The central dogma of molecular biology is that the sequence of the nucleic acids in (the coding sequence of) a gene specifies the structure and function of a protein, and that mutations in the nucleotide sequence may result in the evolution of protein variants with changed properties. An important and frequently observed feature of protein evolution is gene duplication, followed by divergence and/or domain fusion events. Following duplication, the original gene product retains its original function, but the duplicated gene product has no functional constraints and is allowed to evolve and develop new and different functions. Some homologous proteins may even diverge beyond the point of observable sequence similarity. In contrast, protein structures are much better conserved throughout evolution than the protein's sequence. Hence, 3D structures allow for easier recognition of evolutionary relationships. Thus, to increase further understanding of the nature of evolutionary processes of proteins the combination of structural and sequence data will be helpful.

This thesis discusses the crystal structures of six different enzymes. These six enzymes belong to different protein families and utilize different catalytic mechanisms. Understanding their 3D structures is essential for understanding the molecular determinants of their function and may shed a light on how they have evolved.
MANNURONAN C-5-EPIMERASE

Alginate is synthesized by brown algae (*Phaeophyceae*, seaweed, kelp) and by *Azotobacter* and *Pseudomonas* species (Tondervik *et al.*, 2013), and is a major constituent of mature biofilms produced by the latter bacteria. Alginate is industrially harvested all over the world. Commercially, alginates are interesting because of their gelling and viscosifying properties. Alginate is capable of absorbing 200-300 times its own weight in water, which makes it useful as an additive in products such as slimming aids, and for thickening drinks and cosmetics. It is also used for the water- and fireproofing of fabrics, for surface sizing in the paper industry and for textile printing. The most important application is probably in ice cream production, where alginates are used to prevent crystallization and shrinkage and to give a homogeneous product (Clementi, 1997).

Alginate is a family of linear copolymers of (1→4)-linked β-D-mannuronic acid (M) and its C-5 epimer α-L-guluronic acid (G). The polymer is first produced as polymannuronic acid and the guluronic acid residues are then introduced at the polymer level as blocks by mannuronan C-5-epimerases. In *Azotobacter vinelandii* a family of 7 secreted and calcium-dependent, mannuronan C-5 epimerases (AlgE1-7) has been identified (Aachmann *et al.*, 2006). These enzymes are highly homologous and consist of one or two catalytic A-modules and one to seven regulatory R-modules, of which at least one is needed for full activity (Ertesvåg *et al.*, 1999). The mannuronan epimerases probably evolved through divergent evolution so that different epimerases introduce different G distribution patterns. The smallest of the AlgE proteins, AlgeE4, comprises one A-followed by one R-module and strictly forms alternating MG sequences (MG-blocks).

The 2.1-Å resolution 3D structure of the catalytic A-module of the *Azotobacter vinelandii* mannuronan C-5-epimerase AlgE4 is described in chapter 2. AlgE4A folds into a right-handed parallel β-helix structure, originally found in pectate lyase C and subsequently in several other polysaccharide lyases and hydrolases. The AlgE4 β-helix is composed of four parallel β-sheets, comprising 12 complete turns, and has an amphipathic α-helix near the N terminus. The catalytic site is positioned in a positively charged cleft formed by loops extending from the surface encompassing Asp$^{152}$, an amino acid previously shown to be important for the reaction. Site-directed mutagenesis further implicates Tyr$^{149}$, His$^{154}$, and Asp$^{178}$ as being essential for activity. Tyr$^{149}$ probably acts as the proton acceptor, whereas His$^{154}$ is the proton donor in the epimerization reaction.

Interestingly, a striking agreement was noted in the spatial arrangement of these four amino acid residues in the center of the substrate binding clefts of alginate lyases from different polysaccharide lyase families (PL). AlgE4A displays the same spatial arrangement of Asp$^{152}$, His$^{154}$, Lys$^{117}$, and Tyr$^{149}$ in its active site as the equivalent Gln/Asn, His, Arg, and Tyr in the alginate lyases from family PL-7 (β-jelly roll), PL-5, ((α/α)$_6$ barrel) and PL-18 (β-jelly roll). This suggests that the catalytic mechanisms of
the four lyase families are the same. Recently, 3D structures of the periplasmic alginate epimerase form *Pseudomonas aeruginosa* (β-helix, PDB ID 4NK6 (Wolfram et al., 2014) and the exotype PL17 alginate lyase from *Saccharophagus degredans* ((α/α)_6 barrel + β-sandwich, PDB ID 4NEI (Park et al., 2014)) were published. Both enzymes also show a similar conservation of the active site residues. This is a clear example of evolutionary convergent active sites, which evolved independently at least three times.

The structure of the R-module of Alge4, a β-roll structure, was solved by NMR (Aachmann et al., 2006). The complete AlgE4 protein might form a single elongated structure with a long, positively charged patch for substrate binding (see the Figure below), in which the R-module appears as a natural extension of the A-module (Aachmann et al., 2006). However, the crystallization of the complete AlgE4 protein failed so far. Alternatively, electron microscopy or small angle X-ray scattering could validate this proposal. The research on alginates and mannuronan C-5-epimerases is continued at NTNU in Trondheim, Norway (Tondervik et al., 2013).

\[ \alpha-1,4 \text{ GLUCAN LYASE} \]

\( \alpha-1,4 \) Glucan lyase (GLase) was first purified and characterized by Shukun Yu (Yu et al., 1993) while purifying amylases from the red seaweed *Gracilariopsis lemaneiformis*. Subsequently, other GLases were isolated from the fungi *Morchella costata*, *M. vulgaris* and *Peziza ostracoderma* (Bojsen et al., 1999b).

The enzyme degrades α-1,4-glucans to yield the keto-monosaccharide sugar 1,5-D-anhydrofructose (AF). This is the first reaction step in the anhydrofructose pathway, which is believed to be an alternative glycogen catabolic pathway of α-glucans, such as glycogen and starch (Yu et al., 2006). The physiological importance of the AF pathway in red algae is likely related to its involvement in stress responses associated with carbon starvation, removal of reactive oxygen species, and defense against infections by other organisms (Yu et al., 2006). Anhydrofructose is a promising compound for application as a modifier of gluten and gliadin proteins to enhance their solubility, and as a calorie-free
sugar. In addition, AF has an antibrowning effect in green tea and is able to slow turbidity development in black currant wine (Yuan et al., 2005).

![Diagram of Glycogen metabolism]

Scheme 1. The anhydrofructose pathway of glycogen catabolism (Yu, 2008).

The crystal structure of GLase, described in chapter 3, shows that similar to glycoside hydrolase family 31 (GH31) hydrolases, the enzyme contains a (β/α)8-barrel catalytic domain with B and B’ subdomains, an N-terminal domain N, and two C-terminal
domains, C and D. Domain N showed binding of an oligosaccharide, explaining its starch binding capacity. Furthermore, the 3D structures of GLase with inhibitors bound in the active site suggest that, as in GH31 hydrolases, the aspartic acid residues Asp553 and Asp665 are the catalytic nucleophile and acid, respectively. The active site of GLase superimposes well on those of other GH31 members. Our results show that the first step of the reaction, the formation of a covalent enzyme-substrate intermediate, is conserved in both lyases and hydrolases. Yet, the second step is different, since GLases catalyze a lyase reaction, and not a hydrolysis reaction as done by the GH31 glucan hydrolases.

It appears that a subtle change, a Val to Glu substitution at position 556, promotes the lyase reaction. In GH31 hydrolases the strictly conserved Arg649 has a salt bridge interaction with the conserved Glu556, but in the GH31 lyases the equivalent residue is a Val or Thr. As a result, the arginine side chain in GLase can now take up a position to pick up the proton abstracted from the sugar's C2 carbon atom by the correctly positioned carbonyl oxygen atom of Asp553 and relay it to the solvent. This suggests that residue 556 (being either a Glu or a Val) is a major determinant of the hydrolase vs lyase reaction specificity of GH31 enzymes. As a consequence, in GH31 lyases the Asp553 nucleophile has acquired a dual function, acting also as the base that abstracts the proton from the C2 atom of the -1 glucose residue.

To support the putative dual role of the nucleophile we made a quadruple mutant, including the Glu323(556) to Val mutation, in the active site of the homologous GH31 α-glucosidase from Sulfolobus solfataricus. The mutant protein showed detectable lyase activity and a complete loss of glucosidase activity. Thus, indeed the E323V mutation in the glucosidase shifts the reaction from hydrolase toward lyase activity. This change in activity demonstrates the importance of position 323(556) for hydrolase/lyase activity in GH31, likely by modulating the conformation of the fully conserved arginine in the active site. Evolving the glucosidase into a genuine, full-activity lyase with a $k_{cat}$ similar to that of GLase will require further mutational studies.

Likely, the GH31 enzymes appear to have evolved from a common ancestor comprising a domain N, a catalytic domain A, domain C and D. and have subsequently diverged. Domains B and B', the most diverging domain among the GH31 members, make up the entrance to the active site and are important for substrate specificity. Whereas domain N contributes residues essential for oligosaccharide binding and maintenance of the active site architecture, the roles of the C-terminal domains are not clear (Ernst et al., 2006). It is surprising that with only single amino acid substitution a major change in reaction specificity can be achieved.
PYRROLOQUINOLINE QUINONE DEPENDENT ALCOHOL DEHYDROGENASE

Pyrroloquinoline quinone dependent alcohol dehydrogenase (PQQ-ADH) from *Pseudogluconobacter saccharoketogenes* is a versatile dehydrogenase oxidizing various alcohols and sugars. The enzyme requires pyrroloquinoline quinone (PQQ) as a cofactor. As detailed in chapter 4 the structure of the PQQ-ADH consists of an eight-bladed β-propeller fold with the PQQ located in the center of the molecule near the 8-fold pseudosymmetry axis. The enzyme binds one PQQ and one Ca\(^{2+}\) ion like other PQQ-dependent methanol/ethanol dehydrogenases. However, three of the four ligands of the Ca\(^{2+}\) ion differ from those of related dehydrogenases and they come from different parts of the polypeptide chain. The only conserved amino acid residue is Asp333. This residue is proposed to be the catalytic base, which is supported by the bound substrate mimic polyethylene glycol. The active site is open to the solvent and easily accessible. This explains why PQQ-ADH can oxidize a broad range of substrates, extending for example from as small as methanol to as large as cyclodextrins.

One of the closest homologues of PQQ-ADH is a type II Quinohemoprotein alcohol dehydrogenase QH-ADH (Oubrie *et al.*, 2002), of which the structure was also determined in the Laboratory of Biophysical Chemistry. The largest difference between the two structures is that QH-ADH contains its own cytochrome c domain necessary for transfer of the electrons generated during the reaction. The relatively open active site of PQQ-ADH suggests that it can bind a cytochrome acceptor in a location that is similar to that of the cytochrome c domain of QH-ADH dehydrogenase.

Another major difference of PQQ-ADH compared to type I Quinohemoprotein methanol dehydrogenases and type II enzymes is that PQQ-ADH lacks the vicinal disulfide bridge, implicated in electron transfer. The different calcium ligands, the absence of a cytochrome c domain, and the absence of the vicinal disulfide bond together create the open active site characteristic of this enzyme. Because of these major differences PQQ-ADH may be regarded as a phylogenetic outgroup in the tree of PQQ-dependent alcohol dehydrogenases.

Probably the eight-bladed β-propeller PQQ dehydrogenases share a common ancestor. Over time, PQQ-ADH has accumulated differences compared to the other structurally characterized types, and has *diverged* into an enzyme with a very broad substrate specificity suitable for many applications.
Plant cell wall polysaccharides are the most abundant organic compounds found in nature. They make up 90% of the plant cell wall and can be divided into three groups: cellulose, hemicellulose, and pectin (de Vries et al., 2001). Pectins are heteropolysaccharides and consist of a backbone of α-1,4-linked D-galacturonic acid residues (polygalacturonan) decorated with various side chains or modifications. The “smooth” regions consist of polygalacturonan, of which the galacturonic acid residues may be acetylated at O-2 or O-3 or methylated at O-6. The “hairy” regions consist of xylogalacturonan (XGA), a D-xylose-substituted galacturonan backbone (de Vries et al., 2001).

As XGA contributes to membrane fouling in the ultra-filtration process for fruit juice clarification, it is crucial that XGA is completely degraded during enzymatic treatment of the fruit pulp. Xylogalacturonan hydrolase (XghA) from *Aspergillus tubingensis* has been found to be an efficacious enzyme to degrade XGA (Zandleven et al., 2006, van der Vlugt-Bergmans et al., 2000). The enzyme belongs to glycoside hydrolase family 28 (GH28).

In chapter 5 the crystal structure of endo-xylogalacturonan hydrolase (XghA) is described. XghA folds into a right-handed parallel β-helical structure comprising 10 complete turns with an active site that is located in an open groove/cleft on the surface of the protein. The high degree of structural conservation of the active site and catalytic apparatus, which is shared with polygalacturonases, indicates that cleavage of the substrate proceeds in essentially the same way as found for the other GH28 enzymes. However, the substrate-binding cleft is much wider than in other polygalacturonases, in agreement with its specificity for xylosylated substrates. The most extensive interactions appear to occur at subsite +2. Validation of the assigned subsites for XghA binding is a subject for future mutational experiments.

XghA has probably evolved through gene duplication of an ancestral GH28 enzyme followed by divergent evolution. GH28 enzymes contain a β-helix structure observed in many carbohydrate-binding proteins and sugar hydrolases. The need for an extensive repertoire of enzymes in *Aspergillus* that can efficiently degrade pectin is illustrated by large amount of GH28 enzymes that have evolved. XghA plays a unique role in GH28 as it specifically degrades xylogalacturonan.

**CARBOXYLESTERASES FROM BACILLUS SUBTILIS**

Although the crystallization of naproxen esterase (NP) from *Bacillus subtilis* was already reported in 1993 (van der Laan et al., 1993), it took some 20 years to solve its structure (chapter 6). In 1999, diffraction data from a native NP crystal were collected by M.
Nardini. However, no heavy atom derivatives for the enzyme could be found, and in the absence of close structural homologues, structure determination by molecular replacement was neither feasible. Only after an ensemble of different search models with less than 20% identity were used a partial structure was obtained consisting mainly of the central \( \beta \)-sheet structure of the \( \alpha/\beta \) hydrolase fold. The structure determination of carboxylesterase CesB from *Bacillus subtilis* (chapter 6) was straightforward using NP as search model.

NP and CesB show the typical \( \alpha/\beta \) hydrolase structure that contains a catalytic triad comprising the active site. The structures consist of a core of 8 mostly parallel \( \beta \)-strands that are surrounded on both sides by \( \alpha \)-helices, with an extra N-terminal helix stabilizing the cap subdomain. The catalytic triad is composed of Ser\(^{130} \), Glu\(^{245} \), and His\(^{274} \). The access/exit route for substrate and product to/from the active site and the solvent is not obvious from the structures. Flexibility of the cap domain might provide such an access/exit route. A Molecular Dynamics study may support this proposal.

Surprisingly, the search models that were used for molecular replacement were not structures of esterases but of C-C bond breaking enzymes from the meta-cleavage product (MCP) hydrolase family. Despite their different activity, MCP hydrolases are most similar to the carboxylesterases. Recently, enzymes with dual esterase / MCP hydrolase activity were found (Alcaide *et al.*, 2013), indicating that these enzymes show substrate promiscuity. It would be an interesting experiment to establish whether NP and CesB have promiscuous C-C bond cleaving activity.

**CONCLUSION**

In conclusion, the research described in this thesis has shed a light on the molecular details of the various evolutionary processes leading to divergence of reaction specificity of glucan lyase, divergence of substrate specificity of xylogalacturonan hydrolase and of PQQ-dependent alcohol dehydrogenase, divergence of enantioselectivity in carboxylesterases and to the convergent reaction specificity of mannuronan epimerases. How subtle changes in amino acid sequence lead to important changes in functionality is a continuous source of wonder that motivates crystallographers, and attests that evolution of a new function is still hard to predict.