1. Introduction

In chemical biology, bioactive small molecules are applied to biological systems such as living cells for studying and manipulating biological processes.[1] Such a small molecule-based approach for studying biology has several advantages over more classical research strategies such as genetic knock-out studies. The biological effect induced by small molecule modulation is usually rapid and can be spatially and temporally controlled. It is often reversible and conditional as small molecule application can be initiated at any developmental stage of the studied biological system. Moreover, the observed biological effect is tunable by varying the concentration of the small molecule probe. Finally, small molecule probes can be tagged with additional functionalities that allow to accumulate further information in biological experiments. For example, the proper attachment of fluorophores on a bioactive small molecule targeting a protein of interest generates a probe that can directly be employed in imaging applications. Besides these straightforward experimental advantages, small molecule interference has at least one additional conceptual advantage over genetic manipulations. While gene mutations act on the gene-level, small molecules modulate the function of the gene products. Consequently, small molecules interfere at a much more sophisticated and biologically relevant level; accordingly, they are in contrast to genetic studies able to differentiate between different protein forms that result from the same gene, e.g. by alternative splicings or post-translational modifications.

Despite these huge advantages of small molecule manipulation of biological systems, the widespread use of chemical biology methods in biological research is limited by an inherent major disadvantage. In order to be feasible, such an approach requires the availability of suitable small molecules that induce a desired biological effect with high selectivity and potency. To date, only a limited number of proper small molecules for a restricted number of protein targets are though known.

To overcome this limitation, several concepts for identifying proper small molecules have evolved in the last years. Among them, high-throughput screening of compound collections to elucidate proper small molecule modulators is nowadays most regularly used. As a consequence, approaches for synthesizing ‘intelligent’ compound collections such as Biology-oriented synthesis (BIOS) or Diversity-oriented synthesis (DOS) have been devised to reduce compound synthesis and screening efforts.[2] In addition, more rational, structure-based approaches such as ‘virtual screening’ or fragment-based drug discovery are often employed.[3]

An alternative and in the past widely used method for identifying proper small molecules is the use of biologically active natural products as starting points for chemical probe synthesis.[4] In this work, such an approach has been followed with the aim to develop a small molecule for probing the plant immune system. To this end, a peptidic natural product named syringolin A was chosen as a target structure for small molecule development. Syringolin A was reported to induce a plant-protective effect *vs.* various plant pathogens upon application on plants. In contrast to most agrochemicals, this interesting effect however was not caused by an antimicrobial activity against the plant pathogens but by an apparent modulation of the plant immune system. The molecular target of syringolin A and a synthetic route to this natural product class though was unknown at the beginning of this work.

During the course of this PhD thesis, in which several chemical biology studies with syringolin A and derivatives thereof were performed, several unexpected findings either from the own research work or reported in literature, opened up possibilities for promising 'side projects' that were originally not even anticipated. Accordingly, the search for a suitable small molecule probe of the plant immune system ended up with an access to a promising new class of potent anti-cancer agents. Even more particular, the discovery of a new scaffold structure for covalently labeling glyceraldehyde-3-phosphate dehydrogenase has been accomplished. This work therefore gives a striking example on the unexpected but nevertheless promising results that can be achieved by basic science research.

2. Background

2.1. Natural products as starting structures for small molecule development

Since the first description of chemotherapeutic treatments, Nature has always played a prominent role for the identification of therapeutic agents.[5] In fact, in the beginning of commercial drug discovery, the vast majority of chemotherapeutics were derived from natural products. However, with the advance of high-throughput screenings (HTS) in the 1990's, natural products stepped out of the focus of the pharmaceutical industry, resulting in a significant decrease of natural product-based drugs released on the market.[6] The emergence and technological improvement of high-throughput screening (HTS) nowadays enables to test extremely large compound libraries and to evaluate their potency against various biological targets in a very small time frame. The build-up of these huge screening capacities however demanded the allocation of compound libraries of appropriate size. To overcome this bottleneck, combinatorial chemistry soared up to provide a facile and rapid access to drug-like screening libraries of simple compounds with a high diversity and in a low cost fashion.[7,8] Nevertheless, despite the big investment of the pharmaceutical industry and the generation of millions of compounds, it turned out that the biological relevance of combinatorial libraries was rather low, delivering only unexpected low numbers of lead structures suitable for further drug development.

These disappointing results led to a renaissance of natural products in drug discovery. This new awakened interest in natural products stem from their inherent biological relevance. As natural products are synthesized by living organisms to protect and defend themselves and are therefore the result of a long process of the natural selection, their chemical structures are biologically prevalidated for protein interactions. Moreover, their use is not restricted to their original biological function in the producing organism but can exert a biological effect also in other organisms. For example, large varieties of new scaffolds have been isolated from deep-sea organisms such as sponges or marine bacteria and are currently used to treat human pathologies and disorders.[9,10] Consequently, natural products are nowadays again in the focus of drug discovery efforts, either as 'direct' lead structures or as scaffolds for compound library synthesis as in Biology-oriented synthesis (BIOS).

While chemical biology approaches in mammalian cell studies are today well established, the use of small molecules in plant biology research is still in its infancy.[11] Only a limited number of compounds are used to date, although more-and-more small molecule studies have appeared in the last years. Of these probes, most are natural product-derived,

exemplified by Brefeldin A that is used in plant vesicular transport studies or Concanamycin A or Bafilomycin A for elucidating the role of V-ATPases on plant cell tugor pressure.

2.2. The ubiquitin-proteasome system and its modulation by proteasome inhibitors

As described below, the syringolin natural products that are in the focus of this PhD work modulate the ubiquitin-proteasome system by inhibiting the eukaryotic 20S proteasome. For this reason, a short overview on the constitution and biological role of the ubiquitin-proteasome system is given.

Since its discovery in the late 1970's, the ubiquitin-proteasome system (UPS) has been widely studied due to its direct involvement in almost all biological processes of living cells. It consists of several protein components that act in concert to achieve a regulated proteolysis of proteins. The significance of this degradation system is highlighted by the award of the Nobel Prize in Chemistry "for the discovery of ubiquitin-mediated protein degradation" to the UPS research pioneers Aaron Ciechanover, Avram Hershko and Irwin Rose in 2004.

One of the key player and name giving factors of the UPS is the protein ubiquitin. It is a small protein of about 8.5 kDa, highly-conserved in all eukaryotic species with a sequence similarity between the human and yeast variant of 96%. As indicated by its name, ubiquitin is ubiquitously expressed in all eukaryotes. In combination with the 26S proteasome and the ubiquitin-loading enzymes E1-E3, ubiquitin forms the ubiquitin-proteasome system which main function is the proteolysis of proteins which are destined for degradation and the cleavage of unfolded or misfolded proteins (Figure 1).[12,13,14] To this end, a cascade of enzymes known as E1, E2s and E3s tag such proteins with at least four ubiquitin units, thereby directing them to the proteasome which then acts as the proteolytic core system.

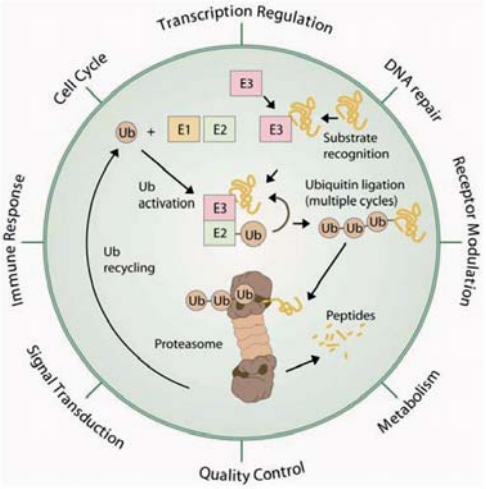


Figure 1. Ubiquitin-mediated proteolysis and its relevance in different biological functions[15]

The 26S proteasome is a 2000 kDa complex of various proteins that can be classified into the 20S core particle structure and two 19S regulatory caps (Figure 2). The proteolysis occurs in the interior of the barrel-shaped 20S structure. The two 19S regulatory subunits overtake the recognition of the ubiquitin-tagged proteins, cleavage of the polyubiquitin chains, unfolding and translocation into the 20S proteasome. The 20S proteasome consists of four stacked ring systems that are arranged in a $\alpha_7\beta_7\beta_7\alpha_7$ manner. The α-subunits, in combination with the 19S subunits, regulate the access to the proteolytic sites which are located at the inner side of three of the seven β-subunits. The catalytic residue is a *N*-terminal threonine moiety which is situated at the β1-, β2- and β5-subunit.

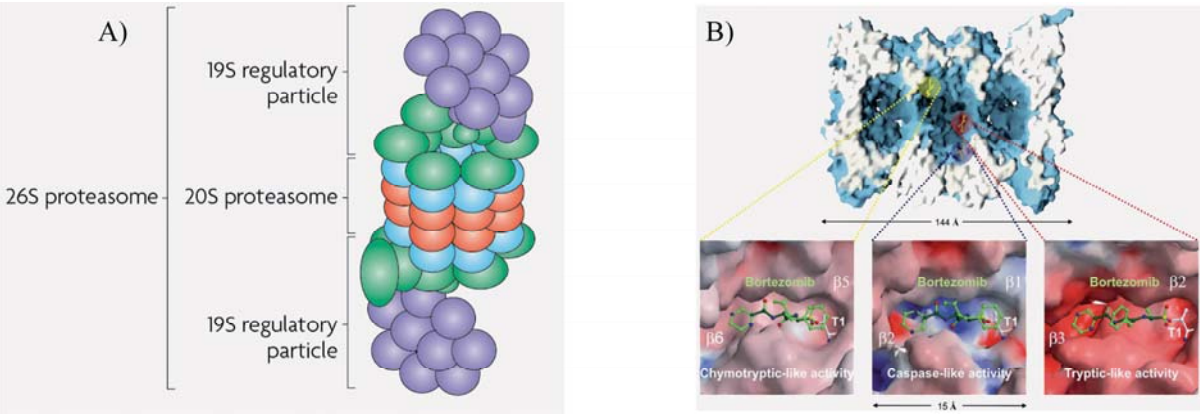


Figure 2. A) Schematic diagram of the 26S proteasome.[16] B) Surface representation of the yeast 20S proteasome crystallized in the presence of bortezomib. The various proteolytic surfaces are marked by a specific color coding: blue = subunit β1; red = subunit β2; and yellow = subunit β5. The nucleophilic threonine (T1) and bortezomib are presented as a ball-and-stick model.[17]

The three different proteolytic sites of the 20S proteasome feature distinct substrate specificities. The β5-subunit shows a chymotrypsin-like activity which cleaves peptide bonds after hydrophobic residues (e.g. tyrosine, tryptophan, and phenylalanine); the β2-subunit exerts a trypsin-like site which cuts after basic residues (lysine and arginine) whereas the β1-subunit possesses a caspase-like activity resulting in preferential cleavage after acidic residues (aspartic acid, glutamic acid). Protein degradation by the proteasome normally results in peptide fragments with an average length of 7 to 9 amino acid residues which are either used for antigen presentation at MHC class I complexes or are further degraded by additional proteases to refill the amino acid pool for new protein synthesis.

Although the UPS has major implications in various cellular processes such as regulation of cell cycle, DNA repair, immune and inflammatory responses and cancer,[18,19,20,21,22,23,24] inhibition of the UPS has surprisingly emerged as a valuable approach for the development of novel anticancer therapies. In fact, the peptide boronic acid PS-341 (bortezomib) has recently been approved by the U.S. Food and Drug Administration for the treatment of relapsed and/or refractory multiple myeloma and is commercially available under its trade name Velcade ®.

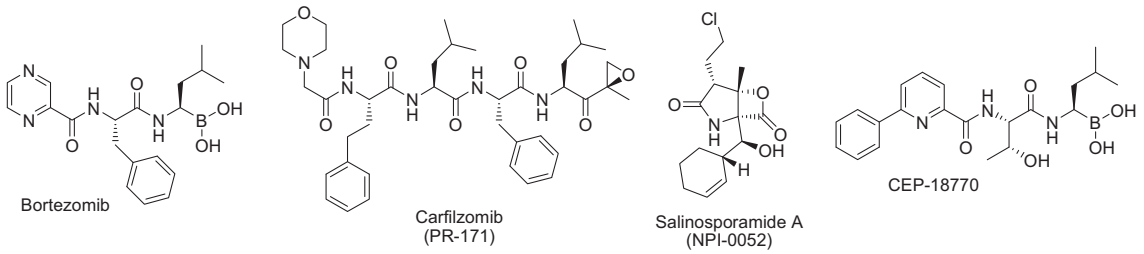


Figure 3. Chemical structure of the marketed proteasome inhibitor Bortezomib and of the proteasome inhibitors currently evaluated in clinical trials.

Although the efficiency of Bortezomib in monotherapy has been considered as astonishing and exciting, side effects and resistances are frequently observed and combinations of bortezomib with other therapeutics are now under evaluation.[25] Consequently, three more proteasome inhibitors are currently evaluated in clinical trials, i.e. carfilzomib, NPI-0052 (also known as salinosporamide A) and CEP-18770 (Figure 3).[26,27,28]

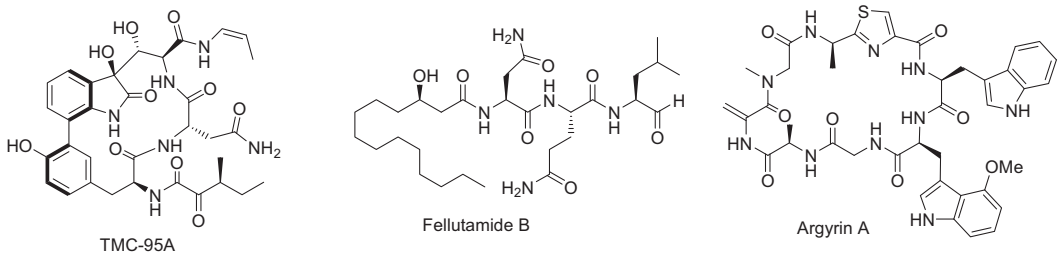
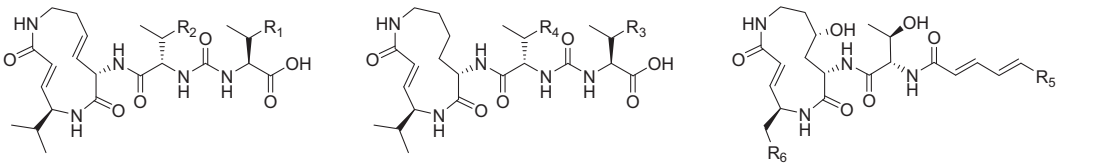


Figure 4. Recently discovered proteasome inhibitors from natural sources.

Despite these advanced clinical studies, alternative proteasome inhibitors are still urgently sought. Natural products have been an invaluable source for proteasome inhibitors, exemplified by the three natural products TMC-95A, fellutamide B or argyrin A (Figure 4) that have recently been disclosed.[29,30,31]

2.3. The natural product syringolin A

The peptidic small molecule syringolin A (SylA, **1**) (Figure 5) was isolated in 1998 from bacterial strains of *Pseudomonas syringae* pv. *syringae* (*Pss*). This plant pathogen causes the brown spot disease on various plant species. In 1999, syringolins B-F were subsequently reported as minor metabolites of *Pss*. SylA is biosynthesized by a mixed non-ribosomal peptide synthetase (NRPS) / polyketide synthetase (PKS) gene cluster exclusively under infection conditions. The syringolins (in particular syringolin B and E) have related chemical structures to another class of natural products known as glidobactins and cepafungins (Figure 5). The main structural differences between syringolins and glidobactins are an additional (*E*)-configured double bond in the syringolin A macrocycle and the presence of an exocyclic urea moiety in the syringolins while glidobactins feature a dipeptide residue.



SylA, **1**: R_1= Me, R_2= Me
SylC: R_1= Me, R_2= Et
SylD: R_1= Et, R_2= Me
SylF: R_1= Et, R_2= Et

SylB, **2**: R_3= Me, R_4= Me
SylE: R_3= Et, R_4= Me

GlbA, **3** or Cepafungin II: R_5= -$(CH_2)_6CH_3$, R_6= H
GlbB: R_5= -$(CH_2)_2CH{=}CH(CH_2)_4CH_3$, R_6= H
GlbC: R_5= -$(CH_2)_8CH_3$, R_6= H
GlbD: R_5= -$(CH_2)_4CH(OH)CH_2CH_3$, R_6= H
GlbE: R_5= -$CH_2CH(OH)(CH_2)_4CH_3$, R_6= H
GlbF: R_5= -$(CH_2)_4CH_3$, R_6= H
GlbG: R_5= -$(CH_2)_6CH_3$, R_6= OH
Cepafungin I: R_5= -$(CH_2)_5CH(CH_3)_2$, R_6 = H
Cepafungin III: R_5= -$(CH_2)_3CH(CH_3)_2$, R_6 = H

Figure 5. Chemical structures of the major metabolite syringolin A of the syringolin family and the additional minor metabolites produced by *Pss*. The chiral centers of the amino acid residues were not assigned in the original structure elucidation and were elucidated only by our syntheses. For clarity, they have been added to this figure. In addition, the chemical structures of the structurally-related glidobactins and cepafungins are represented.

In the original structure elucidation of the syringolins, the chiral centers at the amino acids could not be elucidated. Although the analysis of the biosynthesis gene cluster suggested an overall (L)-configuration, definite prove of the stereochemistry was only achieved by chemical synthesis performed during this PhD project (see section 4.1.2.2.4 for details).

The interest to develop syringolin A into a small molecule probe for studying the plant immune system stemmed from a report in literature, in which the initiation of a plant immune response was described after application of syringolin A to non-host plants. Interestingly, while syringolin A application to healthy plants resulted in no evident phenotype, application of syringolin A to rice plants (*Oryza sativa*) infected with the blast fungus *Pyricularia oryzae* led to a plant protective and infection-counteractive plant immune reaction known as hypersensitive response. The same protective and curative properties of SylA were observed with wheat and powdery mildew as the plant infection.[32,33,34] In plant biology, compounds with such an immune-response promoting activity are generally referred to as elicitors. Most elicitors however are of high molecular weight, being whole proteins or large carbohydrate structures. Only a limited number of small molecule elicitors are known to date, thereby highlighting again the interesting biological activity of SylA.

In 2006, an interesting paper from the Bachmann group at the Hawaiian Cancer Center appeared; describing an anti-proliferative and apoptosis-inducing property of SylA in certain cancer cell lines.[35] However, no mode-of-action of SylA was mentioned.

During the course of this PhD thesis, Groll *et al.* then finally reported SylA's molecular target and biological function. SylA induces an irreversible inhibition of the eukaryotic 20S proteasome by a novel mechanism, thereby providing a previously unrecognized link of the plant immune response with the ubiquitin proteasome pathway. This inhibition facilitates infection of plants by a yet unknown mechanism, suggesting that SylA acts as a bacterial virulence factor.[36]

In addition, the authors disclosed that the structurally related natural product glidobactin A (GlbA, **3**) also potently inhibited the 20S proteasome.[37,38,39,40] Glidobactins have been isolated from *Polyangium brachysporum* and reported as antifungal compounds and antitumor antibiotics. *In vitro* testing of glidobactins (principally A, B and C) revealed a broad activity against clinically important pathogenic fungi such as *Candida albicans*, *Cryptococcus neoformans* or *Aspergillus fumigatus*, although their *in vivo* antifungal potency was considered as only marginal. More interestingly, a strong cytotoxicity was detected against various tumor cell lines. Subsequent studies with mice implanted with P388 leukemia revealed a dramatic prolongation of lifetime under glidobactin treatment.[41]

Biochemical inhibition assays were then used to evaluate the inhibition potency and subsite selectivity of SylA and GlbA (Table 1).

Table 1. Apparent K_i' values and rates of covalent inhibition ($k_{association}$) over inhibitor concentrations.

Inhibitor	Inhibited activity	K_i'	k_{assoc}, M^{-1} s^{-1}
SylA	Chymotrypsin-like	843 ± 8.4 nM (n = 3)	863 ± 106 (n = 6; 100 – 200 nM)
	Trypsin-like	6.7 ± 0.7 µM (n = 6)	94 ± 12 (n = 6; 150 – 600 nM)
	Caspase-like	n.d.[a]	6 ± 0.3 (n = 6; 20 – 40 µM)
GlbA	Chymotrypsin-like	49 ± 5.4 nM (n = 3)	3,377 ± 341 (n = 6; 40 – 60 nM)
	Trypsin-like	2.0 ± 0.6 µM (n = 6)	141 ± 21 (n = 6; 250 – 500 nM)
	Caspase-like	n.a.[b]	n.a.[b]

[a] n.d.: not determined. [b] n.a.: not active.

It turned out that both inhibitors acted in a non-competitive manner, suggesting an irreversible inhibition. SylA inhibited all three proteolytic activities, but with different potencies. While the chymotrypsin-like activity with a K_i' of 843 ± 8.4 nM was most potently inhibited, the trypsin-like activity was inhibited with a K_i' of 6.7 ± 0.7 µM, while no K_i' value for caspase-like inhibition could be quantified despite its apparent inhibition. In contrast, GlbA was much more potent in inhibiting the chymotrypsin-like activity with a K_i' of 49 ±

5.4 nM and the trypsin-like activity with a K_i' of 2.0 ± 0.6 µM while however the caspase-like activity was not inhibited at all in the concentration range tested.

Subsequent structural studies of the binding mode of SylA and GlbA to the yeast 20S proteasome confirmed the suggested irreversible binding mode (Figure 6). These studies revealed that the observed irreversible inhibition is indeed a result of a Michael-type addition of the threonine active site residue to the α,β-unsatured amide moiety of SylA and GlbA. Interestingly, these studies also validated the noticed subsite selectivity. While SylA binding was found in all three proteolytically active subsites, GlbA was only accommodated in the β2 and β5 subunits. Due to their similar binding mode, SylA and GlbA were grouped into a collective class of natural product, named syrbactins.

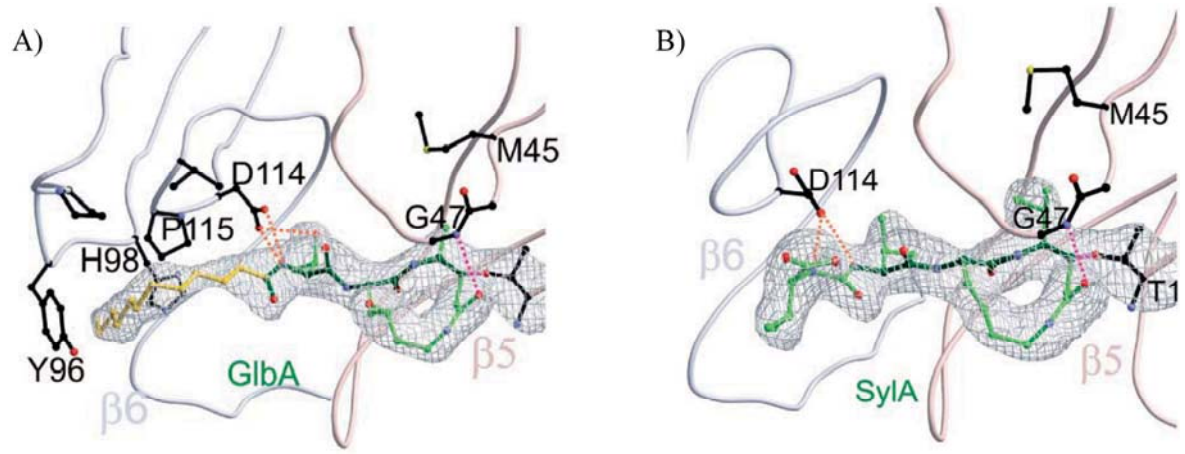


Figure 6. Stereo representation of the chymotryptic-like active site in complex with: A) GlbA and B) SylA. Magenta, covalent linkage of inhibitors with active site Thr1; dotted lines indicate hydrogen bonds. Black, residues performing specific interactions with SylA and GlbA.

Accordingly, the findings of Groll *et al.* also gave a mechanistic explanation to the previously observed anti-proliferative properties of SylA in cancer cells. As described in section 2.2, proteasome inhibitors are potential anticancer agents. Consequently, SylA and analogues thereof might represent promising new chemotherapeutics for the treatment of certain cancer types.

2.4. Previous synthetic approaches to the synthesis of syrbactin-like macrocycles

It has long been recognized that macrocyclic natural products often demonstrate high biological activities in conjunction with favorable pharmacological properties, turning them into interesting compounds for drug discovery efforts.[42] Their observed potent biological potencies seem to be to the consequences of different aspects. First, the entropic loss while forming the protein-ligand complex is dramatically reduced by using macrocyclic inhibitors, resulting in higher target affinity. Second, it has been shown that macrocyclic inhibitors are more selective than linear ligands due to a fixation of target-specific conformations. Third, the physicochemical properties of macrocycles such as cyclic peptide-like drugs have empirically been shown to be more appropriate for membrane passage than of related linear peptide-like candidates. Fourth, in the case of peptides, the proteolytic and metabolic stability of cyclopeptides is usually dramatically enhanced in comparison to linear ones.[43,44]

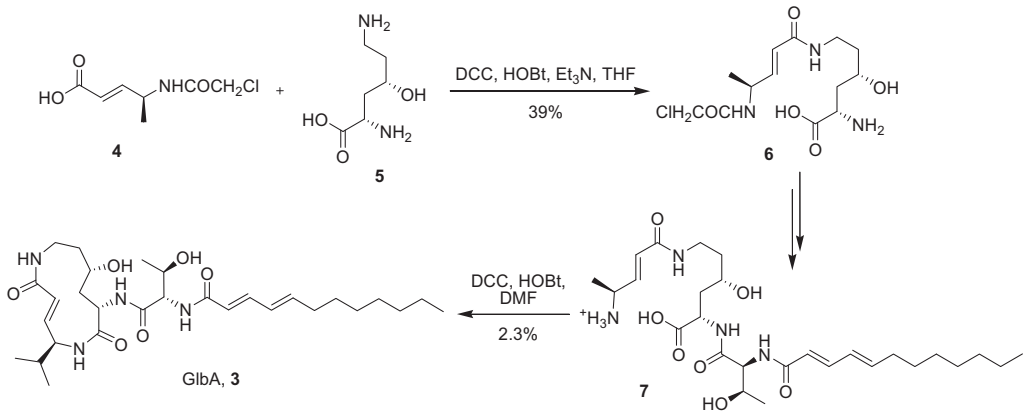
Despite the unambiguous potential of this class of natural products, their development into drugs was underexploited by the pharmaceutical industry for several reasons. One reason is that the molecular mass of these biologically active macrocycles often exceeds 500 Da; thereby violating the commonly adopted Lipinski's rule of 5 for evaluation of the druglikeness of small molecules. Additionally, macrocyclic compounds frequently display complex molecular architectures, thereby limiting their synthetic accessibility and discouraging a possible industrial development. With the advance of new synthetic methodologies for macrocyclization such as ring-closing metathesis,[45] multi-component reactions,[46] metal-templated chelation[47] or ring-closing-contraction sequences,[48] this chemistry-derived disadvantage however starts to vanish.

Accordingly, syringolins seem as attractive target molecules for drug discovery. They feature a highly rigid 12-membered macrocycle with two (*E*)-configured double bonds and two amide functionalities. Due to their limited flexibility, high affinity binding to the active subunit of the proteasome can be assumed which contributes to the observed potent inhibition. Additionally, syringolins possess a moderate number of stereocenters that are derived from amino acids and have a molecular mass close to 500 Da, suggesting favorable pharmacokinetic properties.

To date, only two chemical syntheses of syrbactins have been reported. While syringolins have not yet been prepared, two total syntheses of glidobactin A have been reported by Oka *et al.* in 1988 and the Schmidt group in 1992. In addition, different

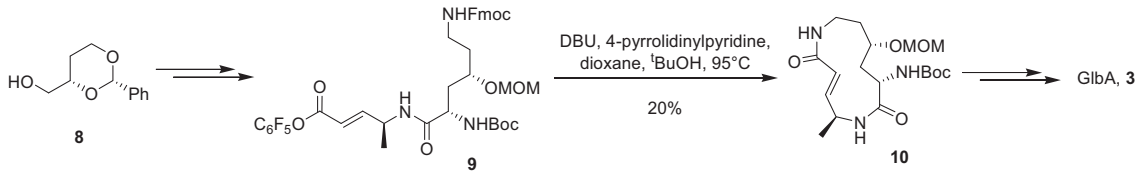
unfortunately unsuccessful approaches towards the synthesis of the glidobactin core structure were reported by the group of Hesse in 1991.[49,50,51] In all these syntheses, the macrocyclization proved as the most difficult step.

The first total synthesis performed by Oka *et al.* was based on successive peptide couplings (Scheme 1). In their approach, *erythro*-γ-hydroxy-(L)-lysine **5** was directly employed without any protective groups leading to poor yields of only 0.2% for the overall synthesis of Glidobactin after several purification rounds.



Scheme 1. First total synthesis of Glidobactin A by Oka *et al.* in 1988.

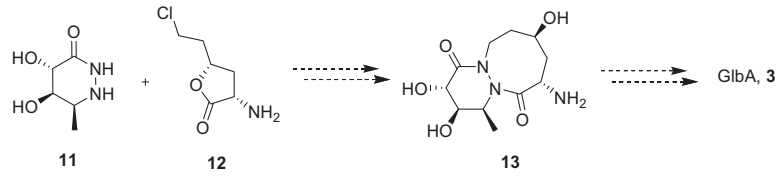
In the total synthesis of the Schmidt group, *erythro*-γ-hydroxy-(L)-lysine was generated in an orthogonally protected form by starting from the chiral butanetriol **8** (Scheme 2). In order to obtain this intermediate, 13 synthetic steps were necessary. In contrast to the Oka strategy, the macrolactamization was however performed between the C-terminal end of the alanine derivative and the ε-amino group of the hydroxylysine analogue. This key step proceeded in a much better yield of 20%. The overall yield of the total synthesis of GlbA however was only 0.8% due to the required 19 steps.



Scheme 2. Second total synthesis of Glidobactin A by Schmidt *et al.* in 1992.

Synthetic studies to the synthesis of the glidobactin macrocycle were also reported by the group of Hesse (Scheme 3). In their work, a strategic reductive N-N bond cleavage of a bicyclic intermediate **13** was examined to install the desired 12-membered ring system. While

model studies with simplified analogues led to promising results, the synthesis of GlbA macrocycle itself could not be accomplished.



Scheme 3. Attempted synthesis of Glidobactin A macrocycle by Hesse in 1991.

In summary, an overall analysis of the so far achieved syntheses of Glidobactin A indicate that although these ring systems are synthetically accessible, they are obviously synthetically challenging as overall yields were always below 1%. In fact, the macrocyclizations proved as a very difficult step and could be obtained in the best case with only 20% yield, leaving plenty of space for an improvement of the synthesis by the employment of alternative synthetic methodologies.