NPR

Cite this: *Nat. Prod. Rep.*, 2012, **29**, 1176

www.rsc.org/npr

REVIEW

The impact of structural biology on alkaloid biosynthesis research†

Santosh Panjikar,*[ab] Joachim Stoeckigt,[c] Sarah O'Connor[d] and Heribert Warzecha[e]

Received 30th April 2012
DOI: 10.1039/c2np20057k

Covering: 1997 to 2012

In the recent past, macromolecular crystallography has gone through substantial methodological and technological development. The purpose of this review is to provide a general overview of structural biology and its impact on enzyme structure/function analysis and illustrate how it is modifying the focus of research relevant to alkaloid biosynthesis.

[a]*Australian Synchrotron, 800 Blackburn Road, Clayton, Victoria 3168, Australia. E-mail: santosh.panjikar@synchrotron.org.au*
[b]*Department of Biochemistry and Molecular Biology, Monash University, Victoria 3800, Australia*
[c]*Institute of Materia Medica, College of Pharmaceutical Sciences, Zhejiang University, 388 Yu Hang Tang Road, Hangzhou, 310058, P. R. China. E-mail: joesto2000@yahoo.com*
[d]*John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK. E-mail: Sarah.O'Connor@jic.ac.uk*
[e]*Technische Universität Darmstadt, Plant Biotechnology and Metabolic Engineering, Schnittspahnstrasse 3-5, 64287 Darmstadt, Germany. E-mail: warzecha@bio.tu-darmstadt.de*
† This paper is part of an *NPR* themed issue on Structural Aspects of Biosynthesis.

1 Introduction

Alkaloids are a therapeutically valuable class of natural products that exhibit an impressive variety of complex and diverse carbon skeletons[1] and most of these secondary compounds are produced by higher plants. Characterization of medically important molecules and their associated structurally complex scaffolds has always been a great challenge for chemists, biochemists and biologists who have been seeking to explore

alkaloids synthetic and biosynthetic routes.[2] Early attempts to provide a basic understanding of their biosynthetic pathways were undertaken in the 1960's and 70's by feeding potential precursors labeled with radioactive or stable isotopes *in vivo*, a technique subsequently largely replaced by cell-free systems[3,4] and isolated single enzymes[2,5] obtained from plant cell-suspension cultures.[6] In fact, introducing enzymology into alkaloid biosynthetic research opened the way to the detailed elucidation of multi-step pathways of monoterpenoid indole alkaloids,[7] such as ajmalicine from *Catharanthus roseus* G. Don[5] and ajmaline from the Indian medicinal plant *Rauvolfia serpentina* Benth. ex Kurz[8,9] or of isoquinoline-type alkaloids,[2,10] such as morphine, berberine or the benzophenanthridines sanguinarine/macarpine.

Since the end of the 1980's, knowledge of these enzymes and their partial amino acid sequences has facilitated the application of "reverse genetics approaches", resulting in highly efficient recombinant enzyme production in bacteria and other organisms[11] and allowing biochemical determination of enzyme properties.

The large quantities of pure enzymes that could be produced due to this progress paved the way for crystallization and elucidation of the enzymes' three-dimensional (3D) architecture.[12,13] 3D structures then provided the basis for understanding detailed enzyme mechanisms at the molecular level, leading to rational site-directed enzyme mutations, metabolic engineering of biosynthetic pathways, development of enzyme-mediated "white chemistry" and finally to the construction of novel alkaloid libraries – topics which are discussed in this review article.

2 Introduction of current methods in structural biology

Macromolecular crystallography is a popular technique in structural biology for elucidating the structure of macromolecules.

Santosh Panjikar

Santosh Panjikar earned a PhD in Biotechnology from Friedrich-Schiller University (Jena, Germany) in 2001. He held a postdoctoral, Senior Technical Officer and Staff Scientist appointments at the EMBL Hamburg Outstation, Germany with Dr Paul A. Tucker between 2001 and 2011. Currently he is working as a Scientist at the Australian Synchrotron and he is an Adjunct Research Fellow at the Department of Biological Science and Biochemistry, Monash University, Australia. His research interests focus on structural biology, method developments in structural biology, synchrotron instrumentation and software development in macromolecular crystallography.

Sarah O'Connor

Sarah O'Connor received her BS in chemistry from the University of Chicago and a PhD in organic chemistry from MIT and Caltech with Professor Barbara Imperiali. She held a postdoctoral appointment as an Irving Sigal postdoctoral fellow at Harvard Medical School with Professor Christopher T. Walsh. She held assistant and Associate professor positions between 2003 and 2011 in the chemistry department at MIT. Currently, she is a project leader at the John Innes Centre, UK and Professor at the University of East Anglia, UK. Her research interests include understanding natural product biosynthetic pathways and enzyme mechanisms.

Joachim Stoeckigt

Joachim Stoeckigt received a PhD in organic chemistry from Muenster University (Germany) with Professor Burchard Franck. He has worked with Professor Meinhart H. Zenk at the Faculty of Biology (Bochum University, Germany) and the Faculty of Chemistry and Pharmacy (Munich University, Germany) and is Professor at the College of Pharmaceutical Sciences, Zhejiang University (Hangzhou, China) and retired from the Institute of Pharmacy (Mainz University, Germany). His research interests include natural products biosynthesis (phytochemistry, enzymology, molecular and structural biology).

Heribert Warzecha

Heribert Warzecha holds a PhD in pharmaceutical biology and is heading the research group Plant Biotechnology and Metabolic Engineering at the Technische Universität Darmstadt, Germany. He is currently chairing a pan-European research network (COST) devoted to plant metabolic engineering (www.plantengine.eu). His main research focus is the production of pharmaceuticals in plants and the engineering metabolic pathways in plants.

The establishment of the Protein Data Bank (PDB)[14] as the single repository for crystal structure and structure determined by other techniques (*i.e.* NMR, EM) provided a unique resource for the scientific community. The pace of structure determination has accelerated in the last decade due to the launch of structural genomics projects around the world and due to the development of powerful new algorithms and computer programs for diffraction data collection, structure solution, refinement and presentation. This also had an immense impact on alkaloid biosynthesis research. In the last couple of years, a number of crystal structures of enzymes and enzyme-complexes from alkaloid biosynthetic pathways have been determined. This has helped tremendously in understanding the enzyme mechanisms and substrate specificities, as well as paving the way for a rational design of novel alkaloids and metabolic engineering. Here, some of the key steps in macromolecular crystallography are outlined with some details.

2.1 Cloning and recombinant protein production

Obtaining highly purified protein is a crucial prerequisite for any protein crystallography project. That implies that the protein can be obtained from a heterologous source since endogenous sources, like whole plants or tissue cultures thereof, usually do not yield sufficient quantities of protein. Once the coding sequence of a protein is identified either the cDNA can be utilized or, more common today, a synthetic gene which is codon-adapted for the expression host, is used. Overproduction of the recombinant protein is easily checked, by preparing crude extracts of total protein from the expressing cells at different time intervals after induction of expression, and analyzing them by SDS PAGE. The band corresponding to the induced protein increases in intensity after induction, showing that the amounts of the other proteins detected on this gel are not altered during the experiment. Regardless of the expression method, to yield sufficiently pure protein which is devoid of most host proteins without the need for multiple purification steps requires utilization of purification tags. By far the most common fusion is the hexa histidine tag for purification on metal-chelate resins. This tag provides a substantial purification handle while being relatively unobtrusive as a fusion partner. Beyond purification, translational fusions often provide means to enhance expression. The larger fusion tags, such as thioredoxin and GST, often are superior in this respect.

Many more options exist for this purpose and a comprehensive review is beyond the scope of this article but can be found here.[15,16] The large amounts of protein which usually can be obtained from *E. coli* enable the screening of several different crystallization conditions.

2.2 Protein crystallization methods and automation

Production of protein crystals suitable for structural studies poses one of the major bottlenecks in the entire process of crystallography. The success of protein crystallization depends on the availability of homogenous, ideally monodisperse solutions of protein molecules with a uniform, rigid three-dimensional fold. Finding crystallization conditions to yield diffraction-quality crystals can be sometime very challenging. Protein crystallization occurs when the concentration of protein in solution is greater than its limit of solubility so that the protein solution becomes supersaturated. In order to crystallize a protein it undergoes slow precipitation from an aqueous solution. As a result, individual protein molecules align themselves in a repeating series of "unit cells" by adopting a uniform orientation. One unavoidable aspect of crystallizing a newly expressed protein is the need to carry out a large number of experiments in order to find suitable conditions in which the protein crystallizes. It can be extremely tedious and time consuming setting up a broad array of different crystallization experiments manually. Currently, there are no systematic methods to ensure that ordered three-dimensional crystals will be obtained in a particular solution based on the protein behaviour.[17] With the advent of high-throughput liquid handling and crystallization systems, it is relatively easy and less time consuming to prepare a thousand or more crystallization experiments in which crystallization parameters, such as the ionic strength, pH, protein and precipitant concentration and temperature, are varied systematically.

Methods used for crystallization include vapor diffusion, batch crystallization, dialysis, seeding, free interface diffusion and temperature induced crystallization. The most popular method for setting up crystallization experiments is vapor diffusion, which includes hanging drop (for smaller volumes), sitting drop (for larger drops), the sandwich drop, reverse vapor diffusion and pH gradient vapor diffusion methods. A drop containing a mixture of precipitant and protein solutions is sealed in a chamber with pure precipitant. Water vapor subsequently diffuses from the drop until the osmolarity of the drop and the precipitant are equal. The dehydration of the drop causes a slow concentration change of both protein and precipitant until equilibrium is achieved, ideally in the crystal nucleation zone of the phase diagram.[18] Batch crystallization relies on bringing the protein directly into the nucleation zone by mixing protein with the appropriate amount of precipitant. The batch method is usually carried out under oil to prevent the diffusion of water out of the drop.[19] Many of these methods can be performed using high-throughput automated instrumentation and miniaturization of crystallization experiments and have had huge impacts on protein crystallization in terms of saving time and conserving precious sample. For example, crystallization robots such as the *Phoenix*™ *RE* (Rigaku Corporation) and the *Mosquito*® (TTP Labtech), which can accurately and reproducibly dispense very small volumes (nl in size) into 96-well plates for automated screening and optimization of crystallization conditions, are now commonplace in many laboratories. The Rigaku CrystalMation™ system was setup to fully automate the crystallization process whilst dealing with sample volumes of 100 nl per experiment. Crystallisation on the nanoscale has one major draw back in that it is often difficult to translate hits into higher volume solutions in order to grow crystals for diffraction experiments. Crystal optimization aims to turn poor quality crystals into diffraction quality crystals that can be used for structure determination. A popular strategy for the optimization of crystallization conditions in vapor diffusion is microseeding. Seeding decouples nucleation

from crystal growth and involves transferring previously obtained seed crystals into undersaturated drops. Homogeneous seeding techniques include microseeding, streak seeding and macroseeding. Seed stock for microseeding can be conveniently generated using Hampton Research's Seed Bead kit. More recently, a simple, automated microseeding technique based on microseed matrix screening has been developed.[20] This method consists of the addition of seeds into the coarse screening procedure using a standard crystallization robot and has been shown to not only produce extra hits but also to generate better diffracting crystals. Successful cases for a simple semi-automated microseeding procedure for nanolitre crystallization experiments have also been recently described.[21] Furthermore, crystallization plate storage and inspection at periodic intervals are now fully automated. For example, the *Minstrel*™ *drop imager family* (Rigaku) and *Rock Imager* (Formulatrix) combine imagers with gallery plate hotels/incubators that simplify the inspection and management of experimental plates. Such a system has been installed at EMBL-Hamburg and the high-through put crystallization facility has been opened to the general scientific community.[22] The facility covers every step in the crystallization process from the preparation of crystallization cocktails for initial or customized screens to the setup of hanging-drop vapour-diffusion experiments and their automatic imaging.

2.3 Crystal handling

In the past, protein crystals were mounted in glass capillary tubes. In order to collect data, the capillary tube was mounted on a goniometer and exposed to X-rays at room temperature using laboratory X-ray source (low flux, sealed tube sources). Nowadays, data collection is handled using automated sample changers and micro-diffractometers in a cryo (100 K) environment utilizing brighter synchrotron radiation as the X-ray source. Significant progress has been made in automating the crystal mounting, crystal centering and energy scan in order to find metals or ions present in crystals that can be used for phasing.[23–25] Automated sample mounting systems allow users to mount samples on the beamline without entering the experimental hutch. These systems minimize the need for manual intervention and facilitate the rapid and systematic screening of dozens of samples.[26,27]

2.4 Post-crystallization treatments

Among the biggest problems in macromolecular crystallography is the relatively weak diffraction power of protein crystals and their sensitivity to ionizing radiation damage. Cryogenic methods provide great advantages in macromolecular crystallography, especially when synchrotron radiation is used for diffraction data collection. Apart from reducing the problem with radiation damage and enabling the storage and safe transport of frozen crystals, there are a number of additional benefits. Diffraction can be dramatically improved. Many factors contribute to improvements in data quality: obvious benefits are reduced thermal vibrations, enhanced signal-to-noise ratio, reduced conformational disorder and, in many cases, higher resolution. Of primary practical importance is the decrease in secondary radiation damage in the crystal caused by the diffusion of free radicals, permitting a complete data set to be collected from one single crystal. Cryogenic data collection has allowed efficient phasing using multi-wavelength methods.

When a crystal of a biological macromolecule is cooled to cryogenic temperatures, the main difficulty is to avoid the crystallization of any water present in the system, whether internal or external. Therefore a cooling procedure has to be chosen that leads to a glass-like amorphous phase of the solvent. In principle there are four possibilities: (a) cooling on a timescale too fast for ice formation to occur,[28] (b) cooling at high pressure by which the formation of common hexagonal form ice is circumvented,[29] (c) replacing the liquid surrounding the crystal with a water-immiscible hydrocarbon oil, such as *Paratone-N*,[30] *Paraffin* oil[31] and *LV CryoOil*™ (MiTeGen), (d) modifying the physicochemical properties of the solvent by addition of cryoprotectants in a way that a vitrified state can be reached at moderate cooling rates.

To prevent the nucleation of ice crystals, the last method is currently the most widely used. The crystal is permeated with a diffusible solvent containing cryoprotectants, such as glycerol, sucrose or other organic solvent.[32–34] Determining the initial and optimal cryoprotectant concentration is often a process of trial and error. One must find suitable cryoprotectant concentrations that do not destroy the crystalline order whilst, at the same time, allowing the solvent to form an amorphous glass upon rapid cooling. Recently, trimethylamine *N*-oxide (TMAO) has been shown as a very versatile cryoprotectant for macromolecular crystals.[35]

It has been shown that diffraction properties of flash cooled macromolecular crystals can often be improved by warming and then cooling a second time – a procedure known as crystal annealing. Two different crystal-annealing protocols have been reported[36–40] and many variants of these have been tried. The first method involves removing a flash cooled crystal from the cold gas stream and placing it in a cryoprotectant solution (either glycerol, MPD or Paratone-N oil) for several minutes before refreezing.[36,39] In the second method, the cold stream is blocked for a fixed amount of time before the crystal is allowed to re-cool.[37] Both annealing protocols can improve crystal resolution and mosaicity, although substantial crystal-to-crystal and molecule-to-molecule variability has also been observed. Recently, the flash annealing technique has been automated using a cryo-shutter,[41] a device that blocks the 100 K nitrogen stream that bathes the crystal, for a specific amount of time. The main advantage of the shutter system is that it allows a controlled instant re-cooling of the crystal and the user can perform the flash annealing experiment remotely without entering the experimental hutch.

Diffraction quality can also be improved by post-crystallization treatments, such as controlled dehydration[42] in order to attempt to improve the crystal diffraction properties. A user-friendly apparatus for crystal dehydration has been designed and implemented at the ESRF/EMBL beamlines.[43,44] In addition, Proteros biostructures GmbH has developed a *Free Mounting System* (*FMS*™) that precisely controls the humidity around a crystal, which can lead to dramatically improved diffraction data.

2.5 Data collection and processing software packages

In many cases, diffraction properties of crystals are not known in advance, especially when crystals are small (in the micrometer range) and cannot be prescreened using an in-house X-ray source prior to a synchrotron trip. It often takes a significant amount of time at the synchrotron to screen these sub-micron crystals to identify a well diffracting crystal suitable for data collection. Whilst collecting data at the synchrotron beamline, the user must make decisions about the parameters of the experiment - exposure time, oscillation angle, rotation range, detector distance, beam size and wavelength based on their experience, the visual inspection of the diffraction images and information output by data-processing packages. Most of the instrumentation in the experimental station is computationally controlled using software packages, such as *Blu-Ice*,[45] *CBASS*,[46] *MxCube*,[47] *JBlue-Ice*[48] and *MxDC*.[49] However, very often an intuitive decision is made by the user on the exposure time to use. In cases where this has been overestimated, it can lead to significant radiation damage before the completion of data collection. In addition, an inappropriate data collection strategy can lead to the failure of an experiment. Computationally efficient modeling of the data statistics for any combination of data collection parameters provides a foundation for making a rational choice. The modeling of data statistics as implemented in the program BEST, using a few test images allows one to quantitatively select which screened crystal gives the highest resolution using an appropriate radiation dose prior to data collection.[50]

The evaluation of the collected reflection intensities on the diffraction images involves the integration of the total intensity within all pixels of the individual spot profiles. Commonly used data processing packages include *XDS*,[51] *MOSFLM*[52] and DENZO/SCALEPACK.[53] These programs all give excellent results with high-quality diffraction data although their treatment of imperfect data differs owing to different approaches to indexing, spot integration and the treatment of errors. All these programs require crystallographers to make informative decisions and to input the correct experimental parameters in order to process the data successfully. There are ongoing attempt to develop expert systems that aim to automate the data collection strategy using the software *BEST*,[50] *RADDOSE*,[54] *MOSFLM* and *XDS* in order to reduce the time required to successfully collect high quality X-ray data at the synchrotron.

2.6 Structure determination methods

Experimental phasing (isomorphous replacement and anomalous scattering), molecular replacement and direct methods are used to solve protein structures. The general requirement for the exploitation of the anomalous signal for the determination of phase estimations *via multiple* or *single-wavelength anomalous diffraction* (*MAD* or *SAD*) techniques is that the protein crystal should contain anomalously scattering atoms (*e.g.* Hg, Au, Pt or Se). With the advent of tunable X-ray sources and improved data collection techniques, it is now possible to measure the intensities of diffracted X-rays with very high precision. The small differences in intensities between Bijvoet pairs due to the presence of heavy atoms can be used to located heavy atom positions and calculate initial estimates of the protein phase angle. One of the strategies widely used for the determination of novel protein structures is selenomethionine incorporation, where selenomethionine replaces methionine in the protein during expression. This method has been popular in protein X-ray crystallography and it is estimated that over two thirds of all new crystal structures have been determined using either the Se-SAD or Se-MAD technique. Novel structures can also be solved using the weak anomalous signals from atoms such as sulphur present in protein molecule. *Multiple* or *single isomorphous replacement* (*MIR* or *SIR*, respectively) methods also require the introduction of heavy atoms, such as mercury, platinum, uranium or gold, into the macromolecule under investigation. These heavy atoms must be incorporated into protein crystals without disrupting the lattice interactions, so that it remains *isomorphous* with respect to the native crystal. In the *SIR* method, intensity differences between the heavy-atom derivatized and native crystal are used to calculate experimental phases. Very recently, the *SIR* phasing concept has been re-applied in the *radiation damage-induced phasing* (*RIP*) *technique*, where the difference in intensities induced by radiation damage is used as a phasing tool.[55] Limitations of these phasing protocols are mainly due to the deleterious effect that a high X-ray dose has on a protein crystal. X-ray radiation damage induces many changes to the protein structure and to the solvent, resulting a number of damaged sites and a decrease in the diffraction quality of the crystal. Recently, as an alternative to X-rays, *Ultra-Violet* (*UV*) radiation has been used to induce specific changes in the macromolecule, which only marginally affects the quality of the diffraction[56] whilst inducing specific changes to the protein structure. This method is known as *UV-RIP* (*ultraviolet radiation-damage-induced phasing*). The most striking effect of UV radiation damage on protein crystals, as for X-ray radiation, is the breakage of disulphide bonds.[56] Recently it has been shown that crystals of selenomethionine (MSe) proteins can also be damaged when exposed to UV radiation.[57] The damage was very specific to Se atoms. The differences in intensities recorded before and after exposing crystals to UV radiation from a 266 nm laser were sufficient to locate the Se atom substructure and to phase the protein structure by the *UV-RIP* method. This method has considerable potential and selenium-specific UV damage could serve as an extra source of phase information, being an additional or even an alternative way of experimental phasing in macromolecular crystallography.[58]

The *Molecular Replacement* (*MR*) technique requires a search model for the protein under investigation, either determined from X-ray crystallography or from homology modeling, in order to calculate initial estimates of the phases of the new structure. The use of MR has become method of choice with the expansion of the database of known structures. MR is currently used to solve up to 70% of deposited macromolecular structures and the advantage of this method is that it is fast and does not require experimentally determined estimates of the phases. In cases when there are up to four molecules in the asymmetric unit of the crystal, the search model is structurally similar to the target protein and its oligomeric state is known, the MR method is fairly straightforward using programs such as *MOLREP*[59] and *AMoRe*.[60] However, in many cases the best available search models may be dissimilar in parts and in such cases the problem can become time consuming and laborious. Generally, structures that share high sequence identity exhibit low root mean square

deviations in their Cα atom positions. Potential model templates are therefore identified by sequence comparison searches. It is best to subsequently improve the model structure templates by omitting regions of large sequence diversity, which are likely to differ in structure and merely add noise to the search, and additionally truncating different side chains to common atoms,[59] CG atoms[61] or alanine. Since the *B* factors of the atoms also determine the scattering, modifications to the *B* factors, for example lowering the *B* factors for the hydrophobic core of the protein and increasing them in the surface-exposed residues, can also improve the model.[62] Where several possible models exist, none of which is expected to be significantly better than another *a priori*, the search can be repeated iteratively with each model or alternatively all the models can be grouped together as an ensemble in *Phaser*.[63] This program has shown improvements over traditional MR algorithms through the introduction of the maximum likelihood function for the rotation and translation searches and by scoring these searches based on the calculation of the log likelihood gain. To further streamline the MR procedure, a number of automated MR pipelines have been developed. These include the Bias Removal Server,[64] *CaspR*,[65] *BRUTEPTF*[66]and JCSG MR pipeline.[61] Other developments include *AUTO-RICKSHAW*,[67] which is principally used for experimental phasing but also uses phased MR as well as enabling a standard MR phasing protocol using *BALBES*,[68] *MrBUMP*[69] and a scheme for using comparative models in MR.[70] Recently MR phasing has been demonstrated for 2.0 Å data based on the combination of localizing model fragments such as small helices with *Phaser* and density modification with *SHELXE*.[71]

It is worth noting that if an MR search is difficult primarily because the model is extremely poor and the resolution of the X-ray data is limited (lower than 2.0 Å) then the significant amount of time require to obtain a solution. This is partly because the model suffers from bias and often requires iterative, time consuming manual correction using computer graphics in combination with model refinement. Interestingly, the determination of the substructure becomes easier when an anomalous difference Fourier synthesis can be calculated using preliminary phases from a MR solution. The subsequent use of this substructure to generate an unbiased electron density map is often referred to as *MRSAD* (*Molecular Replacement with Single Wavelength Anomalous Diffraction*).[72] A combination of MR and SAD has been automated and incorporated into the structure determination platform *AUTO-RICKSHAW*. The complete *MRSAD* procedure includes molecular replacement, model refinement, experimental phasing, phase improvement and automated model building and it has been shown that poor MR or SAD phases with phase errors larger than 70° can be improved using this described procedure[73] and a large fraction of the model can be determined in a purely automatic manner.

2.7 Computational resources for crystal structure determination

New software packages have been developed in order to accelerate the effort for determining the 3D structure of proteins.[74] The software pipelines have varying degrees of automation deriving from different aims, but all require minimum user input in order to facilitate the automated location of heavy atom sites, phase determination and phase improvement by solvent flipping/flattening, model building and refinement. *AutoSHARP*,[75] *HKL-3000*,[76] *CRANK*,[77] *BnP*,[78] the *PHENIX* suite[79] and *HKL2MAP*[80] for *SHELX*[81] are highly automated and provide all the tools necessary to proceed from substructure solution and phasing through to displaying and interpreting the resultant electron density map. The *AUTO-RICKSHAW* suite executes many widely used programs for automatic protein structure determination.[67,73] The software pipeline provide automated protocols to enable protein models to be built rapidly without user intervention, providing feedback on the success of the experiment whilst the crystal is still at or near the beamline. *AutoSHARP* includes various *CCP4*[82] programs,[83] uses the *SHELXD* software[81] for locating heavy atoms and carries out density modification using either *DM*[86] or *SOLOMON*,[87] whilst *ARP/wARP*[88] or *BUCCANEER*[89] are used for automated model building. This pipeline can be run without user intervention once suitable input has been provided and can be rerun from any of the structure solution steps by the user whenever desired. The *CRANK* package invokes *BP3*,[85] *CRUNCH2*,[84] *SHELXD*, *SOLOMON*, *DM*, *RESOLVE*,[90] *BUCCANEER* and *ARP/wARP* along with a few *CCP4* programs and uses standard XML input at every step of the structure solution. The process may be invoked either using the *CCP4* graphical user interface (*CCP4i*) or offline and the user must choose the defined path through the pipeline. The *BnP* pipeline includes *SnB*[91,92] and the *PHASE* package for structure solution. *HKL-3000* is commercially available software, which includes the data processing programs *DENZO*/*SCALEPACK* along with structure solution programs, including modified versions of *MLPHARE*, *SHELXC*/*D*/*E*, *DM*, and *ARP/wARP*. The *PHENIX* software suite is a highly automated system for macromolecular structure determination that can rapidly arrive at an initial partial model of a structure without significant human intervention, given moderate resolution and good quality data. The *AUTO-RICKSHAW* pipeline has been developed with its primary aim to validate the X-ray diffraction experiment whilst the crystal is still at or near the synchrotron beamline.[67] The software pipeline is optimized for speed so that the user has the ability to evaluate their data in the minimum possible time. It makes use of publicly available macromolecular crystallography software and it runs as a web server. The entire process in the pipeline is fully automatic. Each step of the structure solution process is governed by the decision-making module within the pipeline, which attempts to mimic the decisions of an experienced crystallographer for a number of phasing protocols (SAD, MAD, RIP, MR and variations thereof). Once the input parameters (number of amino acids, heavy atoms, molecules per asymmetric unit, probable space group and name of phasing protocol) and X-ray intensity data have been input into the web server, no further user intervention is required. It proceeds step by step through the structure solution using the computer coded decision makers. In cases where a problem is encountered during the structure solution process, the user is informed so that the data collection, the data quality, space group ambiguity or optimization of the anomalous signal is flagged as a problem. Once all the steps have been run successfully, the web server provides a tar ball which includes all the necessary files to evaluate the electron density map and model including ready-made scripts for the graphics programs *COOT*,[93]

O[94] and *XtalView*.[95] All of the above mentioned software except *HKL-3000*[76] is freely available to academic users and the *AUTO-RICKSHAW* web server (http://www.embl-hamburg.de/Auto-Rickshaw) is freely accessible to the scientific community to aid their protein structure determination effort.

3 Structure determination and architecture of major enzymes from various alkaloid biosynthesis pathways

3.1 Monoterpenoid indole alkaloids

The Indian plant *Rauvolfia serpentina* is called Sarpagandha (the snake root) in Hindi, and belongs to the Apocynaceae family. This plant is listed in the earliest Ayurvedic medicinal text, the Charaka Samhita (*ca.* 700 B.C.), and has been used to treat mental illness and insomnia. The roots of *R. serpentina* contain numerous alkaloids, which are therapeutically used and ajmaline is one of them, serving as an antiarrhythmic drug (Fig. 1). The biosynthesis of the alkaloid ajmaline consists of a 10-step reaction sequence which is catalysed by several enzymes of different classes[96,97] all of them exhibiting a high substrate specificity. The pathway is also one of the best known examples in modern proteomics research for which experimental evidence is available not only for all enzymes directly involved in the pathway but also for those catalyzing side routes.[98] Together, this yields a comprehensive knowledge of alkaloid metabolism in *R. serpentina* at the enzymatic level. In recent years we have systematically delineated the biosynthetic pathway leading to ajmaline in *Rauvolfia*. The complex biosynthetic pathway starts with tryptamine and the monoterpene secologanin, contains ten different enzyme catalysed steps and leads finally to the six-membered ring system of ajmaline, bearing nine chiral carbon atoms.

3.1.1 Strictosidine synthase (STR1) (2V91, 2VAQ, 2FP8, 2FPB). The enzyme strictosidine synthase (STR1) (EC 4.3.3.2) initiates all biosynthetic pathways leading to the entire monoterpenoid indole alkaloid family representing an enormous structural variety of about 2000 compounds in higher plants. It catalyses the condensation of tryptamine and secologanin to yield strictosidine. Cloning of the cDNA for the enzyme from *R. serpentina* is a first such example in the entire field of alkaloid biosynthesis.[11,99] STR1s from two other sources, *Catharanthus roseus* G. Don[100] and *Ophiorrhiza pumila*[101] could also be identified and the biology of the enzyme has been reviewed several times during the last one and a half decades.[102,103] To obtain sufficient amounts STR1 was overexpressed in *Escherichia coli* and the enzyme purified at the mg-scale using nickel-NTA affinity chromatography. The N-terminal His-tag was cleaved *via* enzymatic digestion and the His-tag-free enzyme was screened against large number of crystallisation conditions. The best crystals of STR1 and its complexes were grown using the

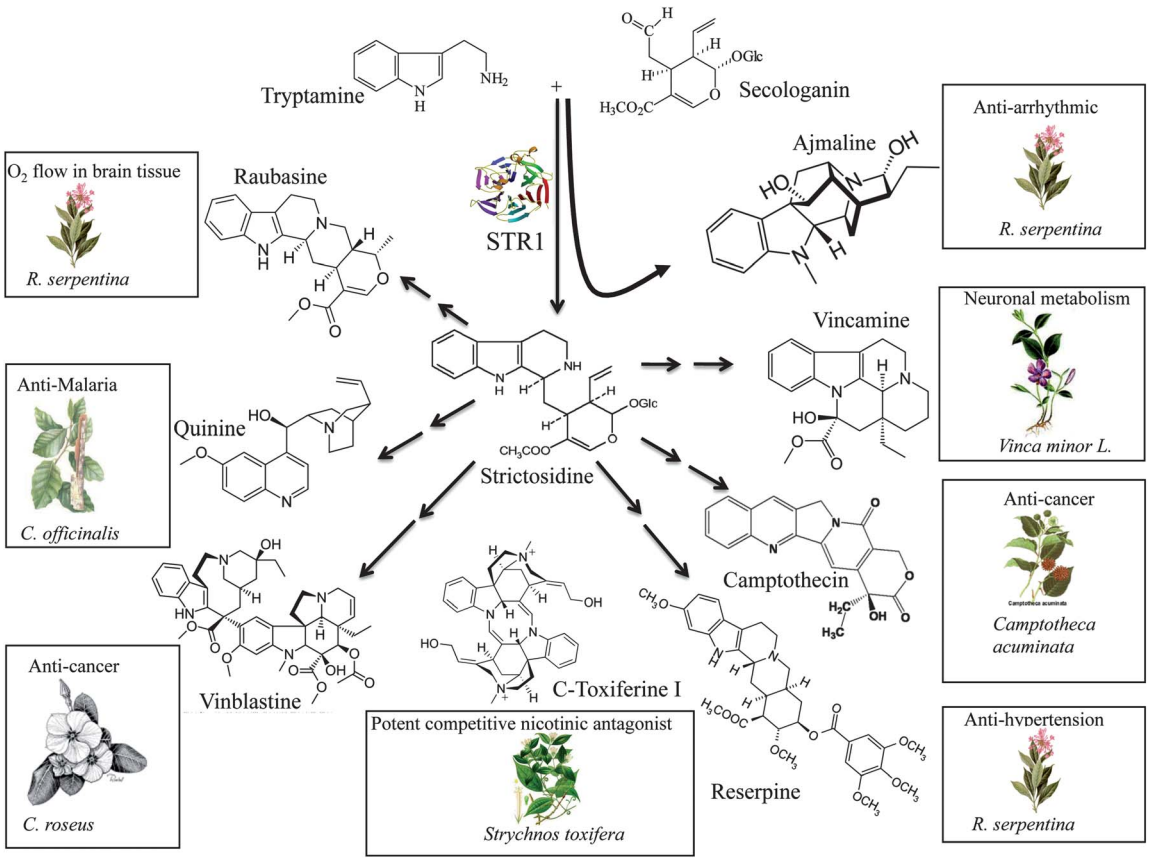

Fig. 1 STR1 catalyzes the stereo-specific condensation of tryptamine and secologanin leading to 3a(*S*)-strictosidine, a central reaction in the biosynthesis of the entire family of monoterpenoid indole alkaloids in plants. (STR1 = Strictosidine synthase) (The figure was partially adapted from ref. 161).

hanging-drop vapor diffusion method.[104] Crystals were cryoprotected with glycerol before being flash frozen at 100 K temperature and the X-ray diffraction data were collected on a synchrotron beam line. Crystals of STR1 belong to space group *R3*. At that time there were no protein structures available showing significant homology to STR1, consequently the enzyme was labeled with seleno-methionine in order to use the multiple wavelength anomalous diffraction (MAD) approach to obtain experimental phases. The structure was solved using *AUTORICKSHAW*, as described in Ma, Panjikar *et al.*, 2006[105] and refined to 2.3 Å resolution. STR1 was the first structure of an enzyme which catalyses a Pictet–Spengler type reaction. STR1 also represents the first 6-bladed four-stranded β-propeller fold found in plant proteins (Fig. 2). Each blade of the propeller consists of four-stranded antiparallel beta-sheets A–D (Fig. 2a). The blades are arranged around a six-fold axis to form the binding pocket in which the Pictet–Spengler reaction is catalyzed. The binding pocket consists of hydrophobic residues and a charged residue. Structure-based sequence alignment revealed a common repetitive sequence motif (three hydrophobic residues are followed by a small residue and a hydrophilic residue) indicating a possible evolutionary relationship between STR1 and several sequence-unrelated six-bladed β-propeller structures. Structural analysis and site-directed mutagenesis experiments demonstrate the essential role of Glu309 in catalysis.[105] The structure of the complex obtained by cocrystallization of STR1 with the substrate tryptamine (Fig. 2b) allowed further insight into the nature of the reaction center.[105] The substrate is located deep in the pocket with the amine group coordinated to Glu309. The indole part is sandwiched between the aromatic rings of Tyr151 and Phe226. This orientation most probably helps to keep the tryptamine in an active position for the condensation reaction with the second substrate, secologanin. The structure of the complex with secologanin shows that 1) the hydrophilic glucose part points out of the catalytic center towards the solvent, and 2) the aldehyde group points towards the nearest residue Glu309, which is close to the amino group of tryptamine. Clearly this is the favored position of both substrates for the primary reaction of the Pictet–Spengler condensation. The position of the substrates was demonstrated by the structure of STR1 with its reaction product strictosidine.[106] As shown in Fig. 2b–d, the location of the enzyme product is nearly identical to that of both substrates. The superimposition of all three ligands bound in the active center shows they are in nearly identical locations, with only relatively small differences (~1.5 Å), whereas the indole part of the inhibitor complex of STR1[107] (showing IC_{50} value of 3 ± 0.5 nM) has a different conformation compared to the conformation when tryptamine alone was bound to strictosidine synthase (Fig. 2d). In contrast, the secologanin moiety of the inhibitor overlaid well with the enzyme–secologanin complex.

3.1.2 Strictosidine glucosidase (SG) (2JF6, 2JF7). Strictosidine β-D-glucosidase (SG) follows the enzyme strictosidine synthase (STR1) at the beginning of the biosynthesis of the entire monoterpenoid indole alkaloid family. Initially SG has been detected and partly characterized from cell suspension cultures of *Catharanthus roseus* by Treimer and Zenk (1979).[108] Its cDNA has been cloned from various sources, including *R. serpentina*.[109] The enzyme shows high substrate specificity and, based on its

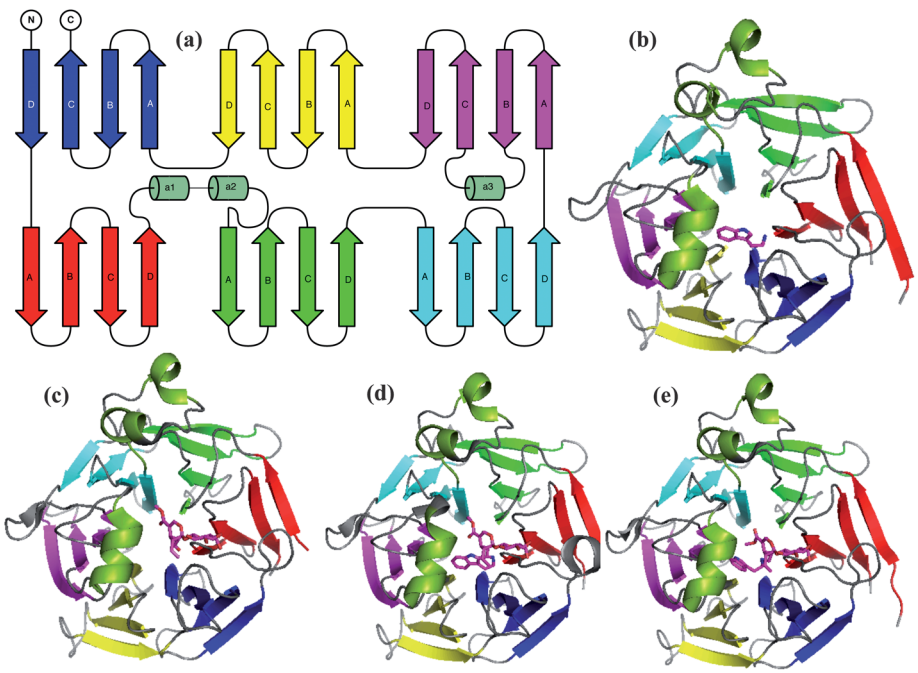

Fig. 2 (a) The topology of the STR1 structure. Each blade (consisting of four β-strands) of the propeller is shown in a different color (blue, red, green, cyan, magenta and yellow), the connecting loop is shown in grey and helices in light green. (b) Front view of the six-bladed β-propeller in complex with tryptamine (PDB code: 2FP8), (b) in complex with secologanin (PDB code: 2FPC), (c) in complex with strictosidine (PDB code: 2V91) and (d) in complex with inhibitor (PDB code: 2VAQ). The color code is consistent with the topology of STR1 structure. Each ligand complex is shown in the same orientation.

primary structure, it belongs to the family 1 of glycosyl hydrolases. Since a detailed explanation of the enzyme reaction and the substrate acceptance could not be determined by sequence alignment alone we planned to crystallize SG after overexpression in *E. coli* in order to establish the 3D-structure of the enzyme from *R. serpentina*.

The cloning of the cDNA of SG was based on primer development from the SG sequence of *C. roseus* and raucaffricine glucosidase of *R. serpentina*.[110] The expression vector pQE-2 resulted in good expression in *E. coli*. This strategy delivered sufficient amounts of pure SG after His-tag cleavage for crystallization experiments.

Crystallisation screening was performed using various commercially available crystallisation kits. The first hit condition was optimised for good diffracting crystals.[111] Crystals of SG belonged to tetragonal space group $P42_12$. The structure of SG was determined using the molecular replacement method and using the structure of the homologous maize enzyme[112] (PDB code 1E1E) as a search model. The structure was refined to resolution 2.48 Å.[113] The fold of the structure is characterized by 8 parallel β-strands forming a β-barrel and the barrel is surrounded by helices. Each β/α repeat is connected by either loop or in combination with loop and helices or strands at the top of the barrel. The barrel hosts a binding site for the natural substrate strictosidine. The Glu207Gln mutant showed no detectable catalytic activity with strictosidine compared with wild-type SG, allowing soaking experiments with the natural substrate strictosidine to be performed. The three dimensional structure of the mutant in complex with strictosidine was resolved at 2.82 Å and is identical to that of wild-type enzyme (Fig. 3). The crystal structures of native SG and the complex of its inactive mutant Glu207Gln with the substrate strictosidine provide structural understanding of substrate binding and identify amino acids lining the active site surface. The glycone part is bound deep in the pocket, and the aglycone part points away toward the solvent, even though the aglycone part of the substrate is hydrophobic in nature (Fig. 3). The aglycone part is surrounded by residues Thr-210, Phe-221, Met-275, Met-297, Gly-386 and Trp-388 and, the majority of which are hydrophobic. Important residues for recognition of the aglycone unit seem to be Gly-386 and Trp-388. Gly-386 is in very close proximity to the indole system of strictosidine; therefore, mutation to larger residues at this position decreases the enzyme activity.[113] Trp-388 is discussed in the next paragraph. The glucose moiety interacts with a number of hydrophilic residues (Tyr-48, Gln-57, His-161, Asn-206, Gln-207, Asn-343, Tyr-345, Glu-416, Trp-465, Glu-472 and Trp-473) (Fig. 2a). Comparison of the catalytic pocket of SG with that of other plant glucosidases indicates the structural importance of Trp388. In SG, Trp388 occupies a different conformation compared to all other plant glucosidases, most likely due to the large area of the substrate and stacking with the hydrophobic residue. Structural analysis and site-directed mutagenesis experiments demonstrate the essential role of Glu207, Glu416, His161 and Trp388 in catalysis. The most important functional role of SG is its biosynthetic potential to "activate" the glucoside strictosidine to enter multiple indole alkaloid pathways. SG represents the first structural example within the synthesis-related (as opposed to defence-related) glucosidases. The enzyme delivers intermediates for the synthesis

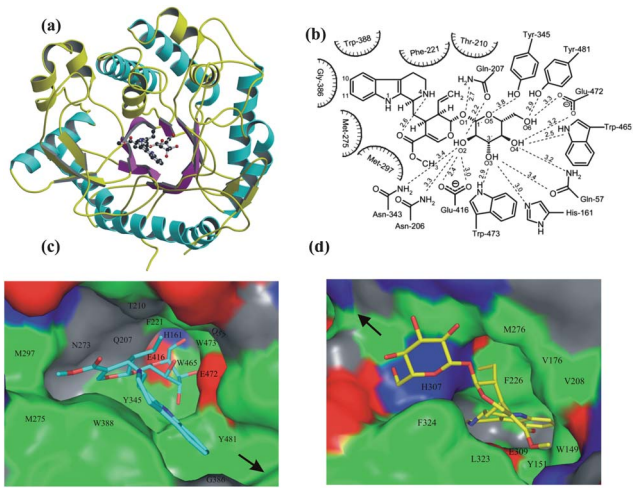

Fig. 3 The overall structure of SG (PDB code: 2JF7) from *R. serpentina* illustrating its $(\beta/\alpha)_8$ fold, the groove leading to the catalytic centre is harbouring its substrate strictosidine and highlights secondary structure elements. (a) (β/α) repeats are illustrated in magenta and cyan, respectively. The extra helices and strands as well as the connecting loop are shown in yellow. (b) The hydrogen-bonding network between the glucosidic part of strictosidine and residues wthin 3.8 Å distance in the ligand structure of SG inactive mutant Glu207Gln (picture adapted from ref. 113) (c) Surface representation of SG-Glu207Gln complex (PDB code: 2JF6) with strictosidine, hydrophobic residues (Tyr, Trp, Phe, Leu, Met, Cys, Ile and Val), positively charged residues (His), negatively charged residues (Glu) and hydrophilic residues (Ala, Gly, Ser, Thr, Pro, Gln, Asn) are shown in green, blue, red and grey, respectively. The surrounding residues are labeled with their single-letter code. (d) Surface representation of STR1-strictosidine complex (PDB code: 2V91), color code is the same as presented in figure (c). The arrows point to the solvent in both figures (c) and (d).

of 2000 different indole alkaloids. The glucosidase hydrolysing coniferin provides the intermediate for the biosynthesis of lignin, which is the second most occurring natural product.

3.1.3 Vinorine synthase (VS) (2BGH). Vinorine synthase creates the ajmalan-skeleton from sarpagine type alkaloid *epi*-vellosimine during the biosynthesis of Rauvolfia alkaloids. The enzyme belongs to the benzylalcohol acetyl-, anthocyanin-*O*-hydroxy-cinnamoyl-, anthranilate-*N*-hydroxy-cinnamoyl/benzoyl-, deacetylvindoline acetyltransferase (BAHD) enzyme superfamily, members of which are involved in the biosynthesis of several important drugs, such as morphine, taxol, or vindoline, a precursor of the anti-cancer drugs vincaleucoblastine and vincristine.[114] The BAHD members identified to date are all monomeric enzymes with a molecular mass ranging from 48–55 kDa. The average number of amino acids in a BAHD protein is approximately 445.[115] All conserved motifs of this superfamily can be found in the enzyme's sequence, at the active site the HxxxD motif and near the C-terminus the conserved sequence DFGWG.

The VS was expressed as a His-tagged fusion protein in *E. coli* M15 cells and purified using an affinity chromatography (Ni-NTA) and anion exchange. After cleavage of the His-tag crystals were grown. The structure of VS was solved by the MAD method using selenomethionine (SeMet) substituted VS at 2.6-Å

resolution.[116] Vinorine synthase is a globular protein and it contains two nearly equal sized domains (domain 1 and domain 2) that are connected by a loop. The structure contains 14 β-strands and 13 α-helices (Fig. 4a). The two domains are connected with a large crossover loop (residues 201–213), which spans nearly 36 Å. Domain 1 contains a mixed six stranded β-sheet, which is covered on both sides by seven helices. The strand (residues 370–372) protrudes out from the domain 2, and forms an anti-parallel strand in domain 1. Domain 1 also contains a pair of β-strands on the surface of the protein at one end of the central β sheet. Domain 2 contains 6 helices and a β-sheet, consists of 6 mixed β-strands. Both domains share a very similar polypeptide backbone fold, however, their topology is different. The interface between the two domains forms a solvent channel, which runs through the VS molecule. The active site HxxxD sequence motif in the VS structure is located at the interface between the two domains and the catalytic residue His160 of this motif is accessible from both sides of the channel (Fig. 4b). Mutation of His160 or Asp164 to Ala results in complete loss in enzyme activity.

Mutation in each of eleven residues (Ser68, Cys89, Ser413, Asp360, Ser16, Asn219, Asp362, Ser29, Ser243, Asp32, Cys149) results in mutant derivative that shows either same activity as wild-type or decrease in enzyme activity (Fig. 5). His160 is hydrogen-bonded with two main chain carbonyl oxygens in addition to the side chain of Asn293. Asp164 points away from His160 and the active site.[116] The residue is rather involved in the formation of a salt bridge with the conserved Arg279, which is most likely to be important for maintaining the geometry of the active site. Only the His160 residue is directly involved in enzyme catalysis process. The residue functions as a general base during catalysis (Fig. 4c). This structural arrangement allows the ligand acetyl-CoA and the substrate *epi*-vellosimine to approach the active site independently from opposite sides of the enzyme, the front face (CoA binding) and the back face (substrate binding), respectively. Surprisingly, the DFGWG motif, which is indispensable for the catalyzed reaction and unique to the BAHD family, is located far away from the active site and seems to play only a structural role. Despite low sequence identity, the

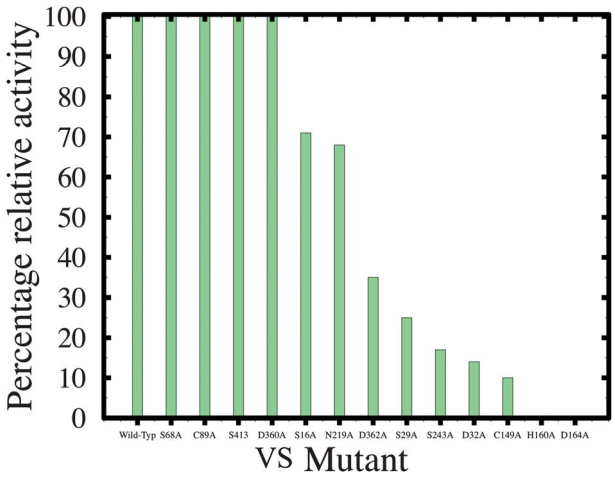

Fig. 5 The 2D diagram shows the VS mutants *versus* their percentage relative activity. H160A and D164A mutants show no activity.

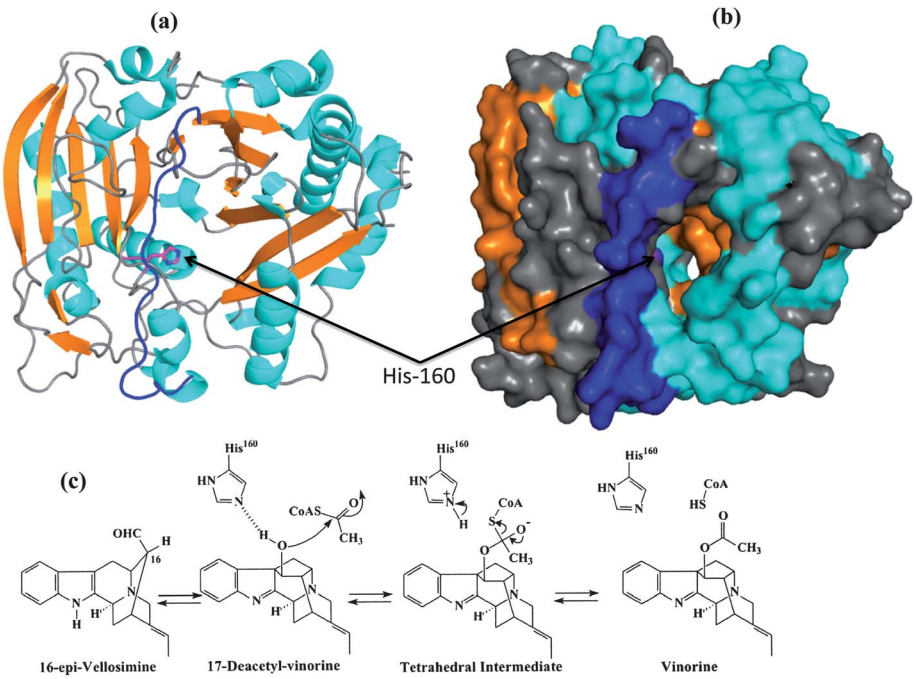

Fig. 4 Structure of vinorine synthase (PDB code: 2BGH). (a) The VS structure is represented as cartoon. α-helices are shown in cyan, and β-strands are shown in orange. The large crossover loop (amino acids 201–213) that connects both domains is marked in blue. The catalytic residue His160 is shown as ball-and-stick. (b) Surface representation of VS structure in the same orientation as shown in the cartoon representation and color code of the surface representation is same as the secondary structure element in the structure, the active site residue His160 is marked in the reaction channel. (c) Catalytic mechanism of vinorine synthase. The involvement of His160 is shown as a general base catalyst in proton abstraction and the subsequent formation of a putative tetrahedral intermediate is depicted (scheme adapted from ref. 116).

two-domain structure of vinorine synthase shows surprising similarity to structures of several CoA-dependent acyltransferases such as dihydrolipoyl transacetylase, polyketide-associated protein A5, and carnitine acetyltransferase.

3.1.4 PNAE (2WFL, 2WFM, 3GZJ). Polyneuridine aldehyde esterase (PNAE, EC 3.1.1.78) catalyses a reaction in the middle of the ajamaline biosynthetic pathway, leading from polyneuridine aldehyde to *epi*-vellosimine. *Epi*-vellosimine is the direct precursor of the ajmaline skeleton-bearing alkaloid vinorine. After spontaneous epimerisation of *epi*-vellosimine to vellosimine it leaves the biogenetic route to sarpagine type alkaloids. Based on the sequence and structure comparison, PNAE shows high similarity to the members of the α/β hydrolase fold enzyme family. The esterase consists of 264 amino acids with a molecular weight of 29.65 kDa. PNAE and its inactive mutant His244Ala were overexpressed into *E. coli* and purified using chromatography (Ni-NTA). Attempts to crystallize PNAE failed for many years until high-throughput crystallisation approaches followed by long term fine-tuning of crystallization conditions revealed suitable crystals for X-ray diffraction experiment.[117] Crystals of PNAE were grown using the hanging drop vapor diffusion method. Crystals of the PNAE mutant His244Ala were also grown at similar conditions. Crystals of PNAE His244Ala were soaked in 1 mM PNA solution (5% ethanol in precipitant buffer) in the crystallisation drop. For X-ray measurement, native, mutant and complex crystals were cryoprotected by addition of 20% (v/v) glycerol to the precipitant buffer (1 mM PNA was added into cryobuffer for the complex) before being flash-frozen in a stream of cold nitrogen at 100 K. The X-ray structure of PNAE was solved at 2.1 Å resolution using the molecular replacement protocol of *AUTO-RICSKHAW* in the C222 space group. The crystal structure of Salicylic Acid-Binding Protein 2 (SABP2) from *Nicotiana tabacum*, (PDB code 1XKL) was used as a search model. Crystal structures of the mutant (H244A) protein and its complex with PNA were also solved using molecular replacement at 2.2 Å resolution with the native PNAE structure as a search model in the C2 space group.[117] The overall structure of PNAE contains 6 β-sheets flanked by 5 α-helices (Fig. 6a–b). Together with the cap domain[118] consisting of 2 β-sheets and 3 α-helices, it is now unequivocally confirmed that PNAE belongs to the α/β hydrolase fold. Opening of the reaction channel is located in the "cap", a region structurally highly flexible in contrast to the canonical core of the α/β hydrolases. Analysis of the structure clearly shows that active centre consists of three residues Ser87, Asp216 and His244 forming the catalytic triad (Fig. 6b). Mechanistically, an additional mutation together with the enzyme-product structure is very helpful in order to elucidate the reaction mechanism. Despite absolute necessity of the triad,[119] Met245 in the binding pocket is also indispensable for hydrolysis because mutant Met245Ala is inactive.[117] PNAE belongs to the subgroup of hydroxynitrile lyases (HNLs), of which the mechanism has been intensively studied[120–122] and HNLs exhibit the same catalytic triad as PNAE.

3.1.5 Raucaffricine glucosidase (RG) (3U57, 3U5U, 3U5Y). Raucaffricine glucosidase (RG) catalyses the deglucosylation of raucaffricine (the substrate of RG). This hydrolysis leads to the aglycone vomilenine, a direct intermediate on a side route to the target compound ajmaline. The strictosidine glucosidase enzyme acts downstream of the synthase STR1 in the ajmaline biosynthesis pathway. Although the amino-acid sequence similarity is as high as 50%, the substrate acceptance of the two glucosidases varies greatly: RG can also hydrolyse strictosidine to some extent (the substrate of SG); however, SG is unable to hydrolyse raucaffricine.

The RG cDNA was cloned into the pQE-2 vector and expressed in *E. coli* strain M15 with an N-terminal His_6-tag and purified by using a Ni-NTA column.[123] Following the protocol, 5 mg pure RG could be isolated per litre of bacterial culture, and was instantly suitable for systematic crystallization experiments.

Crystals of RG were obtained by the hanging-drop vapour diffusion technique.[123] Crystallization of the inactive mutant RG-E186Q and preparation of its complex with the substrate 1,2(*S*)-dihydro-raucaffricine, and secologanin, were carried out with the optimal crystallisation conditions at 20 °C. Once the protein crystallised, freeze-dried ligands 1,2(*S*)-dihydro-raucaffricine and secologanin were added directly to the crystallisation drops.

Prior to X-ray data collection, the crystals of wild-type RG were treated with cryoprotectant and flash-cooled to 100 K. Diffraction data were then collected at a synchrotron beamline to 2.20 Å resolution. The 3D-structure of native RG was determined by molecular replacement using the software pipeline *AUTO-RICKSHAW* with the structure of homologous SG (PDB 2JF7) as a target model.[124] X-ray data from all RG-ligand complex crystals were collected at room temperature after the soaking experiments for 10 min to 3 h since these crystals were sensitive in presence of ligand and cryoprotectant.[196] Changing from cryo to room temperature measurements was, therefore, essential to proceed with the RG structure project.[124] Structures of the RG-ligand complex were solved using either molecular replacement or Fourier synthesis using the native structure of RG. The crystal structure of RG consists of 13 α-helices and 13 β-strands (Fig. 7). The overall structure of the wild-type RG enzyme possesses the expected TIM $(\beta/\alpha)_8$ barrel fold belonging to GH family 1 (http://www.cazy.org), as for SG.[113] RG now represents the 11th (excluding myrosinase) structural example from a total of ~380 identified GH-1 members of plant origin and the 32nd example from all structurally elucidated GH-1 proteins to date. The $(\beta/\alpha)_8$ barrel of RG hosts binding sites for the natural substrate, raucaffricine and its derivative, dihydro-raucaffricine (DHR). The groove leading to the catalytic center is formed by irregular loops between the secondary structures located on the surface of the enzyme. Interactions of amino acid residues in <4.0 Å distance are schematically illustrated in Fig. 7b. RG contains the catalytic residues Glu186 and Glu420 similar to SG. The RG mutants Glu186Asp, Glu186Gln, Glu420Gln and the double mutant Glu186Gln/Glu420Gln do not show any hydrolase activity.[124] Glu186 and Glu420 show a distance of their carboxylcarbons of ~5.1 Å which provides the space for reacting with the glucosidic bond. A total of eleven hydrophilic residues in SG interact with the glucose part of its substrate strictosidine (Fig. 3b), whereas RG requires only seven residues for keeping the glucose in the catalytically desired position (Fig. 7b), all being identical with those in the SG complex. The most striking difference at the catalytic center of

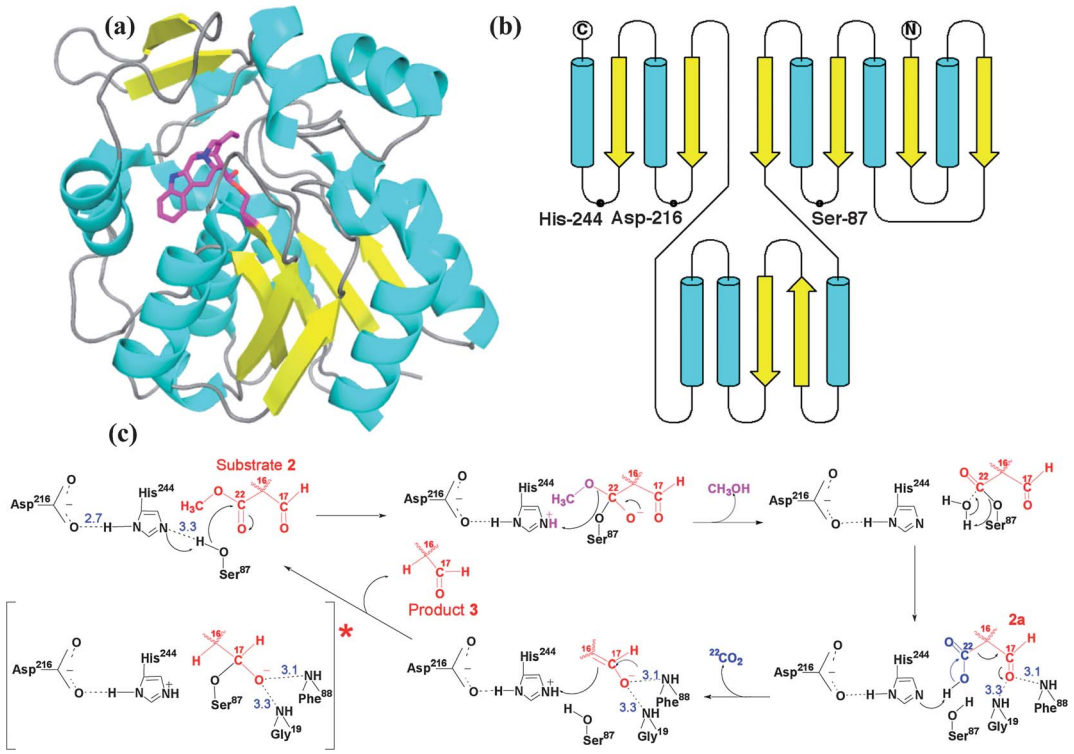

Fig. 6 The overall structure of PNAE (PDB code: 3GZJ) illustrating its (β/α) hydrolase fold, (a) enzyme product 3 (magenta) covalently linked as a hemiacetal to Ser87 in the active site of PNAE mutant His244Ala and secondary structure elements are highlighted in cyan (helix) and in yellow (strand). (b) A 2D-topological diagram of PNAE illustrating the connection of the structural elements ("cap domain" at bottom), color code is the same as shown in the three dimensional structure of PNAE. The active site residues are marked. (c) Proposed reaction mechanism of PNAE (scheme adapted from ref. 117), partial structures of substrate 2 and product 3 are shown; the covalently linked enzyme product 3 is marked with a star; average distances of residues are in Å.

RG and SG, is different conformations of the tryptophan side chains at the identical positions of RG-Trp392 and SG-Trp388 (Fig. 7d). The conformation helps each protein to recognize its natural substrate.

3.1.6 Perakine reductase (PR) (3UYI, 3VOS, 3VOT, 3VOU). Perakine reductase (PR) is a novel member of the aldo-keto reductase (AKR) enzyme superfamily from higher plants. PR from the plant *Rauvolfia serpentina* is involved in a side-route of the monoterpenoid indole alkaloids biosynthesis by performing NADPH-dependent reduction of perakine, yielding the alkaloid raucaffrinoline and it leads to biosynthesis of dihydroperaksine.[125,126]

PR was expressed as a His-tagged fusion protein in *E. coli* M15 cells and purified using an affinity chromatography (Ni-NTA) and anion exchange. The recombinant protein was not suitable for crystallisation due to strong degradation. In order to minimize degradation, the residues Lys 100, 244 and 296 were mutated to Ala by side directed mutagenesis. Moreover, because of unsuccessful crystallization of native PR, reductive methylation was chosen to modify the protein surface as described by Rypniewski *et al.* (1993).[127] Thus, the first crystallization of PR was obtained with methylated triple mutant PR[128] followed by the crystallization of methylated wild type PR.[129] The crystals were derivatised using a platinum derivative and the MAD data collection around the Pt-LIII edge was performed at a tuneable Synchrotron beamline. The structure was solved using AUTORICKSHAW software to resolution 2.31 Å in $C222_1$ space group. The methylated wild type PR folds into an unusual α/β-barrel consisting unexpectedly of only six β-strands with eight α-helices which pack along the outside of the β-strands (Fig. 8a). So far, all other 3D-structures identified for AKR members fold as $(\alpha/\beta)_8$-barrels which contains eight parallel β-strands and eight α-helices and it is also known as a TIM barrel.[130,131] In addition there are some remarkable differences when PR is compared to other AKR members.[129] The barrel structure of PR starts with an α-helix instead of a β-strand. Co-crystallisation of the protein with its ligand or soaking of the ligand into the protein crystals was unsuccessful. This crystal form contained one molecule in an asymmetric unit. C-terminal part of the structure interacts with symmetry related molecule in the unit cell and forms a β-barrel. Interestingly the active site is blocked by the symmetry related molecule that prevents accessing any ligand (Fig. 8b). Therefore another crystal form was needed where the binding site is accessible for the ligand. The structure was analysed and a mutant was predicted which will possibly break the interaction between the molecules. The residue Ala-213 was chosen for mutation, which is located at the interface region between the two molecules, one being symmetry related (Fig. 8a,b). A bulky Trp was selected, and the PR-A213W mutant, which is still functional, was generated to prepare PR complexes with cofactor and/or substrate bound. The mutant was crystallized in different crystal form under the same conditions as methylated His_6-PR. The PR-A213W mutant crystal contained two molecules in the

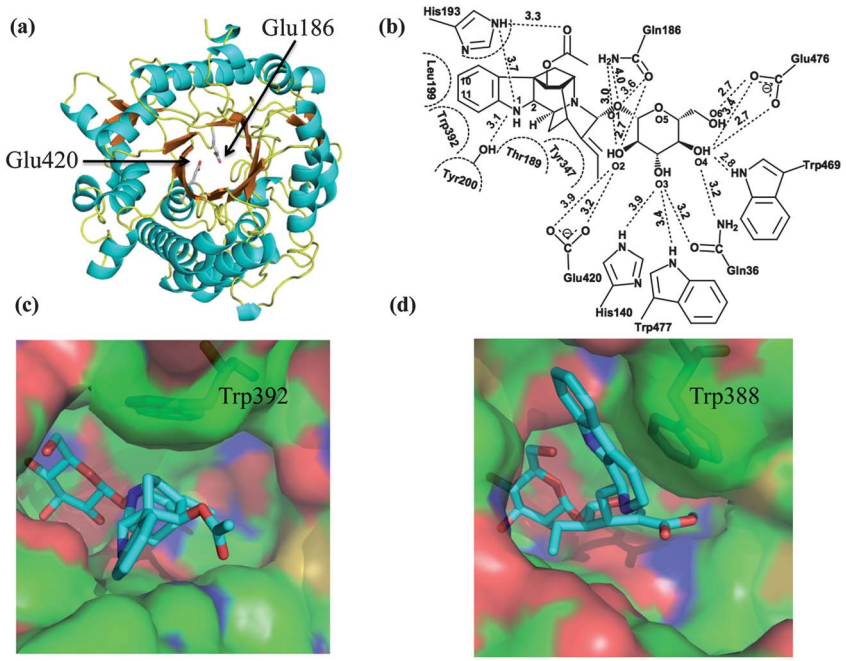

Fig. 7 (a) The structure of the RG-Glu186Gln mutant (PDB code: 3U5U), it resembles the $(\beta/\alpha)_8$ –barrel fold, illustrating binding pocket and the catalytic residue Glu420 is marked in the picture where as second catalytic residue is shown in mutation Glu186Gln. (b) Hydrogen binding network between 1,2(*S*)-dihydro-raucaffricine and residues within 4.0 Å distance in the ligand structure of RG inactive mutant Glu186Gln, illustrating the interactions of amino acids with both the aglycone and the glucosidic part of the substrate (picture adapted from ref. 124). (c) The binding pocket of RG is shown as a surface representation and bound ligand 1,2(*S*)-dihydro-raucaffricine as ball-and-stick (PDB code: 3U57). The residue Trp392 is shown in transparent surface as ball-and-stick. (d) Similar to figure (c), the binding pocket of SG is shown together with strictosidine (PDB code: 2JF6) in the same orientation as the structure of raucaffricrine-RG-Glu186Gln complex.

asymmetric unit with space group $P3_221$. The crystal diffracted to 1.77 Å resolution. The structure was solved using the molecular replacement method and the His_6-PR structure as a search model. The structure was identical to His_6-PR except for the C-terminal part. The cofactor NADPH was visible without the nicotinamide riboside moiety in the C-terminal core of the barrel of PR (Fig. 8c) and points towards the active site, which is formed by the catalytic tetrad Asp52, Tyr57, Lys84 and His126 at the center of the barrel, as in most AKR enzymes.[132] Upon NADPH binding, dramatic conformational changes and movements were observed: two additional β-strands in the C-terminus become ordered to form one α-helix and a movement of up to 24 Å occurs (Fig. 8d). This conformation change creates enough space around the catalytic site for substrate binding. Such a conformational change was not observed in the apo form of the PR-A213W mutant structure. In fact, the C-terminal segments (residues 205–219, 311–337) of the apo A213W mutant were disordered. PR represents the founding member of the novel AKR13D subfamily.[129]

3.2 Tropane and nicotine alkaloids (*Datura stramonium*)

The tropane class of alkaloids, which are found mainly in the Solanaceae, contains the anticholinergic drugs hyoscyamine (the racemate of which is called atropine) and scopolamine. Solanaceous plants have been used traditionally for their medicinal, hallucinogenic, and poisonous properties, which are due to their tropane alkaloids. There are many non-tropane toxic alkaloids from Solanaceae, for example solanine, a glycoalkaloid poison found in berries from species such as *Solanum nigrum* or *Solanum dulcamara*, or green potatoes.[198] The narcotic topical anesthetic and central nervous system stimulant cocaine is a tropane alkaloid found outside of the Solanaceae in the Erythroxylaceae *Erythroxylum coca.*

3.2.1 Tropinone reductase-I/II (1AE1, 1IPE, 1IPF, 2AE2, 2AE1). Two tropinone reductases (TRs) TR-1 and TR-II constitute a branching point in the biosynthetic pathway of tropane alkaloids. Tropane alkaloids are of medicinal application, such as hyoscyamine and scopolamine. TRs catalyze NADPH-dependent reductions of the 3-carbonyl group of their common substrate, tropinone, to hydroxy groups with different diastereomeric configurations: TR-I (EC 1.1.1.206) produces tropine (3a-hydroxytropane), and TR-II (EC 1.1.1.236) produces pseudotropine (c-tropine, 3b-hydroxytropane). TR-I and TR-II share 64% identity on amino acid level. The amino acid sequences of these TRs also have the characteristic motifs of enzymes that belong to the short-chain dehydrogenase reductase (SDR) family.[133] The two enzymes TR-1 and TR-II have been crystallised using hanging drop vapor diffusion method in orthorhombic and tetragonal space group, respectively. The crystal structure of TR-1 was solved using the SIRAS method whereas the structure of TR-11 was determined using the MIR method.[134] The two structures are almost indistinguishable from each other in both subunit folding and their association in dimers. Conservation of the subunit structures between TR-I and TR-II was substantiated when the two structures were superimposed by the least squares method using

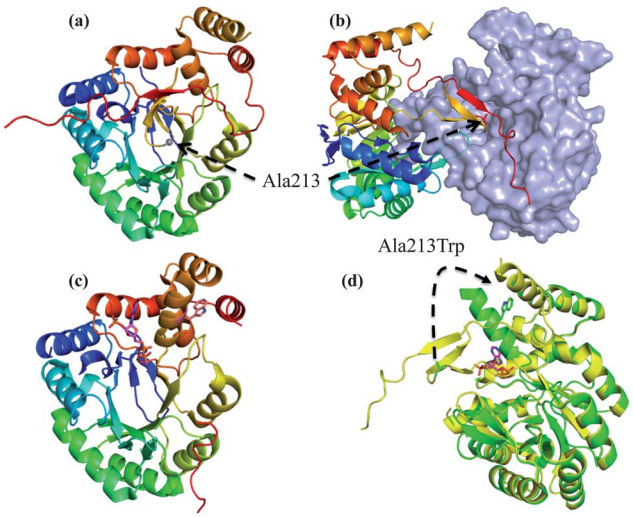

Fig. 8 The overall structure of methylated His_6-PR (PDB code: 3UYI). The structure of methylated His_6-PR displays an unusual α_8/β_6 barrel fold for an AKR enzyme. The residue Ala213 is highlighted as small sphere. N- to C-terminal is colored from blue to red. (b) The two symmetry related molecules from the crystal packing are presented, first molecule (symmetry related) is shown as cartoon with same color code as figure (a) and second molecule as surface representation. The symmetry related molecule shows the blockage of the binding site of the second molecule. The residue Ala-213 on the symmetry molecule is indicated as a stick. (c) Structure of PR-Ala213Trp-NADPH complex (PDB code: 3V0S), binding site is partially occupied by NADPH and the Ala213Trp mutation is highlighted. (d) Comparison between His_6-PR (yellow) and PR-Ala213Trp-NADPH complex (green). The arrow indicates the mutation and change of secondary structure from β-strand to α-helix.

all equivalent Ca positions (rms deviation = 0.78 Å). In the center of the core domain is a seven-stranded parallel beta-sheet, flanked on each side by three α-helices, which constitutes the "Rossmann fold" topology.[135] This core structure is highly conserved among the SDR family members, despite relatively low residue identity between these enzymes. TR-I protein was crystallized in the presence of $NADP^+$, and the bound cofactor molecules in the protein structure are located at the bottom of the cleft between the core domain and the small lobe.[136] The carboxyamide group of the nicotinamide ring is anchored by the main-chain nitrogen and oxygen atoms of Ile-204 and the side-chain oxygen of Thr-206. This tight binding of the carboxyamide group to the protein directs the B-face of the nicotinamide ring toward the void of the cleft, consistent with the observed specificity for the pro-*S* hydride transfer of both TRs.[137] In contrast to TR-I, TR-II did not crystallized in presence of the cofactor. Based on the highly conserved architectures of the cofactor-binding site of the two TRs, it looks as if the TR-II protein binds the cofactor in the same way as TR-I. Although the conformations of the two TR-II side chains (Arg-19 and Arg-41) are very different from those of the corresponding TR-I residues (Lys-31 and Arg-53), the differences are considered to be caused by the binding of the cofactor to TR-I, because these two basic residues have been postulated to be of functional importance in the binding of NADPH preferably to NADH. TR-I catalyzes the reduction of the keto group in tropinone to the hyoscyamine precursor tropine, and TR-II reduces tropinone at the same position to ψ-tropine leading to calystegines.[138]

3.3 Benzylisoquinoline alkaloids

Benzylisoquinoline alkaloids (BIAs) are a complex and diverse group of natural products consisting of more than 2500 known structures, largely restrict to Ranunculales and the Eumagnoliids but also found in distantly related families, including the Papaveraceae, Ranunculaceae, Berberidaceae, Fumariaceae, and Menispermaceae.[139] These compounds are naturally involved in the chemical defense[197] of plants against herbivores and pathogens and they are also pharmacologically active (Fig. 9). Morphine is the most important member of the group of benzylisoquinoline alkaloids and it is a natural product with high medicinal significance. Codeine is used as a cough suppressant; berberine and sanguinarine are used as antimicrobials; papaverine and tubocurarine work as muscle relaxants. Collectively, these alkaloids are found mainly in the Papaveraceae, Ranunculaceae, Berberidaceae, and Menispermaceae; *Papaver somniferum* (opium poppy), *Eschscholzia californica*, *Thalictrum* species, and *Coptis japonica* are the most extensively investigated species.

3.3.1 Norcoclaurine synthase (2VNE, 2VQ5). Like strictosidine in MIA biosynthesis pathways, (*S*)-norcoclaurine is the central precursor in BIA biosynthesis pathways (Fig. 9). The alkaloid is yielded from the first committed step, which consists of the Pictet–Spengler stereospecific condensation of dopamine with 4-hydroxyphenylacetaldehyde (4-HPAA) in the pathway.[140] The enzyme catalyzing this reaction has been identified as norcolaurine synthase (NCS, EC 4.2.1.78). Recently, wild type NCS from *Thalictrum flavum* has been cloned, expressed, purified and crystallised in a trigonal space group.[141] A selenomethionine derivative was overexpressed, purified and crystallized in similar crystallisation conditions and in the same space group as native crystal. Its crystallographic structure has been determined using MAD phasing to resolution 2.7 Å.[142] The crystallographic data, obtained in complex with dopamine, the natural substrate, and *p*-hydroxybenzaldehyde, a non-reactive substrate analogue, provided a snapshot of the initial step of the reaction mechanism. Analysis of these X-ray structures revealed a tetrameric assembly in which the overall fold of each NCS monomer belongs to the Bet v1-like superfamily, which includes plant phytohormone carriers, pathogen-related proteins (PR10), MLN64-START domains, and the recently characterized tetracenomycin aromatase/cyclase. Secondary elements of the enzyme consist of a seven-stranded antiparallel β-sheets wrapped around a long C-terminal helix and two smaller α-helical segments (Fig. 10). Each monomer shows an accessible cleft, located between the seven-stranded antiparallel β-sheets and the three α-helices, that extends through the protein matrix forming long tunnel. The tunnel formed by an array of hydrophobic residues and a polar patch located at the entrance of the cavity. X-ray data obtained on crystals soaked with the dopamine substrate and the nonreactive substrate analogue 4-hydroxybenzaldehyde (PHB) indicates that the two ligands adopt a stacked configuration with the respective aromatic rings lying on almost parallel planes (Fig. 10a–b). The PHB carbonyl oxygen forms a hydrogen bond with Lys-122 amino group, whereas the phenolic oxygen is in

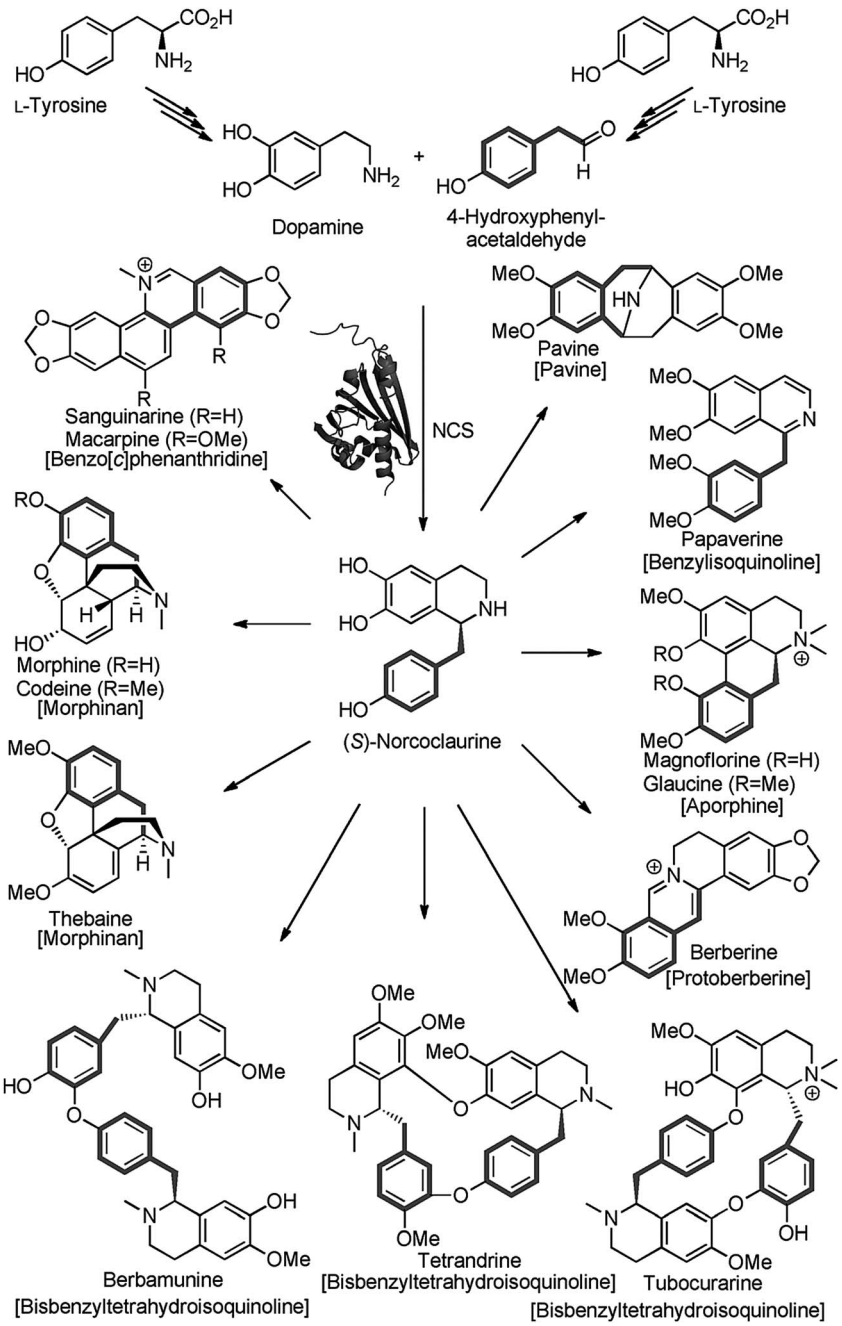

Fig. 9 The key biosynthetic function of NCS and its enzymatic product (*S*)-norcoclaurine together with derived famous alkaloid examples of the large benzylisoquinoline family are summarized (NCS = norcoclaurine synthase) (The figure was reproduced from ref. 161).

contact with the carboxyl moiety of Asp-141. Dopamine is held in place by the stacking interaction with PHB and by hydrogen bonding of the C-1 phenol hydroxyl to the Tyr-108 phenol hydroxyl. Most significantly, the dopamine C-5 carbon atom lies close to the carboxyl group of Glu-110, suggesting a key role for this residue in the catalytic mechanism[142] (Fig. 10b–c).

3.3.2 Berberine bridge enzyme from *Eschscholzia californica* (3GSY, 3FW7, 3FW8, 3FW9, 3FWA). Berberine bridge enzyme (BBE) has a central role in benzophenanthridine biosynthesis and channels its substrate (*S*)-reticuline[143] toward berberine[144] and sanguinarine[145] generation. BBE catalyzes the stereospecific conversion of (*S*)-reticuline to (*S*)-scoulerine[146] (Fig. 11c), thereby forming the so-called berberine bridge (C8) by linking the isoquinoline ring with the benzyl ring system. This reaction requires FAD as a cofactor, which is covalently bound to the enzyme by a cysteine (linked to C6 of flavin) and a histidine (linked to C8α).[147,148] BBE from *Eschscholzia californica* (California poppy) has been expressed in *Pichia pastoris* and purified. Diffraction quality crystals were obtained in monoclinic and tetragonal space groups. The structure was solved by molecular replacement using a partial homology model (comprising only the cofactor binding domain) based on the structure of glucooligosaccharide oxidase from *Acremonium strictum* (PDB-code: 2AXR) as the template.

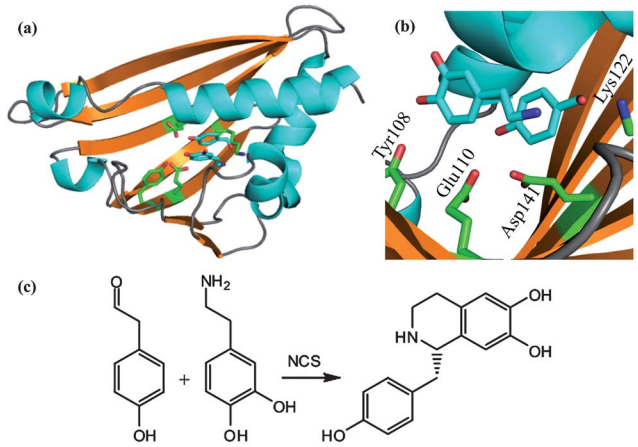

Fig. 10 The overall structure of NCS in the presence of the substrate dopamine and the substrate analogue 4-hydroxybenzaldehyde (PDB code: 2VQ5), secondary structure elements are highlighted in cyan (helix) and in orange (strand). The surrounding residues close to the substrates are shown as ball-and-stick. (b) Close-up view of NCS catalytic site enlarged from figure (a) with a slightly different orentation. (c) Condensation reaction of the analogue 4-hydroxy-benzaldehyde and dopamine catalysed by NCS.

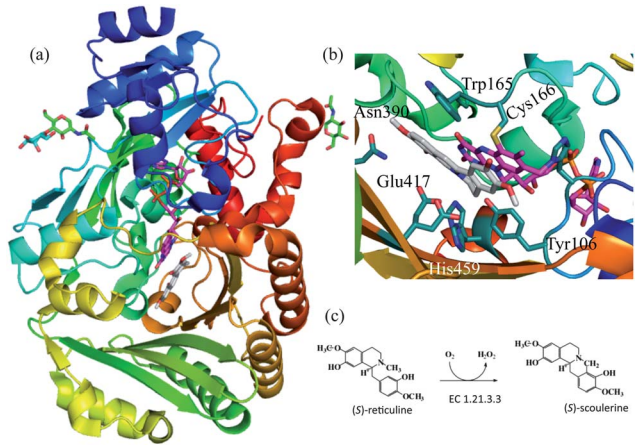

Fig. 11 (a) A cartoon representation of the BBE structure (PDB code: 3GSY) in the presence of FAD (co-factor) and dehydroscoulerine (substrate). The N- to C-terminal is colored from blue to red. The FAD and reticuline are shown in magenta and grey, respectively. N-linked sugar molecules are shown in green whereas BME has been shown in cyan. (b) The binding pocket shown in figure (a), has been enlarged slightly in a different orientation. The co-factor and the substrate are shown as ball-and-stick and the amino acids involved in interaction and covalent binding are shown in cyan. (c) Overall reaction catalyzed by BBE (scheme adapted from ref. 148).

The structure of the complex with (*S*)-reticuline was determined to resolution of 2.8 Å,[149] which shed light on the reaction mechanism. The molecular structure of the enzyme comprises two domains: an FAD binding domain and a central α/β-domain (Fig. 11). The FAD binding domain consists of two N-terminal α/β-subdomains, a C-terminal, mostly α-helical stretch and the central α/β-domain with a seven stranded, anti-parallel β-sheet forming the substrate-binding site. The closest structural neighbors of BBE are members of the *p*-cresol methylhydroxylase (PCMH) superfamily: glucooligosaccharide oxidase from *Acremonium strictum*, 6-hydroxy-D-nicotine oxidase from *Arthrobacter nicotinovorans*, and aclacinomycin oxidoreductase from *Streptomyces galilaeus*. (*S*)-Reticuline is bound in a deep cleft with the phenolic ring pointing toward the bottom of the binding site and is sandwiched between the flavin cofactor and amino acid residues extending from the β-sheet of the central domain (Fig. 11a–b). The active site residue Glu-417 is hydrogen bonded to the C3′OH group of the substrate, and the carboxamide group of Asn-390 interacts with both the OH and the 4′-methoxy group of the phenolic ring (Fig. 11b). The corresponding polar substituents of the isoquinoline moiety are positioned close to Asp-352 at the entrance to the active site.[149]

3.3.3 Salutaridine reductase from *Papaver somniferum* (3O26). SalR belongs to the short chain dehydrogenase/reductase (SDR) family and catalyzes NAD(P)(H)-dependent oxidation/reduction reactions. The SDR family constituting a large protein family contains a single domain. The domain composed of a parallel α/β-fold with a Rossmann fold. At present, roughly 3000 members are known from all living organisms and they exhibit a wide substrate spectrum, including alcohols, sugars, steroids, aromatic compounds, and xenobiotics.[150] SDRs consist of a one-domain subunit of about 250 amino acids with the cofactor binding site in the N-terminal part and substrate binding in the C-terminal part.[133] The two main characteristics of this protein family are the highly conserved TGxxxGhG motif for coenzyme binding and the YxxxK motif, which, together with an upstream Ser residue, represents the catalytic center.[151] The enzyme reduces the C-7 keto group of salutaridine to the C-7 (*S*)-hydroxyl group of salutaridinol in the biosynthetic pathway that leads to morphine in the opium poppy plant *Papaver somniferum*. The cDNA of salutaridine reductase (SalR), reducing the keto group of salutaridine to a hydroxyl as an intermediate step in morphine biosynthesis, was isolated and identified as a member of the classical SDRs with preference for NADPH as cofactor.[152] His-tagged SalR native protein was expressed in *E. coli* BL21 (DE3) Codon Plus RIL cells and the protein purified after thrombin cleavage using Talon and Benzamidine-Sepharose and gel-filtration.[153] Selenomethionine substituted SalR was produced in a similar way as the native protein and crystallised in presence of 0.1 M MES (pH 6.0–6.6), 1.9 M ammonium sulfate, 5% (v/v) PEG 400, 0.1 M LiCl, and 3% (v/v) glycerol.[153] The structure of SalR was solved using SAD phasing from the X-ray data collected from crystals of SeMet-substituted SalR co-crystallized with NADPH.[154] The core structure is highly homologous to other members of the short chain dehydrogenase/reductase family. The structure of SalR is similar to the structure of carbonyl reductase except a part of the structure (residues 105–140) (Fig. 12a). This insertion does not exist in carbonyl reductases (PDB code 1WMA).[154] The rmsd between structure of carbonyl reductase and SalR over 247 residues is 1.60 Å. The bound NADPH in SalR is located at an almost identical position as in the carbonyl reductase and it is buried in the core of the enzyme. Some of the residues interacting with the NADPH are shown in Fig. 12b. Based on the computer modeling and mutagenesis studies, the substrate binding has been proposed. Based on these studies, it is predicted that SalR may undergo significant conformational changes during catalysis.[154]

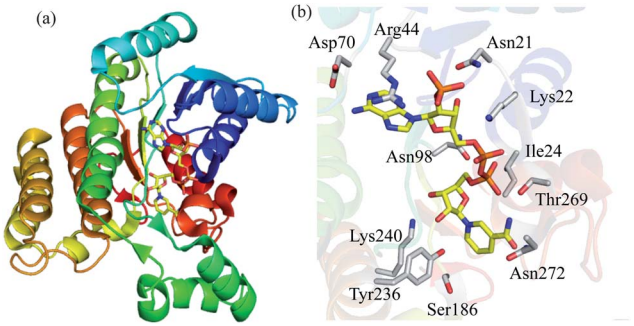

Fig. 12 (a) A cartoon representation of the SalR structure (PDB code: 3O26) looking at the active site. The N- to C-terminus is colored from blue to red. (b) The active site from the figure (a) has been zoomed in a different orientation, showing NADPH and its neighbouring residues in ball-and-stick.

3.3.4 Pavine *N*-methyltransferase from *Thalictrum flavum*. *S*-Adenosyl-L-methionine-dependent *N*-methyltransferases (NMTs) are common in plant secondary metabolism and catalyze key reactions in alkaloid biosynthesis.[139] A cDNA from the plant *Thalictrum flavum* encoding pavine *N*-methyltransferase, an enzyme belonging to a novel class of *S*-adenosylmethionine-dependent *N*-methyltransferases specific for benzylisoquinoline alkaloids, has been heterologously expressed in *Escherichia coli*.[155] The enzyme was purified using affinity and gel-filtration chromatography and was crystallized in space group $P2_1$. The structure was solved at 2.0 Å resolution using a xenon derivative and the single isomorphous replacement with anomalous scattering method[156] using AUTO-RICKSHAW. Structural analysis of the structure of PavNMT and comparison with those of other methyltransferase structures will provide insight into the molecular mechanism of this important enzyme.

4 Product and substrate specificity of enzymes involved in alkaloid biosynthesis pathways

The crystal structures of STR1 in complex with its natural substrates tryptamine and secologanin provide structural understanding of the observed substrate preference and identify residues lining the active site surface that contact the substrates. Recently, the crystal structure of STR1 complex with its product strictosidine was also solved.[106] Based on structures of substrates and product-bound STR1, the first rational site-directed mutagenesis experiments were performed with the enzyme. In the STR1-strictosidine complex, the indole ring of the tryptamine unit is located in the hydrophobic pocket, lined by six residues (Phe226, Val208, Val167, Trp149, Tyr151, including Gly210). The five hydrophobic residues are either invariant or conservatively substituted by other hydrophobic amino acids in STR1 from different plant species. The tryptamine part is stacked between two aromatic residues, Tyr151 and Phe226 keeping the indole ring in place *via* π–π interaction and *via* van der Waals interactions with other three hydrophobic amino acids Val208, Val167 and Trp149. These residues block substituents at C-10 and C-11 of strictosidine (at C-5 and C-6, respectively, in tryptamine). Tryptamine derivatives with bulky groups at these positions therefore act as poor substrates. The crystal structure of STR1 complex with tryptamine demonstrates that the Val208 side-chain is shielding the 5-position, preventing binding of 5-substituted indole bases. Replacement of this particular Val by the smaller Ala broadens the substrate acceptance for 5-substituted tryptamines, delivering novel strictosidine analogues.[106]

The crystal structure of the SG-strictosidine complex from *Rauwolfia serpentina* (PDB code 2JF6) indicates that the indole portion of strictosidine points toward the surface of the enzyme whereas the glucose moiety of strictosidine is buried within the enzyme's active site.[113] The results of these substrate specificity studies suggest that the active site of SG has not evolved to discriminate against substitutions on the indole ring. Notably, 5-methyl and 6-methyl tryptamine are not turned over by the enzyme strictosidine synthase to form strictosidine analogs.[106,157,158] The specificity of the early stages of the MIA pathway therefore appears to be controlled in large part by strictosidine synthase and not by SG.[157] SG and RG share high sequence identity and structural similarity. Two similar enzymes with different biosynthetic functions in one plant species have evolved to catalyze two distinct reactions. In addition to its natural substrate raucaffricine, RG is able to hydrolyse the glucoalkaloid strictosidine (relative activity 1.2%), whereas SG does not exhibit any measurable conversion of raucaffricine. Crystal structures of both enzymes reveal their differences in substrates specificity.[113,124] Structural analysis of the enzymes indicates a "wider gate" of RG that allows strictosidine to enter the catalytic site, whereas the "slot-like" entrance of SG prohibits access by raucaffricine. Trp392 in RG and Trp388 in SG control the gate shape and acceptance of substrates whereas Ser390 directs the conformation of Trp392. Crystal structures, site-directed mutations and kinetic data of RG and SG, provide a structural and catalytic explanation of substrate specificity and deeper insights into *O*-glucosidase chemistry.[124]

5 Elucidation of enzyme mechanism

5.1 Acid/base catalysis

Proton transfer is the most common reaction that enzymes perform. Most enzyme reactions occur by ionic mechanisms, involving the creation or disappearance of charge. Such reactions are typically acid or base catalysed, but acid and base concentrations are minimal under physiological conditions near pH 7. Enzymes have evolved subtle and highly effective solutions to this problem, involving general acid and general base catalysis by the functional groups accessible on the side-chains of amino acids strategically placed in their active sites. Structural analysis of VS and its mutagenesis studies confirm that His-160 is a catalytic residue.[105] It acts as a base and abstracts the proton of the hydroxy group of 17-deacetylvinorine. The 17-oxygen will then attack the carbonyl carbon of acetyl-CoA, resulting in its acetylation and release of CoA. Most likely the reaction proceeds without formation of an acetylated enzyme intermediate (Fig. 4c).

Structure based sequence alignment of SG with glucosidases of various origins indicates complete conservation of amino acid residues Glu-207, Glu-416, and His-161. Based on earlier site-directed mutation experiments in the glucosidase family, Glu-207 is the proton donor that allows nucleophilic attack of Glu-416 at

the anomeric carbon C-1.[159] Both glutamic acids catalyze the concerted hydrolysis of the glucoside bond. The residues Glu-207 and Glu-416 are located within the pocket near the sugar moiety, leaving a distance of 5.2 Å between their carboxyl carbons (Fig. 3b). This distance offers enough space for substrate entry and to place the glucosidic bond in an optimal position for hydrolysis. In fact, mutations Glu207Gln, Glu207Asp, Glu416Gln, and Glu416Asp resulted in loss of enzyme activity, highlighting the importance of both glutamates for the deglucosylation of strictosidine.[113] Although His-161 is not a catalytic residue, mutation of the residue(s) to Leu or Asn nevertheless results in loss of enzyme activity. The structure of SG-strictosidine complex highlights the importance of the residue in holding strictosidine in the correct orientation for deglucosylation as the residue is located at 5.8 Å away from the anomeric C-atom of the glucose moiety but forms a hydrogen bond to O3 of the sugar moiety of the substrate (Fig. 3b).

The corresponding catalytic residues of SG are also conserved in the RG enzyme. The RG catalyzes similar reaction as the SG enzyme for its substrate raucaffricine.

The active site residue Glu417 in BBE is hydrogen bonded to the C3'-OH group of the substrate. It has been determined based on the structure, mutagenesis studies and kinetic measurements that the residue Glu417 is an essential amino acid for substrate oxidation and acts as a catalytic base for C–C bond formation and concomitant transfer of a hydride to the flavin indicating flavin reduction and berberine bridge formation for the concerted processes that is initiated by proton abstraction of the phenolic hydroxyl group.[148]

5.2 Pictet–Spengler reaction

Despite the difference in sequence and in structure of STR1 and NCS, the enzymes share a common chemical reaction known as Pictet–Spengler reaction[160] (for a review see ref. 161) in which a β-arylethylamine, such as tryptamine, undergoes ring closure after condensation with an aldehyde or ketone. As outlined in a recent review, it has remained an important reaction for the synthesis of indole and isoquinoline alkaloid.[161] Based on the structural and enzyme kinetic studies of NCS and STR1, key residues involved in the Pictet–Spengler reaction has been identified. NCS and STR1 adopt different mechanisms to achieve the Pictet–Spengler cyclization. NCS employs the positive charge of Lys122 as a strong polarizing agent for the carbonyl group of the aldehyde substrate upon which the amine substrate acts as a nucleophilic agent (Fig. 10c). Conversely, STR1 employs the negatively charged carboxyl moiety of Glu309 to hold in place and eventually deprotonate the nitrogen atom of the amine substrate that subsequently reacts with the incoming aldehyde substrate.

5.3 Covalent catalysis

PNAE performs catalysis of its substrate PNA *via* covalent binding with one of the catalytic residues. The covalent binding of the product intermediate has been trapped using the inactive mutant His244Ala although the mutant was co-crystallised with its substrate. The crystal structure of the mutant-complex reveals the enzyme mechanism. The analysis of the PNAE structure shows that active centre consists of three residues Ser87, Asp216 and His244 forming the catalytic triad (Fig. 6b). The crystal structure of His244Ala in complex with enzyme product 3 reveals that the indolic part in the molecule interacts with Met113, Phe125, Tyr128, and Lys187. This arrangement enables the alkaloid to be fixed by hydrophobic, sandwich-like interactions providing optimal structural accommodation for catalysis. The optimized geometry of the active center, the substrate as well as the correct positioning of the C17 aldehyde group are crucial for PNAE activity since any slight changes in structure or functionality result in loss of enzymatic hydrolysis. After Ser87-assisted hydrolysis (Fig. 6c), the reaction intermediate decarboxylates to the enolized enzyme product 3, in which the enolate anion is stabilized by hydrogen bonds of the backbone amides of Gly19 and Phe88 forming an oxyanion hole.[117]

6 Rational structure-based enzyme engineering

Molecular details of proteins are determined by their three-dimensional (3D) structure, with the precise configuration of specific amino acid residues contributing to the functional site(s) within the protein. In structure-based protein engineering, appropriate sites are selected for mutation based on analysis of the 3D structure of the protein and the mutants characterized for desired activity. Despite the recent movement toward directed evolution methods for redesign of proteins, it would be insufficient to discount the power of rational concepts utilizing 3D structures and/or homologous sequences. Indeed, in concert with random mutagenesis and directed evolution methods, structure-based protein engineering is a powerful approach.[162] Over the years, a number of examples of rational engineering for the elucidation of enzyme mechanisms, changing substrate specificity, cofactor specificity, introduction of metal binding sites for affinity purification, stabilization and allosteric control *etc.*, have been described (for a review, see ref. 161). Elucidation of various enzyme reaction mechanisms from multiple alkaloid biosynthesis pathways have been derived from it crystallographic structural studies combined with analysis of homologous sequence and examining the enzyme activity of the predicted mutants. For example, the 3D-structure of VS confirms the functional importance of His-160 in the mutagenesis studies for VS activity[116] and has been shown that how the enzyme catalyzes acid/base reaction (Fig. 4c), similarly 3D structures of the SG, RG, PR, TR-I/II and BBE helped to understand its acid/base catalysis. The crystal structure of PNAE inactive mutant complex with the substrate reveals covalent binding of the substrate to Ser87 and shows that the nucleophilic residue reacts with the substrate through a nucleophilic attack, confirming the covalent catalysis of the enzyme (Fig. 6c). The Pictet–Spengler reaction in plant system has been described using the crystal structures of STR1 and NCS. The structure-based engineering of STR1 is a classic example of redesign of substrate specificity with great potential for changing metabolic pathways and generating novel molecules against infectious bacteria and viruses. The enzyme catalyzes the condensation of tryptamine and secologanin leading to the synthesis of numerous monoterpenoid indole alkaloids in higher plants. Using the crystallographic structure of strictosidine synthase in complex with strictosidine, Loris *et al.* (2007)[106] produced and characterized mutants with the capacity

to generate novel alkaloid libraries for pharmacological screening (see below the next section). This example demonstrates potential of the knowledge-based protein engineering of strategic target molecules to alter pathways and make new products. However, it is important to bear in mind that an inherent limitation of structure-based rational mutagenesis is its inability to consider the contributions of side chains that may be distal to the functional site of the molecule. Nevertheless, the structure of a protein is invaluable to a mechanistic understanding of protein function and can provide the template for further engineering by rational or combinatorial genetic approaches.

7 Novel alkaloid generation

Based on the 3D-structures of the STR1-tryptamine and the STR1-strictosidine complexes, a model of STR1-5-methoxytryptamine was generated, and it was noticed that the methoxy-group of the substrate is in close vicinity to Val208 (at a distance of 4 Å from the 5-position). In order to design a His-tagged STR1 mutant which can accept 5-methoxy- or 5-methyltryptamine as a substrate, the bulky Val208 was replaced by Ala. The best substrate for STR1 is tryptamine, benzene-ring-substituted tryptamines generally react at less than 10% of the rate for tryptamine. Interestingly, the resulting mutant STR1-Val208Ala exhibits conversion of 5-methoxy- and 5-methyltryptamine to the 10-methyl- and 10-methoxystrictosidines in the presence of secologanin, which was confirmed using HPLC and mass-spectrometry,[106] indicating the importance of knowing the 3D structure of STR1 for redesigning its enzyme activity. Moreover, the same STR1 mutant (Val208Ala) converted the 6-methyl- and 6-methoxytryptamines into the corresponding strictosidines more efficiently than the wild-type enzyme.[106] Substitution of amino acid Val208 highlights its critical role in substrate recognition at the "western part" of the indole moiety of strictosidine. The STR1 mutant Val208Ala retains its enantioselectivity. ^{1}H NMR analysis of the 10-methylstrictosidine-lactam-tetraacetate obtained from methylated strictosidine by base-catalyzed lactamization and acetylation showed one of the four acetyl signals significantly shifted to higher field (1.18 ppm compared to the remaining three signals between 1.86 and 2.05 ppm), which clearly indicates the 3α-(*S*)-configuration. It should also be emphasized that the methyl ester group and the vinyl side chain of secologanin can, to some extent, be modified to more flexible, longer chains. The strictosidine analog can also be converted into new heteroyohimbine alkaloids, such as serpentine derivatives.[163,164] The array of such products from STR1 mutants can be further extended by the additional generation of SG mutants from *R. serpentina* or other plant species. The combination of various STR1 and SG mutants based on the structural knowledge of the enzymes together with the whole range of commercially available primary amines and secologanin may lead to significant expansion of alkaloid diversity for the approaches described here. Such a combinatorial strategy would represent a chemoenzymatic concept for the generation of large and diverse libraries containing thousands of heteroyohimbine- type alkaloids, potentially exhibiting pharmacological activities. For example, the heteroyohimbine ajmalicine shows positive effects on postischemic hypoxia and cerebral protection, but there is still a therapeutic need for novel drugs with advanced activities.[165–167] It should be mentioned that also new enzyme functions can be detected (*e.g.* for STR1), which allow construction of alkaloid libraries based on the rare piperazino-indole framework,[168] extending significantly the synthetic application of biosynthetic enzymes ("white chemistry").

8 Metabolic engineering of alkaloid biosynthesis pathway

The activity of many natural products can be modulated or improved by subtle changes in chemical structure. While synthetic chemistry can be used to introduce certain changes, hijacking the biosynthetic machinery to generate analogs provides an attractive strategy to ferment these "unnatural natural products". Metabolic engineering efforts to manipulate the biosynthetic machinery to produce unnatural analogs requires a working knowledge of the biosynthetic pathway at the enzymatic and genetic level.[144] Therefore, metabolic engineering of the iridoid-derived monoterpene indole alkaloids relies heavily on the identification and characterization of the genes and corresponding enzymes responsible for producing these compounds.

A number of monoterpene indole alkaloid analogs have improved or altered biological activity. For example, topotecan, a derivative of camptothecin, and vinorelbine and vinflunine, derivatives of vinblastine, are highly successful chemotherapies.[169–172] Manipulation of biosynthetic pathways is a powerful way to make unnatural natural products that are not easily accessible *via* total synthesis. The manipulation of pathways to make unnatural variants of natural compounds, a process often termed combinatorial biosynthesis, has been robustly successful in prokaryotic systems. The development of approaches to generate new-to-nature compounds from plant-based pathways is, in comparison, much less advanced.[173] Its success will depend on the specific chemistry of the pathway, as well as on the suitability of the plant system for transformation and genetic manipulation. Importantly, many metabolic engineering efforts must be performed within the native plant producer, since most plant biosynthetic pathways are incompletely elucidated at the genetic level. As pathways are elucidated, and plant-derived pathways can be heterologously expressed in hosts that are more amenable to genetic manipulation, biosynthetic production of new-to-nature compounds from plant pathways will become more widespread.

When hairy roots cultured in media supplemented with tryptamine analogs were analyzed, LC-MS identified alkaloid derivatives with molecular weights corresponding to the addition of a fluorine, hydroxyl, or methyl group on the indole ring of the unnatural tryptamine substrate, clearly indicating that the monoterpene indole alkaloid pathways can turn over a broad range of tryptamine analogs. The intensities of the MS signals assigned to the alkaloid derivatives suggest that the major products correspond to compounds derived from the unnatural starting material, with the parent (natural) alkaloids present in lower quantities for certain analogs at 1 mM concentration. Hairy root culture extracts could also be fractionated by preparative HPLC and several of the most abundant analogs were purified in milligram quantities and characterized by NMR. *C. roseus* seedlings could be aseptically germinated on solid

medium containing 1 mM of the desired tryptamine analog; incorporation into the monoterpene indole alkaloid pathways was observed as evidenced by LC-MS.

In short, *C. roseus* can produce an array of iridoid-derived alkaloid analogs from unnatural starting materials. Furthermore, the electronic and steric properties of the non-natural substrates impacted how these substrates partitioned among the branches of the monoterpene indole alkaloid pathway of this medicinal plant. For example, secologanin analogs, and tryptamine analogs with substituents at the 4-position, were primarily incorporated into heteryohimbine pathways; 5-substituted tryptamine analogs had an incorporation profile most similar to natural tryptamine; 6- and 7-substituted tryptamines favored incorporation into the strychnos type alkaloids. This different partitioning among the branches of the pathway may lend insight into the mechanism or specificity of downstream enzymes.

In precursor-directed biosynthesis, the producer organism is supplemented with analogs of the naturally occurring starting materials. These non-natural starting materials are, in turn, converted into the corresponding unnatural products. However, the yield and purity of these unnatural products is improved if the biosynthesis of the natural starting material is genetically blocked, and the producing organism is forced to utilize exogenously supplied precursors exclusively for product biosynthesis. This strategy, termed mutasynthesis, was first applied several decades ago to yield novel antibiotics in the soil bacterium *Streptomyces*,[174] and has proven to be highly successful in microbial systems.[175]

Mutasynthesis can also be applied to the monoterpene indole alkaloids in *C. roseus*.[176] Tryptamine, the starting substrate for all monoterpene indole alkaloids,[177,178] is produced from tryptophan by tryptophan decarboxylase.[179,180] If tryptamine biosynthesis is blocked, alkaloid biosynthesis could, in principle, be rescued by introducing exogenous tryptamine or tryptamine analogs to plant cell cultures. Tryptophan decarboxylase was targeted for gene silencing (RNAi) to prevent formation of tryptamine. The plasmid designed to suppress tryptophan decarboxylase was introduced into *Agrobacterium rhizogenes*, which was then used to infect *C. roseus* seedlings to generate hairy root cultures.[181,182] Hairy root lines harboring the silencing plasmid were cultured in liquid medium, where production of all major tryptamine derived alkaloids was substantially decreased in the five representative silenced lines examined. RT-PCR indicated that the expression levels of tryptophan decarboxylase were substantially reduced in cultures harboring the silencing plasmid.

The tryptamine analog 5-fluorotryptamine was chosen to illustrate the potential of mutasynthesis for the monoterpene indole class of alkaloids. A silenced hairy root culture was incubated with varying concentrations (250–2500 μM) of 5-fluorotryptamine. The complex mixture of alkaloid products was greatly simplified, since no natural alkaloids derived from endogenous tryptamine were present. The levels of certain fluorinated alkaloids (ajmalicine and catharanthine) were greater in the silenced culture, compared to precursor directed biosynthesis with a wild type hairy root line, suggesting that some pathway branches could support increased production levels of the desired nonnatural compounds when not challenged with competing natural substrate.

Producing a wide variety of natural product-based compounds by fermentation is an attractive way to obtain a diversity of natural product-inspired molecules. We envision that RNA silencing methods can also be used with additional unknown downstream biosynthetic enzymes, as they become identified, to more effectively tailor the production of the desired unnatural alkaloid in plants and plant culture.

Precursor-directed biosynthesis and mutasynthesis rely on the inherent ability of biosynthetic enzymes to accept non-natural substrates with altered steric and electronic properties. Many biosynthetic genes, however, have tight substrate specificity and will not turn over the desired unnatural substrate analog, as described above. In these cases, alternative biosynthetic genes with altered substrate specificity can be incorporated into existing natural product pathways, enabling the production of unnatural compounds from non-natural starting substrates. This approach has been applied to many microbial systems,[176] but again, the application of this general strategy to plant-derived natural product pathways has been limited.

Precursor-directed biosynthesis studies in *C. roseus* have revealed that one key bottleneck in the production of unnatural monoterpene indole alkaloids in this plant is the stringent substrate specificity of strictosidine synthase, the enzyme that catalyzes formation of the biosynthetic intermediate strictosidine from tryptamine and the iridoid secologanin. Notably, tryptamine analogs with substituents at the 5 position of the indole ring are not accepted by this enzyme. The recently reported crystal structure of strictosidine synthase[105] has enabled design of enzyme mutants with broadened substrate specificities, allowing enzymatic production of a greater variety of strictosidine analogs.[106,183] Specifically, mutation of Val214 allows turnover of a variety of 5-subsituted tryptamine analogs in the corresponding strictosidine products.

The strictosidine synthase mutant gene containing the point mutation Val214Met was introduced into hairy root culture under the control of the strong, constitutive 35S CMV promoter.[182] Transgenic hairy root lines harboring the Val214-Met mutant enzyme were cultured in the presence of 5-chlorotryptamine, 5-methyltryptamine, or 5-bromotryptamine, which are only turned over by the Val214Met mutant enzyme, and are not recognized by natural strictosidine synthase.[183] After one week of culture, LC-MS analysis of the extracts of this plant tissue indicated appearance of novel compounds derived from the exogenous substrates. Control experiments clearly indicated that these compounds were not present when the tryptamine analog was absent from the medium, or when *C. roseus* tissue was transformed with the wild-type strictosidine synthase gene.

Fermentation is a powerful strategy for large-scale production of natural product analogs. This example demonstrates that genetically reprogramming alkaloid metabolism can be achieved in medicinal plant cell culture, even when the genetic, biochemical and regulatory aspects of the pathway are incompletely characterized. Optimistically, this study sets the stage for further metabolic engineering efforts to improve the scope and practicality of unnatural product biosynthesis in plants.

Engineering of metabolic pathways in plants poses significant challenges when compared to the extraordinary advances that have been achieved in prokaryotic systems. Nevertheless, plant metabolic engineering has made substantial strides over the last

decade, and the work on the iridoid-derived monoterpene indole alkaloid systems has comprised a large part of these efforts. Moreover, the increasing speed at which plant pathways are being elucidated bodes well for the future of plant-derived natural products. As genomic technologies have improved, the genetic basis of the natural product chemistry catalyzed within medicinal plants has slowly – but surely – begun to emerge. This genetic information allows detailed mechanistic explorations of plant pathways, and also enables metabolic engineering efforts that could improve production of medically important plant natural products. This review highlights a number of methods that have been used to engineer plant biosynthetic pathways, with an emphasis on the well-studied monoterpene indole alkaloids of *Catharanthus roseus*. Finally, as heterologous reconstitution of plant pathways into tractable host organism such as tobacco or yeast become more routine, this will further enhance our ability to hijack the iridoid and downstream alkaloid pathways.

9 Structural relationships and evolution of enzymes and genes of alkaloid biosynthesis

Although the secondary metabolites occurs in plants with high structural diversity, their formation from abundant primary metabolites is managed by only a few types of reactions, for example condensations, cyclizations, hydroxylations, methyl- and acetyl transfers, to name the most common. Since secondary metabolites are defined by their occurrence in only restricted plant groups or species,[184] the pathways leading to their formation are thought to have evolved basically from primary metabolism in response to specific environmental factors. It is generally believed that novel gene functions arise through gene duplication followed by divergence of gene function.[185] Moreover, single protein domains can also be multiplied, shuffled within a genome and combined in different ways, again giving rise to the evolution of novel protein functions.[186] To analyse a given metabolic pathway or network and to determine its evolutionary relationship to other metabolic pathways it is therefore mandatory, not only to analyse the biochemical reactions catalyzed by the proteins or to determine the enzyme-encoding gene sequences but also to have structural data of these pathway enzymes. The main reason for this requirement is that below a certain level of similarity (30%) the detection of homology between enzymes may not be feasible. Complex multiple sequence alignments as well as the comparison of three dimensional structures could circumvent this obstacle,[187] especially since it became clear that the structure of a protein is more highly conserved throughout evolution than the protein's sequence.[188]

In the case of alkaloid biosynthetic pathways the increasing amount of structural data only recently allows us to have a first look at the potential evolution of these biosynthetic pathways, at least at the level of single enzymes. An intriguing example for a so far not obvious evolutionary relationship is that of strictosidine synthase (STR1). On a sequence level homology for *str1* could only be determined to strictosidine synthases from related *Rauvolfia* species or for the corresponding enzymes from *Catharanthus* and *Ophiorrhiza*.[189] Other genes with sequence similarity to *str1* from GeneBank or with the assigned name strictosidine synthase lacked functional studies and therefore, raised doubt over their functional relatedness to *str1*. However, when the structure of STR1 was solved and reported to be a comprised of a six-bladed β-propeller, a first hint of the evolution of this unique enzyme became evident.[105] This unique fold groups the enzyme into a structurally similar but functionally highly diverse group of proteins with widespread phylogenetic occurrence.[190] The β-propeller fold is constructed from a modular building block – a four-stranded β-sheet – repeated four to eight times around a central axis. This modular and partially repetitive structure suggested that the common ancestors are primordial β-sheet domains that were fused and duplicated and acquired various functions. Here it becomes clear that genes from plant secondary metabolism are not necessarily evolutionarily derived from genes of primary metabolism but might have evolved as new functions from modular domain building blocks. Among the subgroup of the six-bladed β-propeller structures, the closest alignment of STR1 could be performed with phylogenetically and functional unrelated proteins, *e.g.* diisopropylfluorphosphatase from the squid *Loligo vulgaris*,[191] the serum paroxonase,[192] or the low density lipoprotein receptor (LDLR) YWTD domain.[193] The latter contains an internal repetitive sequence (XYWTD) which shows the relatedness of the building blocks comprising the protein structure. For STR1 a similar motif could be identified in β-B3 (XYFTD), which gives a hint to a common evolutionary relationship, showing divergent evolution from a common ancestral β-sheet gene.

Especially interesting is the question of whether a similar fold has been adopted by the other STR1 enzymes which catalyze the Pictet–Spengler reaction between secologanin and tryptamine in numerous other plant species. So far only five of these enzymes have been described in detail, three from different *Rauvolfia* species, one from *Catharanthus roseus* and one from *Ophiorrhiza pumila* (summarized in ref. 189). However, many more plant species contain monoterpenoid indole alkaloids derived from strictosidine and it will be of interest to elucidate the structure–function relationship once the sequences and characteristics of the enzymes becomes available.

Structural data of alkaloid biosynthetic enzymes can also prove that although there is a functional relationship between two enzymes, they might not exhibit any obvious evolutionary relationship. The enzyme (*S*)-norcoclaurine synthase (NCS) catalyzes a similar Pictet–Spengler reaction to STR1, in that it forms from dopamine and *p*-hydroxyphenylacetaldehyde (*S*)-norcoclaurine, therefore initiating the biosynthesis of benzyl-isoquinoline alkaloids.[194] Structure elucidation of NCS from *Thalictrum flavum* shows that the enzyme adopts an overall fold of a Bet v1 protein.[142] Moreover, the structural data indicate that both enzymes, STR1 and NCS, employ different mechanisms to perform the asymmetric condensation of the aldehyde and the amine. In NCS Lys-122 acts as a strong polarizing agent for the carbonyl group that allows the nucleophilic attack of the amine moiety of dopamine.[195] In contrast, in STR1 the negatively charged carboxyl group of Glu-309 deprotonates the amine, which subsequently reacts with the aldehyde.[107] These data are only available through thorough structural and biochemical analysis and could not be predicted from sequence data alone. As a consequence, the different protein folds adopted by STR1 and NCS and the variable onset of the Pictet–Spengler reaction

suggest that the different entry steps to alkaloid biosynthetic pathways for the benzylisoquinoline and monoterpenoid indole alkaloids might have evolved independently and not from an ancestor.

The structural data obtained from alkaloid biosynthetic enzymes can also give an inside into the unique substrate preference of some of the catalysts, as exemplified by the analysis of strictosidine glucosidase (SG) and raucaffricine glucosidase (RG) from *R. serpentina*. Both enzymes catalyze the hydrolyzation of a β-glucosidic bond, SG converting strictosidine to its reactive aglycon and RG hydrolyzing the side product of the ajmaline pathway – raucaffricine – to the intermediate vomilenine, which subsequently can be re-routed to the ajmaline biosynthesis. While SG occupies the central role in this *Rauvolfia* metabolic pathway and presumably many related pathways in other plants, RG catalyzes a very special branch reaction. Both enzymes share a sequence identity of 52% and exhibit different substrate preferences. While SG only accepts strictosidine as a substrate and not raucaffricine, RG is slightly more promiscuous in that it also accepts strictosidine as well as its main substrate raucaffricine. Both enzymes are structurally highly related, bearing a TIM $(\alpha/\beta)_8$ fold of the GH-1 family of glycosyl hydrolases.[124] Also the catalytic residues, Glu186 and Glu420 are identical, proving that the mechanism of action is the same. A striking difference lies in the position of the tryptophane side chain, which occupies identical positions in both RG (Trp392) and SG (Trp388). However, the side chain adopts inverse conformations, restricting the binding of the appropriate substrates. In SG, the conformation of Trp388 allows a tight, sandwich-like binding of the aglycon *via* hydrophobic interactions and also restricts the binding of other substrates. In RG, the perpendicular orientation of Trp392 with the raucaffricine aglycon results in weaker binding and a reduced K_m compared to that of SG with its substrate. Nevertheless, the specific orientation of Trp392 in RG allows promiscuous binding of strictosidine, and indeed mutation of the tryptophan to alanine (RG-W392A) reduced activity against raucaffricine but also increased activity against strictosidine.

Comparing the *Rauvolfia* β-glucosidases, eleven plant β-glucosidases for which structural information is available thus far all show a conformation of the central tryptophan residue comparable to that of RG. Only SG differs in this respect. Taken together the structural data on biosynthetic enzymes of plant secondary metabolism gives valuable insight into catalytic mechanisms and evolution of these interesting molecules and provides further understanding of the biosynthetic capacity of higher plants.

10 Summary

Structural analysis of alkaloid biosynthetic enzymes has enabled modification of the substrate binding sites for the development of biomimetic alkaloid production. It has expanded our understanding of the enzyme mechanism, substrate specificity and structural relationships. The structural biology has advanced alkaloid biosynthesis research in the understanding of a number of enzymes, which open the door for metabolic engineering and the generation of novel alkaloids. Future structural biological studies have the potential to significantly advance our fundamental understanding of the alkaloid biosynthesis pathways at the molecular level in order to identify key residues for reaction catalysis and substrate specificity. This offers new enzyme redesign perspectives for the production of large libraries of alkaloid analogs by combinatorial enzyme-mediated approaches.

11 Acknowledgements

We thank Deutsche Forschungsgemeinschaft (Bad Godesberg, Germany); Fonds der Chemischen Industrie (Frankfurt/Main, Germany); K. P. Chao's High-Tech Foundation at Zhejiang University (Hangzhou, China); Australian Synchrotron and COST Action FA1006, Plant Metabolic Engineering; for financial support of our research programs. We also acknowledge staff members at the EMBL-Hamburg (DESY, Hamburg, Germany), SLS (Paul Scherrer Institute, Switzerland), BESSY (Berlin, Germany) and ESRF (Grenoble, France) beamlines. All of our present and past co-workers are highly appreciated.

12 References

1 *The Alkaloids, Chemistry and Physiology*, ed. R. H. F. Manske, Academic Press, New York, 1981, vol. XX.
2 M. H. Zenk and M. Juenger, *Phytochemistry*, 2007, **68**, 2757–2772.
3 A. I. Scott and S. L. Lee, *J. Am. Chem. Soc.*, 1975, **97**, 6906–6908.
4 J. Stoeckigt, J. Treimer and M. H. Zenk, *FEBS Lett.*, 1976, **70**, 267–270.
5 J. Stoeckigt, T. Hemscheidt, G. Höfle, P. Heinstein and V. Formacek, *Biochemistry*, 1983, **22**, 3448–3452.
6 M. H. Zenk, *Phytochemistry*, 1991, **30**, 3861–3863.
7 R. B. Herbert, The Biosynthesis of Terpenoid Indole Alkaloids, in *Monoterpenoid Indole Alkaloids, The Chemistry of Heterocyclic Compounds*, ed. J. E. Saxton, John Wiley and Sons, Chichester/New York/Brisbane/Toronto/Singapore, 1994, vol. 25 (supplement), pp. 1–13.
8 J. Stoeckigt, Biosynthesis in *Rauvolfia serpentina* – Modern Aspects of an Old Medicinal Plant, in *The Alkaloids*, ed. G. A. Cordell, Academic Press, San Diego, California, 1995, vol. 47, pp. 115–172.
9 M. Ruppert, X. Y. Ma and J. Stoeckigt, *Curr. Org. Chem.*, 2005, **9**, 1431–1444.
10 M. H. Zenk, *Pure Appl. Chem.*, 1994, **66**, 2023–2028.
11 T. M. Kutchan, N. Hampp, F. Lottspeich, K. Beyreuther and M. H. Zenk, *FEBS Lett.*, 1988, **237**, 40–44.
12 J. Stoeckigt, S. Panjikar, M. Ruppert, L. Barleben, X. Y. Ma, E. Loris and M. Hill, *Phytochem. Rev.*, 2007, **6**, 15–34.
13 J. Stoeckigt and S. Panjikar, *Nat. Prod. Rep.*, 2007, **24**, 1382–1400.
14 H. M. Berman, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 2008, **A64**, 88–95.
15 S. Yokoyama, *Curr. Opin. Chem. Biol.*, 2003, **7**, 39–43.
16 Y. Kim, G. Babnigg, R. Jedrzejczak, W. H. Eschenfeldt, H. Li, N. Maltseva and C. Hatzos-Skintges, *Methods*, 2011, **55**, 12–28.
17 N. E. Chayen, *Methods Mol. Biol.*, 2007, **363**, 175–190.
18 M. A. Dessau and Y. Modis, *J. Vis. Exp.*, 2011, **47**, 2285.
19 N. E. Chayen, *Structure*, 1997, **5**, 1269–1274.
20 A. D'Arcy, F. Villard and M. Marsh, *Acta Cryst.*, 2007, **63**, 550–554.
21 T. S. Walter, E. J. Mancini, J. Kadlec, S. C. Graham, R. Assenberg, J. Ren, S. Sainsbury, R. J. Owens, D. I. Stuart, J. M. Grimes and K. Harlos, *Acta Cryst.*, 2008, **64**, 14–18.
22 J. Mueller-Dieckmann, *Acta Cryst.*, 2006, **62**, 1446–1452.
23 U. Heinemann, K. Büssow, U. Mueller and P. Umbach, *Acc. Chem. Res.*, 2003, **36**, 157–163.
24 G. Ueno, H. Kanda, T. Kumasaka and M. Yamamoto, *J. Synchrotron Radiat.*, 2005, **12**, 380–384.
25 W. X. Shi, H. Robinson, M. Sullivan, D. Abel, J. Toomey, L. E. Berman, D. Lynch, G. Rosenbaum, G. Rakowsky, L. Rock, B. Nolan, G. Shea-McCarthy, D. Schneider, E. Johnson, R. M. Sweet and M. R. Chance, *J. Synchrotron Radiat.*, 2006, **13**, 365–372.
26 A. Beteva, F. Cipriani, S. Cusack, S. Delageniere, J. Gabadinho, E. J. Gordon, M. Guijarro, D. R. Hall, S. Larsen, L. Launer,

C. B. Lavault, G. A. Leonard, T. Mairs, A. McCarthy, J. McCarthy, J. Meyer, E. Mitchell, S. Monaco, D. Nurizzo, P. Pernot, R. Pieritz, R. G. B. Ravelli, V. Rey, W. Shepard, D. Spruce, D. I. Stuart, O. Svensson, P. Theveneau, X. Thibault, J. Turkenburg, M. Walsh and S. M. McSweeney, *Acta Cryst.*, 2006, **62**, 1162–1169.
27 F. Cipriani, F. Felisaz, L. Launer, J. S. Aksoy, H. Caserotto, S. Cusack, M. Dallery, F. Di-Chiaro, M. Guijarro, J. Huet, S. Larsen, M. Lentini, J. McCarthy, S. McSweeney, R. Ravelli, M. Renier, C. Taffut, A. Thompson, G. A. Leonard and M. A. Walsh, *Acta Cryst.*, 2006, **62**, 1251–1259.
28 H. Hartmann, F. Parak, W. Steigemann, G. A. Petsko, D. R. Ponzi and H. Frauenfelder, *Proc. Natl. Acad. Sci. U. S. A.*, 1982, **79**, 4967–4971.
29 U. F. Thomanek, F. Parak, R. L. Mössbauer, H. Formanek and P. Schwager, *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.*, 1973, **A29**, 263–265.
30 H. Hope, *Acta Crystallogr., Sect. B: Struct. Sci.*, 1988, **B44**, 22–26.
31 A. Riboldi-Tunnicliffe and R. Hilgenfeld, *J. Appl. Crystallogr.*, 1999, **32**, 1003–1005.
32 E. F. Garman, *Methods Mol. Biol.*, 1996, **56**, 87–126.
33 B. Heras and J. L. Martin, *Acta Cryst.*, 2005, **61**, 1173–1180.
34 E. F. Garman and R. L. Owen, *Acta Cryst.*, 2006, **62**, 32–47.
35 H. Marshall, M. Venkat, N. S. H. L. Seng, J. Cahn and D. H. Juers, *Acta Cryst.*, 2012, **68**, 69–81.
36 J. M. Harp, D. E. Timm and G. J. Bunick, *Acta Cryst.*, 1998, **54**, 622–628.
37 J. I. Yeh and W. G. Hol, *Acta Cryst.*, 1998, **54**, 479–480.
38 E. Garman, *Acta Cryst.*, 1999, **55**, 1641–1653.
39 J. M. Harp, B. L. Hanson, D. E. Timm and G. J. Bunick, *Acta Cryst.*, 1999, **55**, 1329–1334.
40 V. R. Samygina, S. V. Antonyuk, V. S. Lamzin and A. N. Popov, *Acta Cryst.*, 2000, **56**, 595–603.
41 A. Vahedi-Faridi, V. Stojanoff and J. I. Yeh, *Acta Cryst.*, 2005, **61**, 982–989.
42 B. Heras, M. A. Edeling, K. A. Byriel, A. Jones, S. Raina and J. L. Martin, *Structure*, 2003, **11**, 139–145.
43 J. Sanchez-Weatherby, M. W. Bowler, J. Huet, A. Gobbo, F. Felisaz, B. Lavault, R. Moya, J. Kadlec, R. B. G. Ravelli and F. Cipriani, *Acta Cryst.*, 2009, **65**, 1237–1246.
44 S. Russi, D. H. Juers, J. Sanchez-Weatherby, E. Pellegrini, E. Mossou, V. T. Forsyth, J. Huet, A. Gobbo, F. Felisaz, R. Moya, S. M. McSweeney, S. Cusack, F. Cipriani and M. W. Bowler, *J. Struct. Biol.*, 2011, **175**, 236–243.
45 T. M. McPhillips, S. E. McPhillips, H. J. Chiu, A. E. Cohen, A. M. Deacon, P. J. Ellis, E. Garman, A. Gonzalez, N. K. Sauter, R. P. Phizackerley, S. M. Soltis and P. Kuhn, *J. Synchrotron Radiat.*, 2002, **9**, 401–406.
46 J. M. Skinner, M. Cowan, R. Buono, W. Nolan, H. Bosshard, H. H. Robinson, A. Héroux, A. S. Soares, D. K. Schneider and R. M. Sweet, *Acta Cryst.*, 2006, **62**, 1340–1347.
47 J. Gabadinho, A. Beteva, M. Guijarro, V. Rey-Bakaikoa, D. Spruce, M. W. Bowler, S. Brockhauser, D. Flot, E. J. Gordon, D. R. Hall, B. Lavault, A. A. McCarthy, J. McCarthy, E. Mitchell, S. Monaco, C. Mueller-Dieckmann, D. Nurizzo, R. B. G. Ravelli, X. Thibault, M. A. Walsh, G. A. Leonard and S. M. McSweeney, *J. Synchrotron Radiat.*, 2010, **17**, 700–707.
48 S. Stepanov, O. Makarov, M. Hilgart, S. B. Pothineni, A. Urakhchin, S. Devarapalli, D. Yoder, M. Becker, C. Ogata, R. Sanishvili, N. Venugopalan, J. L. Smith and R. F. Fischetti, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2011, **D67**, 176–188.
49 M. Fodje, K. Janzen, R. Berg, G. Black, S. Labiuk, J. Gorin and P. Grochulski, *J. Synchrotron Radiat.*, 2012, **19**, 274–280.
50 G. P. Bourenkov and A. N. Popov, *Acta Cryst.*, 2010, **66**, 409–419.
51 W. Kabsch, *Acta Cryst.*, 2010, **66**, 125–132.
52 A. G. W. Leslie, *Acta Cryst.*, 2006, **62**, 48–57.
53 Z. Otwinowski and W. Minor, *Methods Enzymol.*, 1997, **276**, 307–326.
54 K. S. Paithankar and E. F. Garman, *Acta Cryst.*, 2010, **66**, 381–388.
55 R. B. G. Ravelli, H. K. S. Leiros, B. Pan, M. Caffrey and S. McSweeney, *Structure*, 2003, **11**, 217–224.
56 M. H. Nanao and R. B. G. Ravelli, *Structure*, 2006, **14**, 791–800.
57 S. Panjikar, H. Mayerhofer, P. A. Tucker, J. Mueller-Dieckmann and D. de Sanctis, *Acta Cryst.*, 2011, **67**, 32–44.
58 D. de Sanctis, P. A. Tucker and S. Panjikar, *J. Synchrotron Radiat.*, 2011, **18**, 374–380.
59 A. Vagin and A. Teplyakov, *J. Appl. Crystallogr.*, 1997, **30**, 1022–1025.
60 J. Navaza, *Acta Cryst.*, 2001, **57**, 1367–1372.
61 R. Schwarzenbacher, A. Godzik, S. K. Grzechnik and L. Jaroszewski, *Acta Cryst.*, 2004, **60**, 1229–1236.
62 A. A. Lebedev, A. A. Vagin and G. N. Murshudov, *Acta Cryst.*, 2008, **64**, 33–39.
63 A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J. Appl. Crystallogr.*, 2007, **40**, 658–674.
64 V. Reddy, S. M. Swanson, B. Segelke, K. A. Kantardjieff, J. C. Sacchettini and B. Rupp, *Acta Cryst.*, 2003, **59**, 2200–2210.
65 J. B. Claude, K. Suhre, C. Notredame, J. M. Claverie and C. Abergel, *Nucleic Acids Res.*, 2004, **32**, 606–609.
66 B. V. Strokopytov, A. Fedorov, N. M. Mahoney, M. Kessels, D. G. Drubin and S. C. Almo, *Acta Cryst.*, 2005, **61**, 285–293.
67 S. Panjikar, V. Parthasarathy, V. S. Lamzin, M. S. Weiss and P. A. Tucker, *Acta Cryst.*, 2005, **61**, 449–457.
68 F. Long, A. A. Vagin, P. Young and G. N. Murshudov, *Acta Cryst.*, 2008, **64**, 125–132.
69 R. M. Keegan and M. D. Winn, *Acta Cryst.*, 2008, **64**, 119–124.
70 D. Raimondo, A. Giorgetti, A. Giorgetti, S. Bosi and A. Tramontano, *Proteins: Struct., Funct., Bioinf.*, 2007, **66**, 689–696.
71 D. D. Rodríguez, C. Grosse, S. Himmel, C. González, I. M. de Ilarduya, S. Becker, G. M. Sheldrick and I. Usón, *Nat. Methods*, 2009, **6**, 651–653.
72 J. P. Schuermann and J. J. Tanner, *Acta Cryst.*, 2003, **59**, 1731–1736.
73 S. Panjikar, V. Parthasarathy, V. S. Lamzin, M. S. Weiss and P. A. Tucker, *Acta Cryst.*, 2009, **65**, 1089–1097.
74 D. Jain and V. Lamour, *Methods Mol. Biol.*, 2010, **673**, 129–156.
75 C. Vonrhein, E. Blanc, P. Roversi and G. Bricogne, *Methods Mol. Biol.*, 2007, **364**, 215–230.
76 W. Minor, M. Cymborowski, Z. Otwinowski and M. Chruszcz, *Acta Cryst.*, 2006, **62**, 859–866.
77 N. S. Pannu, W. J. Waterreus, P. Skubák, I. Sikharulidze, J. P. Abrahams and R. A. G. de Graaff, *Acta Cryst.*, 2011, **67**, 331–337.
78 C. M. Weeks, R. H. Blessing, R. Miller, R. Mungee, S. A. Potter, J. Rappleye, G. D. Smith, H. Xu and W. Furey, *Z. Kristallogr.*, 2002, **217**, 686–693.
79 P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **D66**, 213–221.
80 T. Pape and T. R. Schneider, *J. Appl. Crystallogr.*, 2004, **37**, 843–844.
81 T. R. Schneider and G. M. Sheldrick, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2002, **D58**, 1772–1779.
82 Collaborative Computational Project, Number 4, *Acta Cryst.*, 1994, **D50**, 760–763.
83 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. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson, *Acta Cryst.*, 2011, **67**, 235–242.
84 R. A. de Graaff, M. Hilge, J. L. van der Plas and J. P. Abrahams, *Acta Cryst.*, 2001, **57**, 1857–1862.
85 N. S. Pannu and R. J. Read, *Acta Cryst.*, 2004, **60**, 22–27.
86 K. D. Cowtan and Z. Y. Zhang, *Prog. Biophys. Mol. Biol.*, 1999, **72**, 245–270.
87 J. P. Abrahams and A. G. W. Leslie, *Acta Cryst.*, 1996, **52**, 30–42.
88 G. Langer, S. X. Cohen, V. S. Lamzin and A. Perrakis, *Nat. Protoc.*, 2008, **3**, 1171–1179.
89 K. Cowtan, *Acta Cryst.*, 2006, **62**, 1002–1011.
90 T. C. Terwilliger, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2000, **D56**, 965–972.
91 H. Xu and C. M. Weeks, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2008, **D64**, 172–177.
92 H. Xu, H. A. Hauptman and C. M. Weeks, *Acta Cryst.*, 2002, **58**, 90–96.
93 P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta Cryst.*, 2010, **66**, 486–501.
94 T. A. Jones, J. Y. Zou, S. W. Cowan and M. Kjeldgaard, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 1999, **A47**, 110–119.

95 D. E. McRee and M. Israel, *J. Struct. Biol.*, 2008, **163**, 208–213.
96 E. Dogru, H. Warzecha, F. Seibel, S. Haebel, F. Lottspeich and J. Stoeckigt, *Eur. J. Biochem.*, 2000, **267**, 1397–1406.
97 X. Y. Ma, J. Koepke, A. Bayer, V. Linhard, G. Fritzsch, B. Zhang, H. Michel and J. Stoeckigt, *Biochim. Biophys. Acta, Proteins Proteomics*, 2004, **1701**, 129–132.
98 J. Stoeckigt, in *The Alkaloids*, ed. G. A. Cordell, Academic Press, San Diego, CA, 1995, vol. 47, pp. 115–172.
99 D. Bracher and T. M. Kutchan, *Arch. Biochem. Biophys.*, 1992, **294**, 717–723.
100 A. DeWaal, A. H. Meijer and R. Verpoorte, *Biochem. J.*, 1995, **306**, 571–580.
101 Y. Yamazaki, H. Sudo, M. Yamazaki, N. Aimi and K. Saito, *Plant Cell Physiol.*, 2003, **44**, 395–403.
102 T. M. Kutchan, *Phytochemistry*, 1993, **32**, 493–506.
103 J. Stoeckigt and M. Ruppert, *Nat. Prod. Chem.*, 1999, **4**, 109–138.
104 X. Y. Ma, J. Koepke, A. Bayer, G. Fritzsch, H. Michel and J. Stoeckigt, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2005, **D61**, 694–696.
105 X. Y. Ma, S. Panjikar, J. Koepke, E. Loris and J. Stoeckigt, *Plant Cell*, 2006, **18**, 907–920.
106 E. A. Loris, S. Panjikar, M. Ruppert, L. Barleben, M. Unger, H. Schübel and J. Stoeckigt, *Chem. Biol.*, 2007, **14**, 979–985.
107 J. J. Maresh, L. A. Giddings, A. Friedrich, E. A. Loris, S. Panjikar, B. L. Trout, J. Stoeckigt, B. Peters and S. E. O'Connor, *J. Am. Chem. Soc.*, 2008, **130**, 710–723.
108 J. F. Treimer and M. H. Zenk, *Eur. J. Biochem.*, 1979, **101**, 225–233.
109 I. Gerasimenko, Y. Sheludko, X. Y. Ma and J. Stoeckigt, *Eur. J. Biochem.*, 2002, **269**, 2204–2213.
110 H. Warzecha, I. Gerasimenko, T. M. Kutchan and J. Stoeckigt, *Phytochemistry*, 2000, **54**, 657–666.
111 L. Barleben, X. Ma, J. Koepke, G. Peng, H. Michel and J. Stoeckigt, *Biochim. Biophys. Acta, Proteins Proteomics*, 2005, **1747**, 89–92.
112 M. Czjzek, M. Cicek, V. Zamboni, W. P. Burmeister, D. R. Bevan, B. Henrissat and A. Esen, *Biochem. J.*, 2001, **354**, 37–46.
113 L. Barleben, S. Panjikar, M. Ruppert, J. Koepke and J. Stoeckigt, *Plant Cell*, 2007, **19**, 2886–2897.
114 B. St Pierre and V. De Luca, *Recent Adv. Phytochem.*, 2000, **34**, 285–315.
115 J. D'Auria, *Curr. Opin. Plant Biol.*, 2006, **9**, 331–340.
116 X. Ma, J. Koepke, S. Panjikar, G. Fritzsch and J. Stoeckigt, *J. Biol. Chem.*, 2005, **280**, 13576–13583.
117 L. Yang, M. Hill, M. Wang, S. Panjikar and J. Stoeckigt, *Angew. Chem., Int. Ed.*, 2009, **48**, 5211–5213.
118 Y. Cajal, A. Svendsen, J. DeBolos, S. A. Patkar and M. A. Alsina, *Biochimie*, 2000, **82**, 1053–1061.
119 E. Mattern-Dogru, X. Ma, J. Hartmann, H. Decker and J. Stoeckigt, *Eur. J. Biochem.*, 2002, **269**, 2889–2896.
120 J. Zuegg, K. Gruber, M. Gugganig, U. G. Wagner and C. Kratky, *Protein Sci.*, 1999, **8**, 1990–2000.
121 H. Lauble, B. Miehlich, S. Förster, H. Wajant and F. Effenberger, *Protein Sci.*, 2001, **10**, 1015–1022.
122 K. Gruber, G. Gartler, B. Krammer, H. Schwab and C. Kratky, *J. Biol. Chem.*, 2004, **279**, 20501–20510.
123 M. Ruppert, S. Panjikar, L. Barleben and J. Stoeckigt, *Acta Cryst.*, 2006, **62**, 257–260.
124 L. Xia, M. Ruppert, M. Wang, S. Panjikar, H. Lin, C. Rajendran, L. Barleben and J. Stoeckigt, *Chem. Biol.*, 2012, **7**, 226–234.
125 L. Sun, M. Ruppert, Y. Sheludko, H. Warzecha, Y. Zhao and J. Stoeckigt, *Plant Mol. Biol.*, 2008, **67**, 455–467.
126 R. V. Edwankar, C. R. Edwankar, J. Deschamps and J. M. Cook, *Org. Lett.*, 2011, **13**, 5216–5219.
127 W. R. Rypniewski, H. M. Holden and I. Rayment, *Biochemistry*, 1993, **32**, 9851–9858.
128 C. Rosenthal, U. Mueller, S. Panjikar, L. Sun, M. Ruppert, Y. Zhao and J. Stoeckigt, *Acta Cryst.*, 2006, **A62**, 1286–1289.
129 L. Sun, Y. Chen, C. Rajendran, U. Mueller, S. Panjikar, M. Wang, R. Mindnich, C. Rosenthal, T. M. Penning and J. Stoeckigt, *J. Biol. Chem.*, 2012, **287**, 11213–11221.
130 J. M. Jez, M. J. Bennett, B. P. Schlegel, M. Lewis and T. M. Penning, *Biochem. J.*, 1997, **326**, 625–636.
131 Y. Jin and T. M. Penning, *Steroids*, 2006, **71**, 380–391.
132 P. J. Simpson, C. Tantitadapitak, A. M. Reed, O. C. Mather, C. M. Bunce, S. A. White and J. P. Ride, *J. Mol. Biol.*, 2009, **392**, 465–480.
133 H. Jörnvall, B. Persson, M. Krook, S. Atrian, R. Gonzàlez-Duarte, J. Jeffery and D. Ghosh, *Biochemistry*, 1995, **34**, 6003–6013.
134 K. Nakajima, A. Yamashita, H. Akama, T. Nakatsu, H. Kato, T. Hashimoto, J. Oda and Y. Yamada, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 4876–4881.
135 M. G. Rossmann, D. Moras and K. W. Olsen, *Nature*, 1974, **250**, 194–199.
136 A. Yamashita, H. Kato, S. Wakatsuki, T. Tomizaki, T. Nakatsu, K. Nakajima, T. Hashimoto, Y. Yamada and J. Oda, *Biochemistry*, 1999, **38**, 7630–7637.
137 T. Hashimoto, K. Nakajima, G. Ongena and Y. Yamada, *Plant Physiol.*, 1992, **100**, 836–845.
138 B. Draeger, *Phytochemistry*, 2006, **67**, 327–337.
139 J. Ziegler and P. J. Facchini, *Annu. Rev. Plant Biol.*, 2008, **59**, 735–769.
140 N. Samanani and P. J. Facchini, *Planta*, 2001, **213**, 898–906.
141 A. Pasquo, A. Bonamore, S. Franceschini, A. Macone, A. Boffi and A. Ilari, *Acta Cryst.*, 2008, **F64**, 281–283.
142 A. Ilari, S. Franceschini, A. Bonamore, F. Arenghi, B. Botta, A. Macone, A. Pasquo, L. Bellucci and A. Boffi, *J. Biol. Chem.*, 2009, **284**, 897–904.
143 P. Steffens, N. Nagakura and M. H. Zenk, *Phytochemistry*, 1985, **24**, 2577–2583.
144 T. M. Kutchan and H. Dittrich, *J. Biol. Chem.*, 1995, **270**, 24475–24481.
145 D. E. Edmondson, *Structure*, 2007, **15**, 639–641.
146 N. S. Scrutton, *Nat. Prod. Rep.*, 2005, **21**, 722–730.
147 A. Winkler, F. Hartner, T. M. Kutchan, A. Glieder and P. Macheroux, *J. Biol. Chem.*, 2006, **281**, 21276–21285.
148 A. Winkler, T. M. Kutchan and P. Macheroux, *J. Biol. Chem.*, 2007, **282**, 24437–24443.
149 A. Winkler, A. Lyskowski, S. Riedl, M. Puhl, T. M. Kutchan, P. Macheroux and K. Gruber, *Nat. Chem. Biol.*, 2008, **4**, 739–741.
150 Y. Kallberg, U. Oppermann, H. Joernvall and B. Persson, *Eur. J. Biochem.*, 2002, **269**, 4409–4417.
151 U. C. Oppermann, C. Filling, K. D. Berndt, B. Persson, J. Benach, R. Ladenstein and H. Jörnvall, *Biochemistry*, 1997, **36**, 34–40.
152 J. S. Ziegler, J. Voigtlände, J. Schmidt, R. Kramell, O. Miersch, C. A. Amme, A. Gesell and T. M. Kutchan, *Plant J.*, 2006, **48**, 177–192.
153 Y. Higashi, T. J. Smith, J. M. Jez and T. M. Kutchan, *Acta Cryst.*, 2010, **F66**, 163–166.
154 Y. Higashi, T. M. Kutchan and T. J. Smith, *J. Biol. Chem.*, 2011, **286**, 6532–6541.
155 D. K. Liscombe and P. J. Facchini, *J. Biol. Chem.*, 2007, **282**, 14741–14751.
156 A. Jain, J. Ziegler, D. K. Liscombe, P. J. Facchini, P. A. Tucker and S. Panjikar, *Acta Cryst.*, 2008, **F64**, 1066–1069.
157 E. McCoy, M. C. Galan and S. E. O'Connor, *Bioorg. Med. Chem. Lett.*, 2006, **16**, 2475–2478.
158 N. Yerkes, J. X. Wu, E. McCoy, M. C. Galan, S. Chen and S. E. O'Connor, *Bioorg. Med. Chem. Lett.*, 2008, **18**, 3095–3098.
159 J. D. McCarter and S. G. Withers, *Curr. Opin. Struct. Biol.*, 1994, **4**, 885–892.
160 A. Pictet and T. Spengler, *Ber. Dtsch. Chem. Ges.*, 1911, **44**, 2030–2036.
161 J. Stoeckigt, A. P. Antonchick, F. Wu and H. Waldmann, *Angew. Chem., Int. Ed.*, 2011, **50**, 8538–8564.
162 F. Cedrone, A. Ménez and E. Quéméneur, *Curr. Opin. Struct. Biol.*, 2000, **10**, 405–410.
163 S. Chen, M. C. Galan, C. Coltharp and S. E. O'Connor, *Chem. Biol.*, 2006, **13**, 1137–1141.
164 M. C. Galan, E. McCoy and S. E. O'Connor, *Chem. Commun.*, 2007, 3249.
165 P. Chan, *Eur. Neurol.*, 1995, **35**(Suppl I), 23–27.
166 S. W. Li, *Eur. Neurol.*, 1998, **39**, 26–30.
167 S. Li, J. Long, Z. Ma, Z. Xu, J. Li and Z. Zhang, *Curr. Med. Res. Opin.*, 2004, **20**, 409–415.
168 F. Wu, H. Zhu, L. Sun, C. Rajendran, M. Wang, X. Ren, S. Panjikar, A. Cherkasov, H. Zou and J. Stoeckigt, *J. Am. Chem. Soc.*, 2012, **134**, 1498–1500.
169 G. Cragg and D. Newman, *J. Ethnopharmacol.*, 2005, **100**, 72–79.
170 C. Coderch, A. Morreale and F. Gago, *Anti-Cancer Agents Med. Chem.*, 2012, **12**, 219–225.

171 V. K. Ngan, K. Bellman, D. Panda, B. T. Hill, M. A. Jordan and L. Wilson, *Cancer Res.*, 2000, **60**, 5045–5051.
172 Frampton, J. E. Moen and M. D. Vinflunine, *Drugs*, 2010, **70**, 1283–1293.
173 J. Pollier, T. Moses and A. Goossens, *Nat. Prod. Rep.*, 2011, **28**, 1897–1916.
174 W. T. Shier, K. L. Rinehart and D. Gottlieb, *Proc. Natl. Acad. Sci. U. S. A.*, 1969, **63**, 198–204.
175 K. J. Weissman, *Trends Biotechnol.*, 2007, **25**, 139–42.
176 W. Runguphan, J. J. Maresh and S. E. O'Connor, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 13673–13678.
177 S. E. O'Connor and J. Maresh, *Nat. Prod. Rep.*, 2006, **23**, 532–547.
178 R. van der Heijden, D. I. Jacobs, W. Snoeijer and D. Hallard, *Curr. Med. Chem.*, 2004, **11**, 607–628.
179 P. J. Facchini, K. L. Huber-Allanach and L. W. Tari, *Phytochemistry*, 2000, **54**, 121–138.
180 V. De Luca, C. Marineau and N. Brisson, *Proc. Natl. Acad. Sci. U. S. A.*, 1989, **86**, 2582–2586.
181 E. H. Hughes, S. B. Hong, J. V. Shanks, K. Y. San and S. I. Gibson, *Biotechnol. Prog.*, 2002, **18**, 1183–1186.
182 W. Runguphan and S. E. O'Connor, *Nat. Chem. Biol.*, 2009, **5**, 151–153.
183 P. Bernhardt, E. McCoy and S. E. O'Connor, *Chem. Biol.*, 2007, **14**, 888–897.
184 E. Pichersky and D. R. Gang, *Trends Plant Sci.*, 2000, **5**, 439–445.
185 R. A. Jensen, *Annu. Rev. Microbiol.*, 1976, **30**, 409–425.
186 C. Vogel, C. Berzuini, M. Bashton, J. Gough and S. A. Teichmann, *J. Mol. Biol.*, 2004, **336**, 809–823.
187 S. C. G. Rison and J. M. Thornton, *Curr. Opin. Struct. Biol.*, 2002, **12**, 374–382.
188 C. Chothia and A. M. Lesk, *EMBO J.*, 1986, **5**, 823–826.
189 J. Stoeckigt, L. Barleben, S. Panjikar and E. A. Loris, *Plant Physiol. Biochem.*, 2008, **46**, 340–355.
190 Z. Jawad and M. Paoli, *Structure*, 2002, **10**, 447–454.
191 E. I. Scharff, J. Koepke, G. Fritzsch, C. Lücke and H. Rüterjans, *Structure*, 2001, **9**, 493–502.
192 M. Harel, A. Aharoni, L. Gaidukov, B. Brumshtein, O. Khersonsky, R. Meged, H. Dvir, R. B. G. Ravelli, A. McCarthy, L. Toker, I. Silman, J. L. Sussman and D. S. Tawfik, *Nat. Struct. Mol. Biol.*, 2004, **11**, 412–419.
193 H. Jeon, W. Y. Meng, J. Takagi, M. J. Eck, T. A. Springer and S. C. Blacklow, *Nat. Struct. Biol.*, 2001, **8**, 499–504.
194 N. Samanani, D. K. Liscombe and P. J. Facchini, *Plant J.*, 2004, **40**, 302–313.
195 A. Bonamore, M. Barba, B. Botta, A. Boffi and A. Macone, *Molecules*, 2010, **15**, 2070–2078.
196 L. Xia, C. Rajendran, M. Ruppert, S. Panjikar, M. Wang and J. Stoeckigt, *Phytochemistry*, 2012, DOI: 10.1016/j.phytochem.2012.05.009.
197 H. Y. Cho, S. Y. Lee-Parson, S. Y. H. Yoon, H. S. Rhee and J. M. Park, *Biotechnol. Lett.*, 2007, **29**, 2001–2005.
198 *Br Med J*,1979, **6203**, 1458–1459.