH structures returns a root-mean-square-deviation of 2.3 Å for 230 C α atoms and 17% sequence identity. Such a structural relation enables a comparative analysis to pinpoint the structural features that lead to the opposite stereoselectivity of the 7 α - and 7 β -HSDHs, which was the main aim of our study [Fig. 1(A,B)].

Active site location and gating loop

The catalytic center is located in a crevice of the Rossmann fold and is lined by the C-terminal α -helical residues and the so-called substrate-binding loop, which are both crucial elements for enzyme function and substrate recognition [Fig. 2(A,B)]. In particular, the presence of a flexible substrate loop is a feature often found in short-chain dehydrogenases. In *E. coli* 7 α -HSDH, this loop undergoes a large conformational change upon binding to the glycochenodeoxycholic acid substrate, with C α shifts as large as 8.5 Å.¹³ As a result, the binding site becomes sealed, making the core ring of the substrate partly solvent inaccessible [Fig. 2(A,B)]. Remarkably, inspection of the 7 β -HSDH crystals indicates that the substrate loop (residues 200–215) is involved in crystal packing. Specifically, the loops of two symmetry-related subunits belonging to different dimers extensively interact with each other, along the crystallographic two-fold axis (Table I). This interaction is made possible by a highly extended conformation of the loop, which protrudes out of each enzyme monomer to interact with the equivalent residues of a symmetry-related protein (Supporting Information Fig. S1). As a result, the active site of the crystalline enzyme is occluded by the crystal packing, which explains why all our attempts to obtain the structures of the enzyme bound to NADPH(H) and/or the substrate were unsuccessful. The structure of 7 β -HSDH, therefore, represents the enzyme in the apo-form with the active-site loop in an open conformation, very

**Figure 2**

Structural basis for the different substrates selectivity of 7β-HSDH and 7α-HSDH. (A) Superposition of 7β-HSDH (orange) and 7α-HSDH (blue; PDB entry 1AHH) in the apo-protein forms. The overall Rossmann fold structures are conserved. Elements of diversity are the longer C-terminus of 7β-HSDH (residues 240–259) and a slightly altered conformation of the substrate loop, which retains an open conformation in both structures (residues 192–207 in 7β-HSDH and 195–210 in 7α-HSDH). (B) Superposition of 7β-HSDH (apo-form; orange) with the holo-form of 7α-HSDH (cyan; PDB entry 1FMC). NAD⁺ (gray) and chenodeoxycholate (pale green) bound to 7α-HSDH are shown [Fig. 1(A)]. (C) Close-up view of the superposition shown in panel B. The picture has been rotated to highlight the closure movement of the substrate loop and the shorter C-terminus of 7α-HSDH in proximity of the substrate.

similar to that observed in 7α-HSDH structure without ligands [Fig. 2(A)].

The NAD(P)H-binding site

C. aerofaciens 7β-HSDH selectively uses NADP⁺ as hydride ion acceptor for substrate dehydrogenation.³ Structural and sequence comparison with NAD⁺-bound 7α-HSDH shows that the residues involved in NAD(P)H binding are highly conserved (Supporting Information Table S1). This implies that *C. aerofaciens* 7β-HSDH can

be confidentially predicted to bind NADPH in the same conformation observed in the NAD⁺-bound complexes of *E. coli* 7α-HSDH.¹³ However, there is a difference in that 7β-HSDH is specific for NADP(H) as opposed to the preference of 7α-HSDH for NAD(H). This variation can be nicely rationalized by the inspection of the crystal structures. The only evident difference in the cofactor-binding sites between the two dehydrogenases concerns the residues that in 7α-HSDH form the binding site for the adenine ribose of NAD⁺ (Asp42-Ile43-Asn44-Ala45). This region contains an Asp side-chain (Asp42) that

H-bonds to the ribose hydroxyl groups (Supporting Information Table S1). The equivalent loop of 7 β -HSDH displays an Arg pair (Arg40–Arg41) that is perfectly suited to bind the 2'-phosphate group of NADP(H), consistent with the cofactor selectivity of the enzyme (Supporting Information Fig. S2).

The catalytic center

The reaction catalyzed by short-chain dehydrogenases generally involves the transfer of a hydride anion from the substrate CH-OH group to NAD(P)⁺. In order to promote catalysis, the alcohol and nicotinamide moieties must be juxtaposed within the active site. In particular, substrate activation requires the deprotonation of the substrate –OH group to favor dehydrogenation and formation of the ketone (or aldehyde) product. In the case of 7 α -HSDH, this substrate-activating role is played by a Ser–Tyr dyad (Ser146–Tyr159). Specifically, the Tyr is proposed to be in the anionic form, ready to abstract a proton from the substrate.¹³ Indeed, the crystal structure of the *E. coli* 7 α -HSDH ternary complex shows that the substrate binds with its 7 α -OH group in direct contact with these residues to establish H-bond interactions. Furthermore, the substrate β -face is oriented toward the nicotinamide to promote transfer of the β -hydride on the substrate C7 carbon to the C4 atom of the cofactor [Fig. 1(A)]. In this context, a key feature revealed by the structural comparison between the 7 β - and 7 α -dehydrogenases is that the Ser–Tyr is conserved. Indeed, Ser143 and Tyr156 of 7 β -HSDH are homologous with and structurally superimpose to Ser146 and Tyr159 of 7 α -HSDH [Fig. 2(C) and Supporting Information Fig. S2]. This implies that these residues very likely perform the same role in the two enzymes by abstracting the hydroxyl proton from the substrate. Thus, despite the opposite stereochemistry of the reaction, the two dehydrogenases share an evolutionary and structurally conserved pattern of catalytic residues.

The C-terminal residues

As opposed to the conserved amino acids underlying cofactor binding and catalysis, the crystal structure of 7 β -HSDH also displays a unique C-terminal extension. This characteristic region (residues 240–259) forms two α -helices that are connected by a short loop and are located at the rim of the substrate-binding site (Fig. 2). These helices are absent in 7 α -HSDH because of a truncated C-terminal segment that is shorter by 15 amino acids compared to the sequence of the 7 β -HSDH. This is the largest and most evident difference between the two enzymes, as it causes a reshaping of the binding site for the steroid substrate. Therefore, this different morphology logically represents the prime (though not the only)

candidate for rationalizing the opposite stereoselectivity of the 7 α - and 7 β -HSDH-catalyzed reactions.

To probe this notion, two deletion mutants have been constructed and expressed. They lack 14 (7 β -HSDH $_{\Delta 250-263}$) and 27 (7 β -HSDH $_{\Delta 237-263}$) C-terminal residues, respectively. Enzymatic activities against 7 β -HSDH (ursodeoxycholic acid and ursolic acid) as well as 7 α -HSDH (chenodeoxycholic acid and cholic acid) [Fig. 1(A,B) and Supporting Information Fig. S3] substrates were experimentally tested using the extracts obtained from *E. coli* cells over-expressing each mutant or the *wild-type* enzymes (Supporting Information Fig. S3). No activity was detected for 7 β -HSDH $_{\Delta 237-263}$, while weak substrate conversions were found for the cells expressing 7 β -HSDH $_{\Delta 250-263}$. For this reason, the latter variant was further purified by affinity chromatography for a more in-depth characterization. The mutant showed poor activity on ursolic acid of 0.012 U/mg, which is 300-fold lower than that measured for the wild-type 7 β -HSDH (3.7 U/mg). Similarly, the measured activity on ursodeoxycholic acid was only 0.01 U/mg, 1000-fold lower than that of the wild type (17.48 U/mg). Such drastic effects support the direct role in substrate binding of the peculiar C-terminal extension of 7 β -HSDH. However, C-terminal truncations do not lead to the acquisition of 7 α -HSDH activity as no conversion was detected using either chenodeoxycholic acid or cholic acid as a substrate (Supporting Information Fig. S3). This is not surprising, since binding of such bulky substrates will involve the contribution of many more protein elements in addition to the C-terminal segment, and a swap in stereoselectivity can be hardly expected to simply arise from the deletion of such a segment.

Clues about the stereoselectivity of the 7 β -HSDH reaction

In NAD(P)H-dependent enzymes, substrate dehydrogenation generally involves a hydride transfer. This crucial catalytic step requires that the substrate hydrogen atom points toward the C4 atom of the NAD(P)⁺ nicotinamide ring. With reference to the 7 β -HSDH reaction, the substrate must bind with its α -face oriented toward the cofactor ring to allow transfer of the 7 α -hydride [Fig. 1(B)]. Conversely, in the case of the 7 α -HSDH enzyme, it is the β -face of the substrate that must point toward the nicotinamide. This geometry has indeed been observed in the crystal structures of the product/substrate complexes of 7 α -HSDH¹³ [Fig. 2(C)].

The structural comparison shows that both dehydrogenases share a very similar NAD(P)H-binding site as well as a common architecture for the substrate-activating residues [Fig. 2(C) and Supporting Information Fig. S2]. Thus, in both enzymes the substrate will bind with the 7-hydroxyl group pointing toward the conserved Ser–Tyr catalytic dyad and the C7 carbon in direct contact with the C4 carbon of the cofactor nicotinamide, which is predicted to feature an identical binding mode in the

two enzymes. Given these observations, how can the two dehydrogenases perform the same reaction, but with opposite stereoselectivity? Apparently, 7 α -HSDH and 7 β -HSDH must bind the substrates in opposite orientation. “Flipped” substrate orientations in the two enzymes would enable them to retain the interactions of the hydroxyl group with the catalytic Ser–Tyr dyad and, at the same time, orient the substrates with their respective 7 α - or 7 β -hydrogen pointing toward the nicotinamide ring as necessary for stereoselective dehydrogenation.

Given the bulkiness of these dehydrogenase substrates, an inversion of substrate orientation can only be achieved through a substantial reshaping of the active site cleft. With a 17% sequence identity, the differences between the substrate sites of 7 β -HSDH and 7 α -HSDH inevitably result from the global and correlated effects of the various amino acid replacements and associated conformational changes. Nevertheless, the observation that 7 α -HSDH lacks the C-terminal helix present in 7 β -HSDH is striking in that it drastically affects the substrate-binding site. This is in sharp contrast with the high degree of conservation of the NAD(P)H-binding region (notwithstanding the difference in the 2'-phosphate niche) as well as of the residues directly involved in the hydride transfer step. Along this line, although with the caveat that 7 β -HSDH crystallized in the apo-form, a simple modeling experiment interestingly suggests that ursodeoxycholic acid [Fig. 1(B)] can be docked into the substrate site of 7 β -HSDH with the orientation expected for transfer of the 7 α -hydrogen without any apparent clash with the surrounding protein residues (Supporting Information Fig. S2).

CONCLUSIONS

The 1.6 Å crystal structure of *C. aerofaciens* 7 β -HSDH and its comparison with *E. coli* 7 α -HSDH illustrate a remarkable case of inversion of stereoselectivity in structurally and evolutionarily related enzymes. This appears to be achieved by a mix of conservation and divergence. The binding of the redox cofactor is conserved together with the constellation of catalytic residues (including a mobile substrate loop) that promote hydride transfer leading to substrate dehydrogenation. Conversely, the substrate-binding site is substantially different, above all for the presence of C-terminal secondary structure elements that diversifies the two enzymes. These variations are predicted to underlie opposed orientations in the substrate binding modes leading to opposite stereoselectivity.

ACKNOWLEDGMENTS

This work was supported by the University of Pavia (Fondi d'Ateneo). Federico Forneris is supported by a Career Development Award from the Armenise-Harvard Foundation and by the “Rita Levi-Montalcini” award

from the Italian Ministry of University and Education (MIUR).

REFERENCES

1. Kavanagh KL, Jörnvall H, Persson B, Oppermann U. The SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell Mol Life Sci* 2008; 65: 3895–3906.
2. Fossati E, Riva S. Stereoselective modifications of polyhydroxylated steroids. In: Patel RN, editor. *Biocatalysis in the pharmaceutical and biotechnology industries*. Boca Raton, Florida: CRC Press; 2006. pp. 591–604.
3. Liu L, Aigner A, Schmid RD. Identification, cloning, heterologous expression, and characterization of a NADPH-dependent 7 β -hydroxysteroid dehydrogenase from *Collinsella aerofaciens*. *Appl Microbiol Biotechnol* 2011; 90:127–135.
4. Ferrandi EE, Bertolesi GM, Polentini F, Negri A, Riva S, Monti D. In search of sustainable chemical processes: cloning, recombinant expression, and functional characterization of the 7 α - and 7 β -hydroxysteroid dehydrogenases from *Clostridium absonum*. *Appl Microbiol Biotechnol* 2012; 95:1221–1233.
5. Lee J-Y, Arai H, Nakamura Y, Fukiya S, Wada M, Yokota A. Contribution of the 7 β -hydroxysteroid dehydrogenase from *Ruminococcus gnavus* N53 to ursodeoxycholic acid formation in the human colon. *J Lipid Res* 2013; 54:3062–3069.
6. Bovara R, Canzi E, Carrea G, Pilotti A, Riva S. Enzymatic α/β inversion of the C-7-hydroxyl of steroids. *J Org Chem* 1993; 58: 499–501.
7. Carrea G, Pilotti A, Riva S, Canzi E, Ferrari A. Enzymatic synthesis of 12-ketoursodeoxycholic acid from dehydrocholic acid in a membrane reactor. *Biotechnol Lett* 1992; 14:1131–1135.
8. Crosignani A, Setchell KDR, Invernizzi P, Larghi A, Rodrigues CMP, Podda M. Clinical pharmacokinetics of therapeutic bile acids. *Clin Pharmacokinet* 1996; 30:333–358.
9. Monti D, Ferrandi EF, Zanellato I, Hua L, Polentini F, Carrea G, Riva S. One-pot multienzymatic synthesis of 12-ketoursodeoxycholic acid: subtle cofactor specificities rule the reaction equilibria of five biocatalysts working in a row. *Adv Synth Catal* 2009; 351:1303–1311.
10. Eggert T, Bakonyi D, Hummel W. Enzymatic routes for the synthesis of ursodeoxycholic acid. *J Biotechnol* 2014; 191:11–21.
11. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 2011; D67: 235–242.
12. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr* 2007; 40:658–674.
13. Tanaka N, Nonaka T, Tanabe T, Yoshimoto T, Tsuru D, Mitsui Y. Crystal structures of the binary and ternary complexes of 7 α -hydroxysteroid dehydrogenase from *Escherichia coli*. *Biochemistry* 1996; 35:7715–7730.
14. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. Phenix: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr* 2010; D66:213–221.
15. Holm L, Rosenström P. Dali server: conservation mapping in 3D. *Nucleic Acids Res* 2010; 38(Web Server issue):W545–W549.

Mutagenesis

7 β -HSDH $_{\Delta 250-263}$ and 7 β -HSDH $_{\Delta 237-263}$ mutants were generated by PCR amplification of the fragments (1-750 nt) or (1-711 nt) of the gene coding for the *C. aerofaciens* 7 β -HSDH. The PCR was carried out in a 50 μ L total volume by using pET24b-7 β -HSDH as template, Xtra-Taq Pol kit (Genespin, Milan, Italy) and the following primers: 5'-GAAGGAGATATACATATGAACCTGAGGGAGAAGTACGGTG-3' (F1) and 5'-GTGATGGTGGTATGATGGGTGGTGGTTTGCCTTCCAGTC-3' (R1) for the 7 β -HSDH $_{\Delta 250-263}$ mutant, and F1 and 5'-GTGATGGTGGTATGATGGTAGCGGATGACTCGTCCTC-3' (R2) for the 7 β -HSDH $_{\Delta 237-263}$ mutant. Primers include 18 nt of overlap with the ends of the pETite vector for the subsequent cloning into the pETite vector (Lucigen) and allow gene cloning in frame with the C-terminal His-tag. PCR conditions were as follows: 95 °C for 2 min, followed by 40 cycles at 94 °C for 30 s, 58 °C for 30 s, 72 °C for 1 min, and then 72 °C for 10 min. Amplified products were purified from the agarose gel using the Wizard SV Gel & PCR CleanUp System (Promega, Madison, WI, USA). Purified PCR products were cloned into the pETite vector and transformed in *E. coli* 10G chemically competent cells using the Expresso Rhamnose Cloning and Protein Expression System (Lucigen) according to the manufacturer's instructions. The resulting plasmids pETite-7 β -HSDH $_{\Delta 250-263}$ and pETite-7 β -HSDH $_{\Delta 237-263}$ were sequenced on both strands to verify the cloned PCR amplicon (BioFab, Rome, Italy). Mutants were expressed and purified as described for the wild type 7 β -HSDH (see above).

Table S1. Residues involved in NAD(P)⁺ binding

7 α -HSDH (NAD ⁺)	7 β -HSDH (NADP ⁺)	Note ^a
Asp42	Gly39	Side chain H-bonded to adenine ribose; in 7 α -HSDH Arg40-Arg41 are predicted to form the binding site for the 2'-phosphate of NADPH
Ile43	Arg40	Main chain H-bonded to adenine
Cys67	Ala65	Main chain H-bonded to adenine
Asp68	Asp66	Side chain H-bonded to adenine
Ile69	Phe67	Main chain H-bonded to adenine
Asn95	Val91	Interaction with nicotinamide
Tyr159	Tyr156	Side chain main interacts with nicotinamide
Lys163	Lys160	Main chain H-bonded to nicotinamide ribose
Ile192	Thr189	Main chain H-bonded to nicotinamide
Thr194	Thr191	Side chain H-bonded to nicotinamide ring

^a Interactions observed in the crystal structure of 7 α -HSDH bound to NAD⁺ (PDB entry 1FMC).

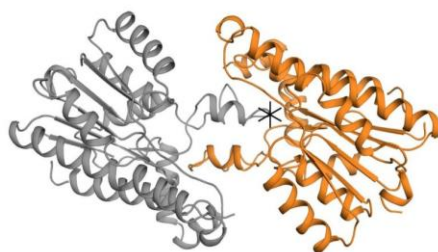


Figure S1. Packing of crystallographically-related 7 β -HSDH subunits belonging to different dimers. The substrate specificity loop (Fig. 2 in the main text) is extensively involved in intermolecular crystal contacts, helping crystallization but also making binding of substrate and cofactor impossible due to steric occlusion of the active site (black asterisk).

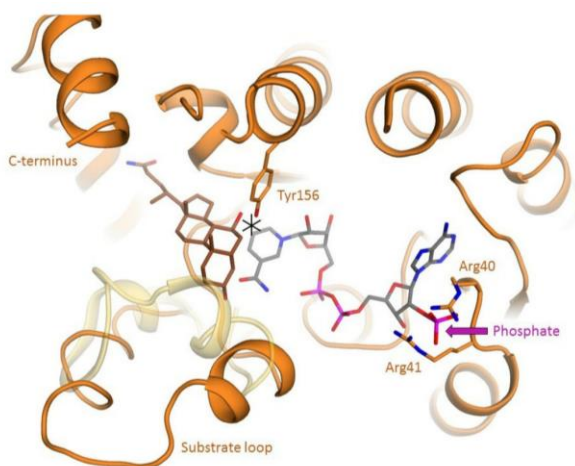
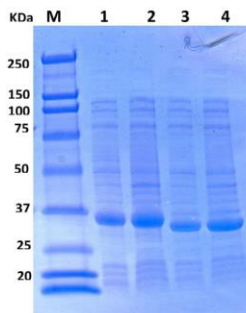


Figure S2. Working model for the binding of ursodeoxycholate (brown; Fig.1B of main text) in 7 β -HSDH (orange) and proposed contraction of the loop upon binding (transparent yellow), based on the structure of 7 α -HSDH holo-form (Fig. 2 of main text). The substrate has been manually located in this position by avoiding clash contacts and following the reaction stereo-chemistry (see main text). The black asterisk indicates the 7 β CH-OH group which is the site of oxidation. The 7-hydroxyl group of ursodeoxycholate points towards Tyr156 side chain whereas the C7 α -hydrogen is oriented towards the nicotinamide (gray) C4 atom. Model also displays the fitting of NADP 2'-phosphate in proximity of Arg40 and Arg41.

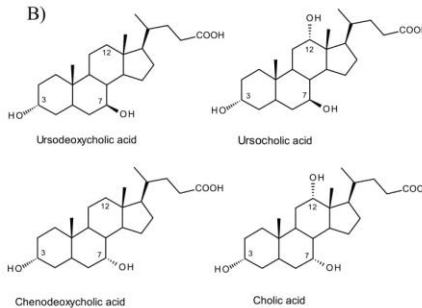
A)



MW= Molecular weight markers.

1= *E. coli* soluble extract expressing 7β-HSDH Δ₂₅₀₋₂₆₃2= *E. coli* total extract expressing 7β-HSDH Δ₂₅₀₋₂₆₃3= *E. coli* soluble extract expressing 7β-HSDH Δ₂₃₇₋₂₆₃4= *E. coli* total extract expressing 7β-HSDH Δ₂₃₇₋₂₆₃

B)



Substrate	Specific activity (U/mg)			
	Wild-type 7β-HSDH	7β-HSDH Δ ₂₅₀₋₂₆₃	7β-HSDH Δ ₂₃₇₋₂₆₃	Wild-type 7α-HSDH
Ursodeoxycholic acid	17.48	0.010	n.d.	n.d.
Ursocholic acid	3.7	0.012	n.d.	n.d.
Chenodeoxycholic acid	n.d.	n.d.	n.d.	47.5
Cholic acid	n.d.	n.d.	n.d.	9.6

Figure S3. A) Overexpression of 7β-HSDH mutants in *E. coli* cells evaluated by SDS-gel electrophoresis and B) comparison of specific activity of 7β-HSDH deletion mutants towards different bile acids with the respect to wild-type 7β-HSDH. Reference activity data for *E. coli* 7α-HSDH for its respective natural substrates chenodeoxycholic acid and cholic acid are shown. n.d.: not detected under the assay conditions.

List of original manuscripts:

- Fürst MJLJ, Savino S, Dudek HM, Gómez Castellanos JR, Gutiérrez de Souza C, Rovida S, Fraaije MW, and Mattevi A (2017) Polycyclic Ketone Monooxygenase from the Thermophilic Fungus *Thermothelomyces thermophila*: A Structurally Distinct Biocatalyst for Bulky Substrates. Journal of the American Chemical Society, 139 (2), 627-630
Fürst MJLJ and Savino S equally contributed

- Savino S, Ferrandi EE, Forneris F, Rovida S, Riva S, Monti D and Mattevi A (2016) Structural and biochemical insights into 7 β -hydroxysteroid dehydrogenase stereoselectivity. Proteins, 84: 859–865

- Dzurová L, Forneris F, Savino S, Galuszka P, Vrabka J and Frébort I (2015), The threedimensional structure of “Lonely Guy” from *Claviceps purpurea* provides insights into the phosphoribohydrolase function of Rossmann fold-containing lysine decarboxylase-like proteins. Proteins, 83: 1539–1546 (Note: my contribution to this work was marginal, therefore I have not described it in the thesis and I have not included the pdf)

Presentations at meetings and conferences:

- Fine understanding of activated sugars biocatalysis by AXS. ESIB 2017, Graz, Austria. S. Savino. Oral presentation.

- Exploring the reaction mechanisms of UDP-D-apiose/UDP-D-xylose synthases. Biotrans 2017, Budapest, Hungary. A. Dennig, S. Savino, T. Eixelsberger, F. de Giorgi, H. Weber, A. Mattevi, B. Nidetzky. Poster session.

- PockeMO and TmCHMO: new insights into biocatalytic Baeyer-Villiger reactivity. ESIB 2016, Graz, Austria. S. Savino, J.R. Gómez Castellanos, E. Romero, M. Fürst, H. Dudek, M. Fraaije, A. Mattevi. Poster session.

- Exploitation of UDP-D-apiose/UDP-D-xylose synthase reactivity as source of different nucleotide sugars. ESIB 2016, Graz, Austria. S. Savino, A. Dennig, T. Eixelsberger, B. Nidetzky, A. Mattevi. Poster session.

- Structural and biochemical insights into 7 β -hydroxysteroid dehydrogenase stereoselectivity. OxiZymes 2016, Wageningen, The Netherlands. S. Savino, E.E. Ferrandi, F. Forneris, S. Rovida, S. Riva, D. Monti A. Mattevi. Poster session.