1 Introduction & Outline

The control of biological processes is a fundamental desire in medicinal chemistry because incorrect regulation causes diseases.[1] In the premodern era little was known about these processes and therefore, only few pharmaceuticals were established. Mainly anesthetics such as alcohol and opium but also other herbal ingredients were used. The first drugs that were synthesized on purpose were developed in the late 19th century. Admittedly, due to the lack of understanding biochemical functions, pharmaceuticals like the soporific chloral hydrate (1869) or the still very popular acetylsalicylic acid (1897) were designed after the principle of trial and error.[2] A rational design of pharmaceuticals was only possible after significant advances in protein crystallography and the development of nuclear magnetic resonance (NMR) spectroscopy. Proceedings in biochemical technologies allowed to understand biological processes, to find the cause of diseases, and to identify new targets for drugs.

Different approaches are made to develop new pharmaceuticals. For the rational design, a ligand, e.g. a natural product or a peptide that binds to the target, must be known. The ligand-target interactions are identified, and beginning from the ligand structure, the pharmaceutical structure is optimized in order to enhance the selectivity, the potency, the stability towards biological degradation mechanisms, and the bioavailability as well as in order to facilitate the synthetic access. This approach is often promising, but also very time-consuming and costly.[2] Further advances in technology offered additional options. Thus, in the 90s the high-throughput screening (HTS), a new technique that allowed to test an incredibly high number of compounds for their biological activity at one time, fuelled the hope that the search for lead structures could be dramatically facilitated.[3] Some positive examples are known, but altogether the output from HTS so far is disappointing measured against the affords that were made.[4]

The techniques are mostly used to develop small molecule drugs. However, due to their small size, a selective interaction with one target and not a whole target family is often difficult to achieve. The small size does not offer many interaction sites and these often flat molecules cannot efficiently fill the three dimensional pockets of enzymes or the binding sites of proteins. In terms of selectivity, natural products often achieve better results due

to their complex structure. Even so, their complexity is also the biggest disadvantage. It makes a synthetic production on an industrial scale impossible and can usually only be gained via extraction from plants or animals.[5] The best selectivity can be achieved by peptidic and protein based drugs. Yet, these drugs are very sensitive towards biological degradation processes. Peptides can be exchanged by peptidomimetics at the expense of a higher synthetic effort. Protein drugs often have the additional drawback that they usually cannot enter the cell, as a result they are mainly used for extracellular targets.[6]

A very interesting class of protein drugs is depicted by antibodies. These pharmaceuticals are mainly used in cancer treatment. Cancer cells differ from normal cells due to genomic mutations, which in some cases give raise to tumor-specific antigens. These cells can then be addressed selectively and destroyed by antibodies.[7] Monoclonal antibodies (mAbs) have already been successfully used in cancer treatment.[8] Nevertheless, many mAbs only have a low cytotoxic activity.[9] For this reason, a new class of anti-cancer drugs that combines the selectivity of antibodies with the potency of chemotherapeutic small molecules called antibody-drug conjugates (ADCs), has been developed.[10]

Besides the high selectivity towards certain antigens caused by the Fab region of the antibody, the long serum half-life also makes antibodies an interesting construct. The Fc region of the antibody is thought to be responsible for the high stability and is therefore an interesting scaffold for protein-binding domains or cell penetrating peptides that can be incorporated instead of the Fab region, and as a result a greater variety of targets can be attacked or cell permeability can be achieved.[11] Fc fusion proteins have already been marked as medicinal products, which shows their therapeutic potential.[12] The use of Fc regions has also been expanded by the design of a strand-exchange engineered domain (SEED) that delivers an Fc platform with two different Fc derived chains and therefore provides up to four different ligation sites.[13] This offers the possibility to design Fc conjugates with interesting combinations of recognition units (protein binding domains, CPPs), effector units (toxins) and reporter units (radiolabels, fluorophores). In order to introduce all kinds of cargo, a chemical modification method would be desirable.

Within the scope of this thesis, a novel ligation strategy shall be developed that later on would allow a chemical one-pot type modification of the Fc region at all ligation sites at the same time. This would open the possibility to design a library of different Fc conjugates.

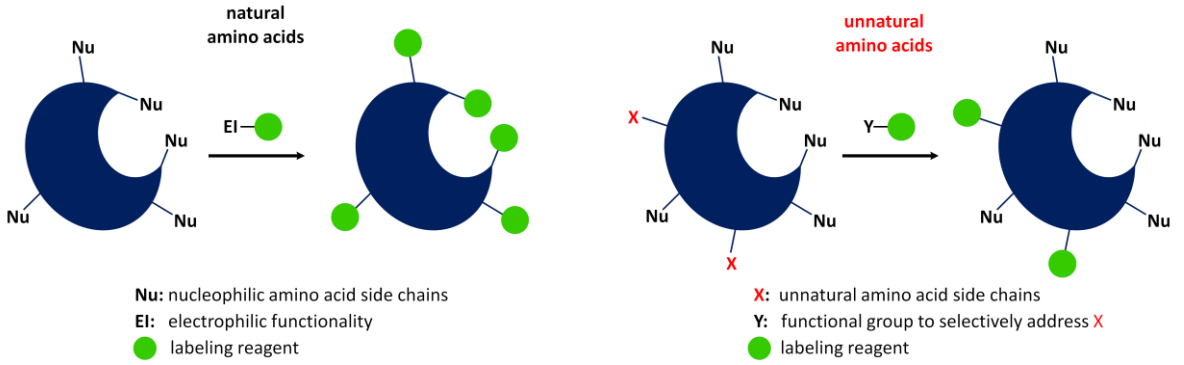
1.1 Protein Modification

Proteins conduct most of the biochemical functions in the cell. In order to understand these functions and to take influence on them, it is important to synthesize and modify proteins. Biochemists found ways to effectively produce proteins by expression techniques.[1] However, these methods are limited. Proteins that need complex post-translational modifications, like glycoproteins, are often excluded. Since only 20 genetically encoded amino acids can be readily incorporated, modifications, including post-translational ones are difficult to achieve and hard to control precisely.[14] Proteins can be modified when cargo with an electrophilic functionality is added, but these functionalities react with all surface nucleophiles of the protein. Therefore, the reactivity of the protein can be influenced in an undesired manner. Furthermore, all proteins abundant in the reaction medium will be labeled.[15]

Nevertheless, in some cases post-translational modifications can be conducted precisely. An *N*-terminal cysteine offers the possibility for a native chemical ligation (NCL)[16] or an *N*-terminal serine provides access to an aldehyde, via a specific sodium periodate oxidation, which can then be used for oxime ligation.[17] However, these examples are few and far between. A different approach is a purely synthetic route to produce proteins. Since the introduction of the solid phase peptide synthesis (SPPS) by MERRIFIELD in 1963, the synthesis of peptides with lengths of ≤ 50 amino acids has been strongly facilitated.[18] Nevertheless, a typical protein found in nature consists of about 300 amino acids.[14] By the development of powerful bioorthogonal ligation techniques, with native chemical ligation leading the way, bigger targets were accessible.[19] This method allows all imaginable kinds of modifications at every position in the protein. Unfortunately, the synthesis is very challenging and time consuming.

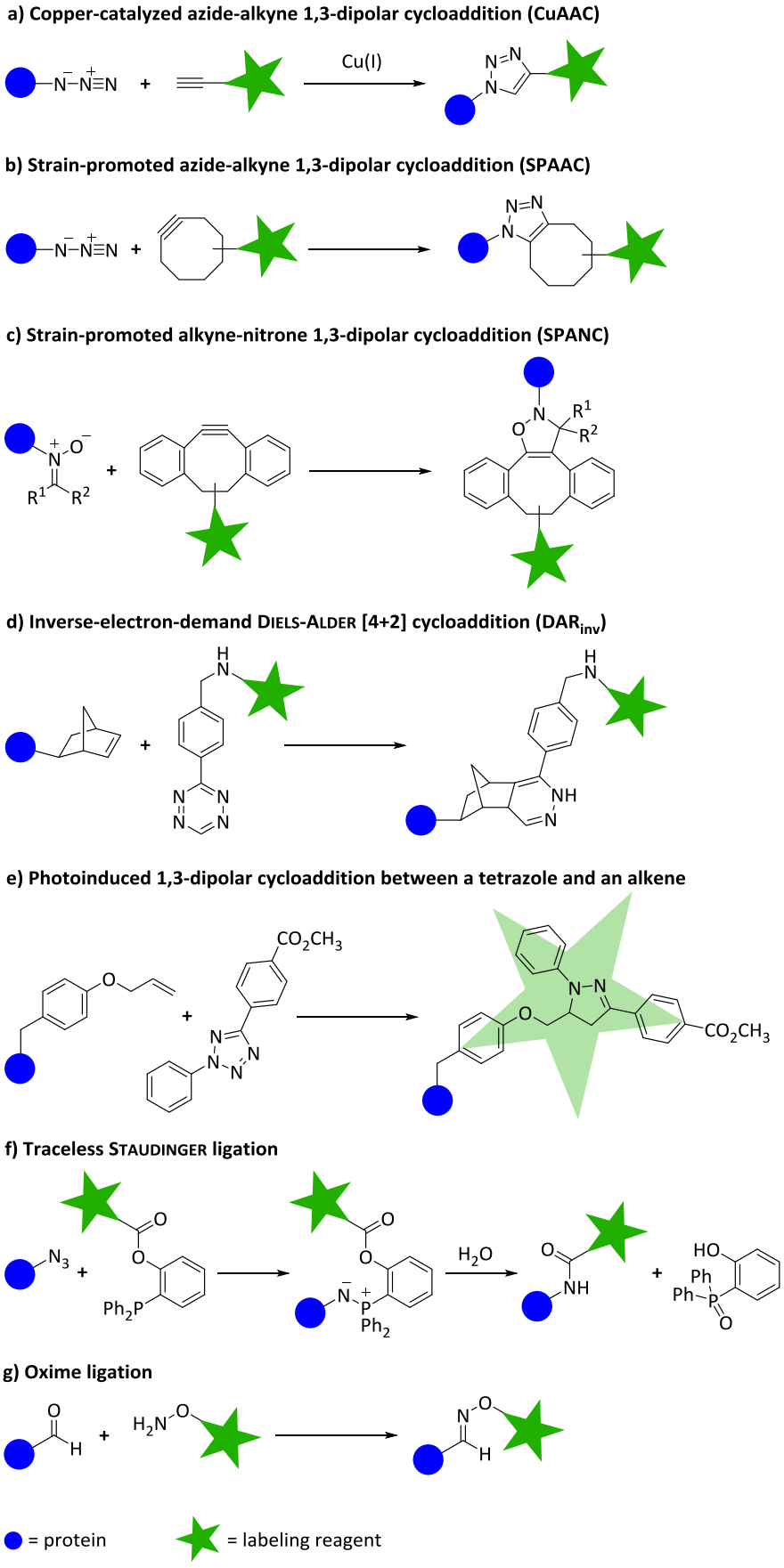
So, in the last few years two methodologies have been established and both can be seen as a compromise between the biological and the chemical approach. Semi-synthetic approaches are often used for big proteins that possess a modification, like glycoproteins, which cannot be added after translation. The part of the protein containing the modification is synthesized chemically, the other parts biologically and the fragments are ligated.[20] Another concept implies the incorporation of unnatural amino acids (UAAs) by expanding the genetic code within optimized protein biosynthetic machineries.[21] During

the last few years, it was possible to incorporate a whole variety of UAAs with different bioorthogonal groups, such as azides,[22] alkynes,[23] cyclooctynes,[24] alkenes,[25] tetrazoles,[26] ketones[27] or even iodides[27] and boronates.[28] These UAAs allow selective post-translational modifications on the proteins (Scheme 1). Bioorthogonal reagents have to perform a difficult balancing act between reactivity and stability. The reactants must be stable under physiologic conditions but still be reactive enough to assure a fast and efficient reaction. Also, the reaction must be chemoselective in a pool of countless different compounds.



Scheme 1. *On the left:* Modification of a natural protein using labeling reagents with electrophilic functionalities. *On the right:* Site-selective modification of a protein with incorporated UAAs (adapted from [15]).

Some reactions have frequently been used for protein modifications and are now established in this field of research. First of all, click reactions have to be named in this context. SHARPLESS' further development of HUISGEN'S azide-alkyne [3+2] cycloaddition[29] yielded the copper-catalyzed azide-alkyne [3+2] cycloaddition (CuAAC).[30] This reaction has become synonymous with the term of click chemistry, although many other reactions belong to this type of reaction (Scheme 2). The term of click chemistry was characterized by SHARPLESS and describes reactions that are high yielding, wide in scope, create only non-toxic by-products that can be removed without chromatography, are stereo- and regiospecific and are simple to perform.[31] These features make click reactions suitable for targeting molecules in complex biological environments. CuAAC reactions fit the requirements perfectly and therefore have been used for the modification of proteins.[32] Also interesting is that the resulting triazoles mimic *trans*-amide bonds because of their similar size, planarity, H-bonding capabilities and dipole moment.[33] Nevertheless, due to the cytotoxicity of copper, an alternative route has been developed.[34] Strain-promoted azide-alkyne 1,3-dipolar cycloadditions (SPAAC) offer a copper-free alternative. Although



Scheme 2. Frequently used bioorthogonal reactions for modification and labeling of proteins.

these reactions are slower compared to CuAAC, the fact that no additional catalysts or additives are needed makes them very attractive and SPAAC has been widely used for protein modifications.[35] Cyclooctyne derivatives do not only react with azides, but also with more reactive 1,3-dipoles such as nitrones via a very fast strain-promoted alkyne-nitrone cycloaddition (SPANC).[36] Inverse electron DIELS-ALDER [4+2] cycloadditions (DAR_{inv}) of tetrazines and strained alkenes belong to the fastest known click reactions (Figure 1). Different alkenes have been established and were used for protein modifications. Norbornene substrates are popular due to their stability against nucleophiles[37] and cyclopropenes offer a small-sized alternative,[38] but by far the most popular alkenes are *trans*-cyclooctene (TCO) derivatives due to their extremely high reaction rate.[39] The major advantage of the photoinduced 1,3-dipolar cycloaddition of tetrazoles and alkenes lies in its inducibility by light. This provides an exciting tool for spatiotemporal labeling reactions in living systems.[40] Besides click reactions, other bioorthogonal reactions like the STAUDINGER ligation[41] or the oxime ligation[42] have been approved for chemoselective bioconjugation reactions. However, it is important to note that these reactions are not useful for *in vivo* labeling, due to their comparable low reaction rate.

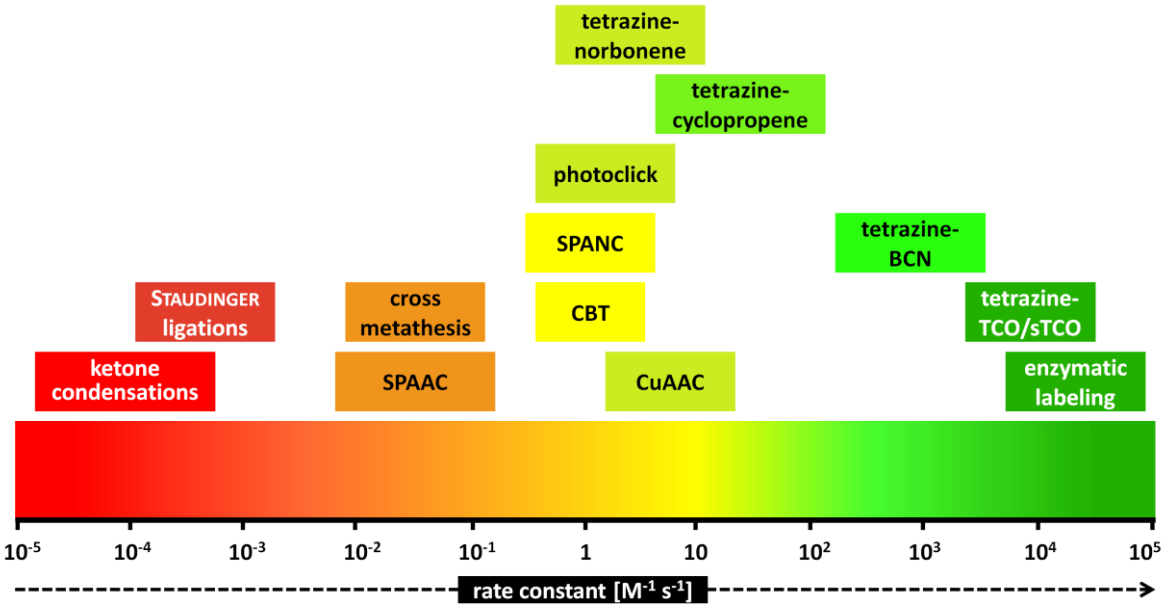
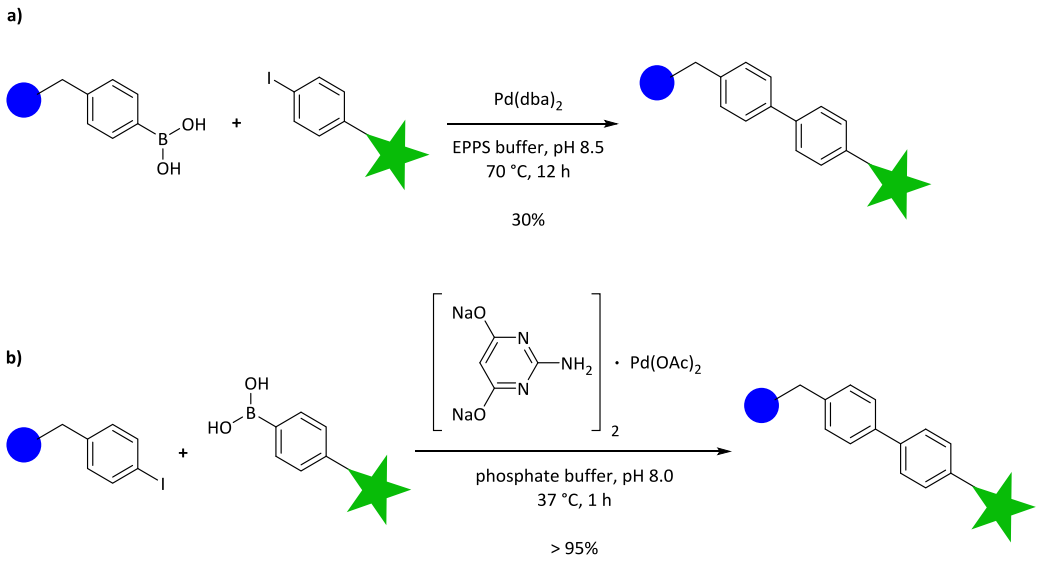


Figure 1. Rate constants of chemoselective reactions (CBT: 2-cyanobenzothiazole ; BCN: bicyclononyne; sTCO: strained *trans*-cyclooctene {adapted from [43]}).

Interestingly, in some examples it could be shown that many of these chemoselective reactions are also orthogonal to each other. This allows to modify proteins selectively at two or more sites. Orthogonality in bioorthogonal reactions could be shown among others

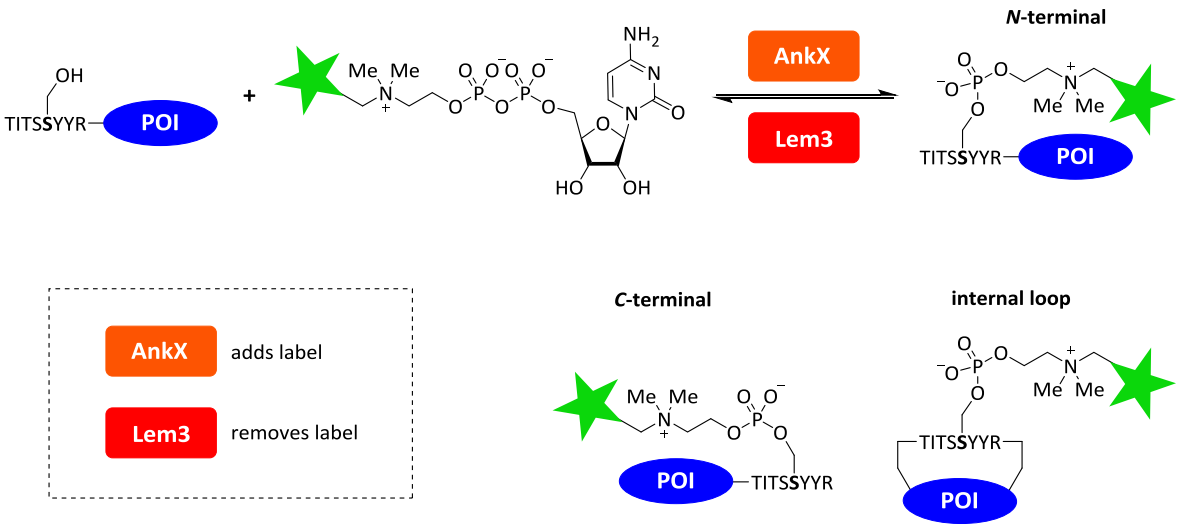
for DAR_{inv} and SPAAC reactions,[44,45] DAR_{inv} and oxime ligation[46] or oxime ligation and sortase A ligation.[47] In order to further exploit the toolbox of bioorthogonal reactions, a high-throughput screening was performed to identify new chemoselective reactions. Indeed, a series of novel 1,3-dipolar cycloadditions has been found that theoretically fits the requirements in terms of chemoselectivity, speed and biocompatibility.[48] Nevertheless, the applicability of these reactions still has to be proven. Besides the well-established concepts, metal-catalyzed cross coupling reactions are also increasing their popularity. Cross-metathesis is a very powerful carbon-carbon bond forming reaction and has been used for chemoselective protein modifications. However, at least a small amount of organic solvents had to be added and huge excesses of the ruthenium catalyst and the additives were necessary to achieve effective product formation.[49] Other alternatives are the palladium-mediated cross-coupling reactions. The absence of palladium from biological systems, its tolerance towards aqueous media and the excellent tolerance of its reaction partners towards nucleophiles make these reactions an interesting tool for bioorthogonal modifications.[43] The first attempts were not very promising, though. A MIZOROKI-HECK reaction between a *p*-iodophenylalanine and a labeled vinyl reagent only yielded 2% product after 50 h. Moreover, DMSO and high concentrations of additives had to be added.[50] A SONOGASHIRA reaction resulted in slightly better yields, but the reaction still did not proceed under physiological conditions.[51] The same can be stated for a SUZUKI reaction, where a reaction temperature of 70 °C was necessary to obtain the labeled product (Scheme 3a).[52]



Scheme 3. Palladium-catalyzed cross coupling reactions for the modification of proteins.[52,53]

However, things changed with the development of a new water-soluble 2-amino-4,6-dihydroxypyrimidine sodium salt (ADHP) ligand. After one hour reaction time, more than 95% conversion was obtained under physiological conditions (Scheme 3b).[53] Also, other Pd-catalyzed cross coupling reactions could be significantly improved by using the ADHP ligand.[54]

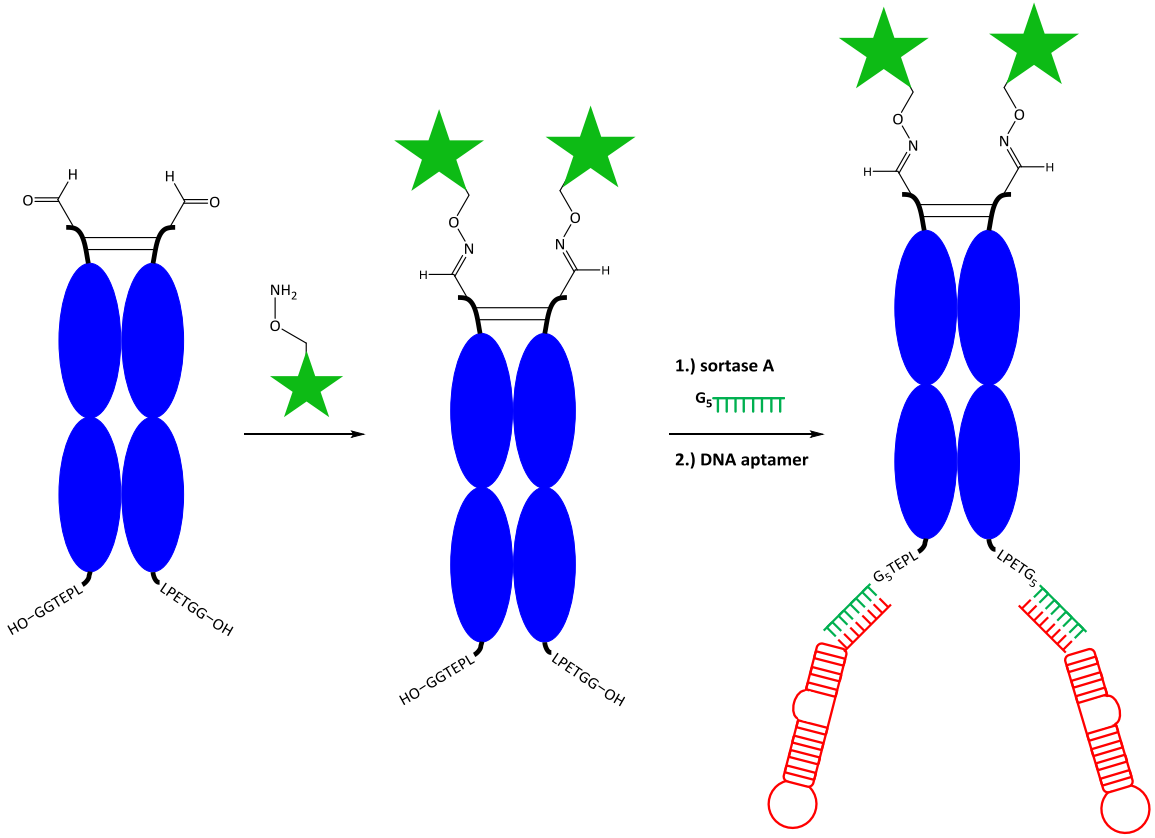
In another approach, enzymes are used to modify proteins. Owing to the very few ligases found in nature, proteases have been used for peptide ligation by reversing their proteolytic activity. Two basic strategies are known to manipulate proteases: the thermodynamically controlled[55] and the kinetically controlled approach[56]. A novel enzymatic protein labeling method was described by the ITZEN and the HEDBERG groups. They took advantage of the ability of the bacterial enzyme AnkX to identify a specific peptide sequence within proteins and to transfer a phosphocholine residue to it. By varying the natural substrate cytidine diphosphate choline and adding a fluorescent residue, it was possible to label the *C*-terminus, *N*-terminus or an internal loop of the protein of interest (POI). Furthermore, it was even possible to detach the label while adding the alkylphosphocholine resistance protein Lem3 (Scheme 4).[57]



Scheme 4. A novel enzymatic protein labeling method. Depending on the position of the octapeptide modification, the POI can be labeled at different positions (adapted from [57]).

One of the very few enzymes that are known to catalyze peptide ligations and have been successfully used for peptide ligations and protein modifications is the *trans*-peptidase sortase A. It cleaves a *C*-terminal LPxTG motif at the threonine residue and ligates it to a peptide containing at least one *N*-terminal glycine.[58] The KOLMAR group used a sortase A

ligation to introduce a peptide nucleic acid (PNA) strand to the *C*-terminus of an antibody Fc fragment and therefore converted it into a modular scaffold. Then, a DNA aptamer was added and attached to the Fc fragment via DNA/PNA base pairing. The *N*-terminus was modified with a fluorophore applying oxime ligation (Scheme 5).[47]

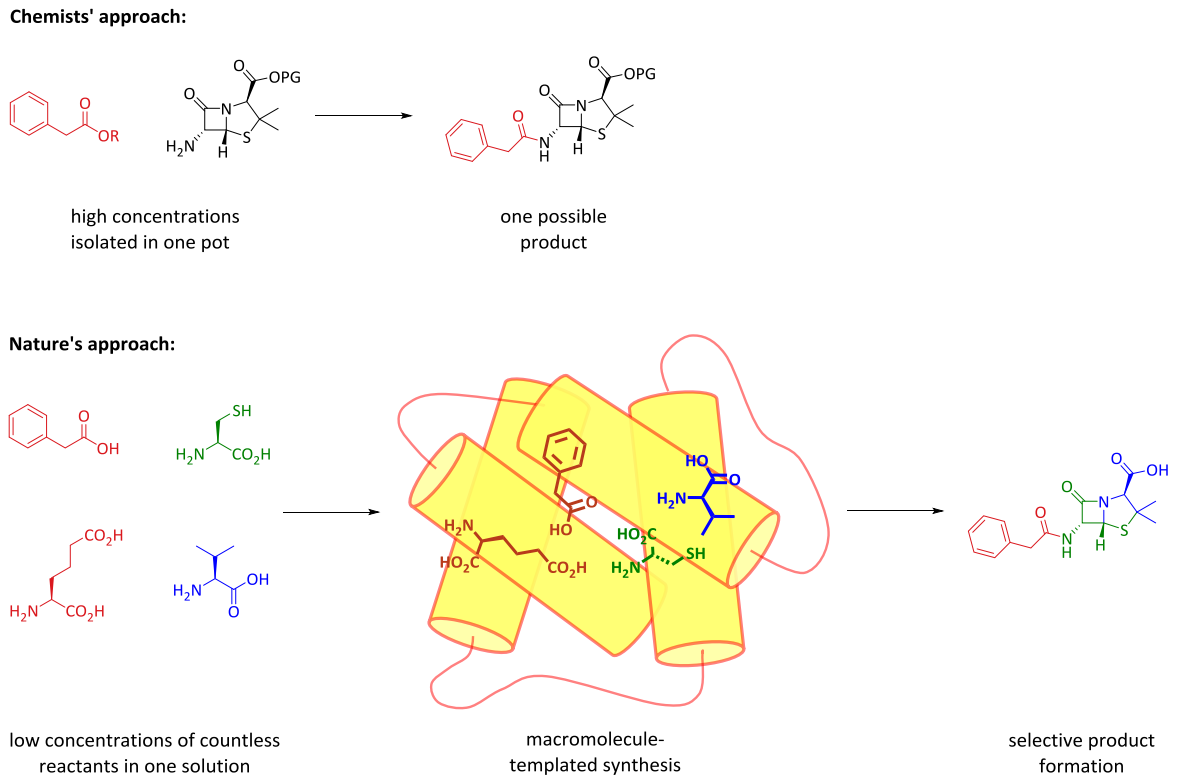


Scheme 5. Self-assembled hybrid aptamer-Fc conjugate (adapted from [47]).

In another attempt, antibody Fab fragments were equipped with PNA strands in order to create self-assembling multimers through PNA base-pairing. With this method, it was possible to access well-defined homo- and heterodimers, as well as higher-order multimers.[59] The WINNSINGER group used a DNA template as a modular scaffold for PNA encoded oligosaccharides that mimic the carbohydrate epitope of HIV.[60] The great advantage of templates lies in their ability to find each other within a medium of a great number of molecules. This ability was also used for covalent bond forming reactions, which will be discussed in the next chapter.

1.2 Template-Directed Reactions

The control of the selectivity and the rate of chemical reactions is a major concern of the natural sciences. In order to achieve this, chemists traditionally use a carefully picked pair of reactants in high concentrations, isolated in one vessel to obtain efficient product formation. Nature on the other hand can achieve highly efficient product formation in a medium with countless of different reactants in very low concentrations. The reason for this is that macromolecules can recognize reactants, bind them non-covalently, bring their reactive centers in close proximity and therefore increase the effective concentration of the reaction partners. These macromolecules are called templates (Scheme 6).[61]



Scheme 6. Chemists use a carefully picked pair of reactants in high concentrations in order to achieve an efficient reaction. Nature can accomplish this in the presence of many different reactants and low concentrations by using macromolecules as templates.[62]

This remarkably well operating machinery has gained a lot of attention from chemists who tried to copy this concept. Among these template based systems, nucleic acid templated synthesis plays an essential role in many biological processes. DNA and RNA, respectively, are polymeric structures that can sequence selectively bind complementary strands and form a great variety of different secondary structures like helices or hairpin structures.[1] They consist of a ribose-sugar unit with a phosphate residue that shapes the backbone and