# 1. INTRODUCTION

Researchers within the rapidly expanding field of chemical biology use the methods of synthetic organic chemistry to answer biological questions by generating and applying chemical tools, which enable the most sophisticated experimental setups. The fundamental goal of this extremely interdisciplinary area of research is to understand processes in biology on a molecular level.[1] Synthetic compounds are used for investigating biological systems in a multitude of different ways. They can function as biosensors for second messengers, with the most prominent example being calcium sensors such as fluo-2,[2] they can block or enhance signaling pathways by inhibition or activation of key enzymes in a number of different ways,[3] or they can be used to investigate enzymatic reactions when applied as isotope labeled substrates or activity probes.[4] A number of the most sophisticated microscopy techniques rely on synthetic dyes with tailor-made properties.[5–7]

One key aspect in chemical biology is the investigation of the dynamics of protein localization and activity. Proteins play key roles in biological processes. They are the final product of the translation of the information coded on DNA and the main components of biological pathways. In biological systems they take part in almost all cellular processes and carry out a multitude of functions, for example as structural elements, enzymatic catalysts, signal transduction elements or molecular transporters.[8,9] After their biosynthesis, proteins are further modified by phosphorylation,[10] glycosylation[11] and ubiquitination.[12] All these post-translational modifications influence protein functionality. Protein function can be investigated by various methods, e.g. the introduction of unnatural modifications, the blocking or enhancing of protein domains, the inhibition of enzymatic activity or the introduction of labels. These modifications have to be bioorthogonal (stable to water, reducing environment, nucleophiles and intracellular enzymatic degradation) to be applied in living cells.[13]

A large number of methods exists to chemically label naturally expressed proteins *in vitro*[14] or *in vivo*.[15,16] Prominent examples are the introduction of fluorophores to visualize proteins,[17,18] site-directed spin-labeling for EPR investigation[19–21] or labeling with radioactive isotopes.[22,23] Labeling can be achieved by introducing artificial amino acids by synthetic modification[24] or site-specific incorporation.[25] The most straightforward labeling approaches make use of the unique reactivity of cysteine residues and their low abundance in proteins.[26] A protein containing a single cysteine can be labeled in a site-selective manner by a straightforward nucleophilic substitution reaction with a haloalkane. Other approaches involve – among many other techniques - lysine functionalization, disulfide formation and click chemistry.[26] However, in many cases the structural homogeneity of the investigated protein sample is of crucial importance. Isolated protein material from biological samples is usually homogenous with respect to amino acid sequence but patterns of posttranslational modifications can be highly variable. As posttranslational modifications are probably the most important cellular mechanisms to regulate protein function,[27] activity and localization, the need of homogenous protein samples for high precision

experiments is apparent. The obvious solution again is chemistry – or more precisely the total or semi-synthesis of functional proteins. Even though the necessity of chemical protein synthesis has been clear since the early days of modern biochemistry, the development of an actually applicable synthetic strategy came as late as 1994 when Dawson and Kent introduced the concept of Native Chemical Ligation (NCL) in a study which is now considered to be one of the hallmark achievements in modern reaction method development.[28] Specifically, this method allows for connecting two unprotected peptide or protein fragments under formation of a native amide bond. Its extreme selectivity also enables the ligation of synthetic peptides with proteins expressed as C-terminal thioesters.[29] By introducing synthetic groups onto proteins, their function can be modulated and thus the importance of distinct domains or posttranslational modifications for the activity of the natural protein can be deduced. NCL is a very robust method with numerous literature known applications,[30,31] but its scope is limited by the requirement of a cysteine as the N-terminal amino acid of its C-terminal fragment. The initial publication almost instantaneously triggered efforts to develop methods for the ligation at other sites, but those published so far have their own limitations and disadvantages.[32] An easily applicable method which enables the formation of a native peptide bond featuring a reaction protocol as mild and selective as classical NCL has yet to be developed.

The central project of this thesis focuses on the development of a new auxiliary-based ligation method which should allow for ligation at other sites than X-Cys while being compatible with all proteinogenic amino acids and posttranslational modifications typically found in proteins. In order to achieve this goal, a novel photocleavable NCL auxiliary[33] is to be applied in ligation reactions. Photocleavage is a very mild method that can also be applied to biological systems.[34] Thus, the new auxiliary would have a great advantage over most known auxiliaries that are only cleavable under harsh acidic conditions and therefore it would be a valuable tool for introducing modifications into proteins.

In the field of protein labeling, this thesis concentrates on the synthesis of histidine-mimicking amino acids that are formed by Cu(I) catalyzed click chemistry of alkynes and azides. They are to be applied for introducing artificial metal binding sites into peptides. Furthermore, the same type of amino acids is used for developing easily variable zinc finger based fluorescent transition metal sensors, which could be applied as molecular tools for the investigation of the cell biology of metal ions.[35]

Another important subject in chemical biology is the investigation of enzymes and the analysis of their reaction mechanisms. In a third project, difluorinated linoleic acid is to be synthesized as a substrate analog for lipoxygenases to investigate the stereo- and regiospecificity of the catalyzed reaction by determining the substrate orientation via cocrystallization.

# 2. HISTIDINE-BASED ARTIFICIAL AMINO ACIDS FOR THE MODIFICATION OF METAL BINDING SITES IN BIOMOLECULES

## 2.1. ZINC FINGER DOMAINS: STRUCTURE AND FUNCTION

Zinc finger domains are one of the most frequent motifs in human transcription factors.[36] They were first characterized in the *Xenopus laevis* transcription factor IIIa, which is the first eukaryotic transcription factor described.[37] TFIIIa was identified as a 40 kDa protein which interacts with a DNA sequence of about 50 nucleotides within its target gene and is required for transcription.[38] It was observed that TFIIIa contains 7 to 11 zinc ions and its sequence is composed of repetitive domains of about 30 amino acids that contain four invariant residues – two histidines and two cysteines, the most common amino acid ligands for zinc ions.[39] As the metal-coordinating residues can be found at the beginning and the end of each domain as pairs separated by only three to four amino acids, it was postulated that the sequences in between would form extended loops called "fingers" which could "grasp" and bind the DNA.[39] The structure of these fingers was not fully determined until the first crystal structure of a zinc finger (Zif268) DNA complex was solved by PAVLETICH and PABO in 1991 (Figure 1a),[40] confirming observations made earlier on synthetic zinc finger peptides by protein NMR.[41,42] Zif268 is a protein that occurs in mice and its zinc finger domain is the best investigated to date.[36,43] The crystal structure shows that the zinc finger domain of Zif268 contains three individual fingers which bind successively in the major groove of B-DNA (Figure 1a).[40] Each finger is held in a ββα-fold by the coordination of the zinc ion. It contains two antiparallel β-sheets at its N-terminal end and an α-helix at the C-terminal end (Figure 1b). Zinc is coordinated tetrahedrally by two cysteine residues from the loop between the β-sheets and two histdine residues from the C-terminal end of the α-helix.

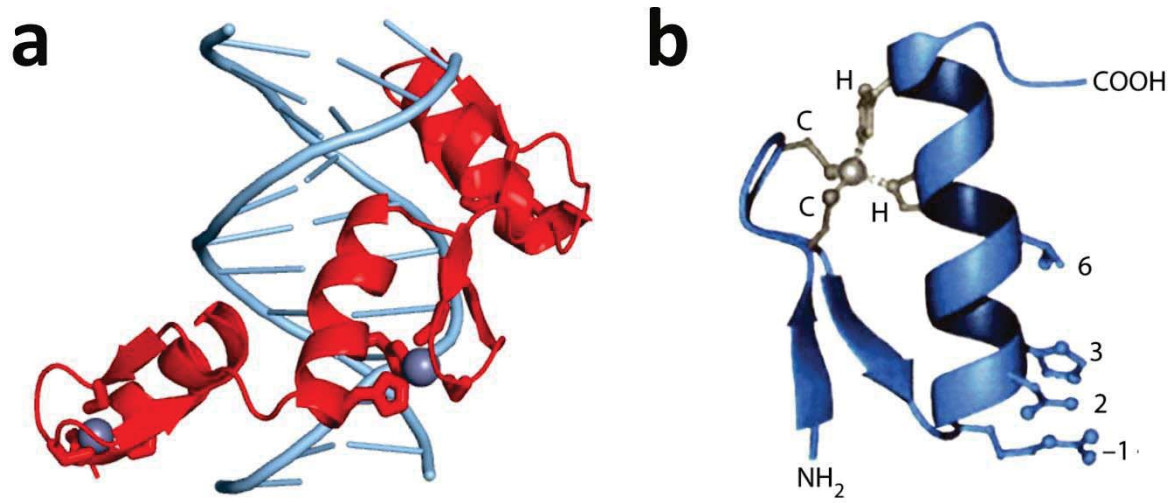



**Figure 1:** a) First crystal structure of zinc finger domain of Zif268 (red) in complex with DNA (blue) by PAVLETICH and PABO.[40] b) A single zinc finger of Zif268 in its characteristic fold featuring two β-sheets at the N-terminal and an α-helix at the C-terminal end. The zinc ion is bound by two cysteines in the loop between the β-sheets and two histidine residues from the C-terminal end of the α-helix. Residues in the α-helix which are responsible for DNA binding are highlighted.[44]

By the same crystal structure it became apparent that four residues at the amino terminal end of the each α-helix maintain specific contacts with the DNA double helix and each finger was found to bind to a three base pair subsite (Figure 1b). Thus, Zif268 recognizes a nine basepair site within the target DNA. Contacts to the DNA are mainly made to the primary DNA strand.[40]

Zinc finger motifs from TFIIIa and Zif268 both belong to the $C_2H_2$ family of zinc fingers, meaning that zinc is coordinated by two cysteines and two histidines. Beside this classical, most prevalent zinc finger family other families are known which are also classified by the coordination sphere of the zinc ion: The $C_3H$, $C_2HC$, $C_4$, $H_2C_2$ and GATA family zinc fingers all coordinate one zinc ion as well, while the GAL4, $C_6$, $C_8$ und $C_3HC_4$ zinc fingers coordinate two.[45–47] All these zinc finger domains show distinct different structures defined by the coordination of the zinc ion and can bind selectively to DNA, as well as to RNA and other proteins.[36,48,49] The structural diversity of zinc finger motifs enables specific recognition of a large variety of DNA sequences.

Over 4000 classical $C_2H_2$ zinc finger domains have been identified in more than 700 proteins.[36] The absence of zinc leads to an unfolded structure.[50] Exchanging one of the cysteine and histidine residues for another amino acid also leads to loss of structural integrity.[51] All classical zinc finger domains from the $C_2H_2$ family share a consensus sequence: F/Y-X-**C**-$X_{2-4}$-**C**-$X_3$-F/Y-$X_5$-Ψ-$X_2$-**H**-$X_{3-5}$-**H**, where X is any amino acid and Ψ stands for an amino acid bearing a hydrophobic side chain, e.g. leucine (Figure 2).[52] The three conserved hydrophobic residues form a hydrophobic core upon the folding of the domain which is thought to shield the zinc binding site thus further stabilizing the structure.[53,54]

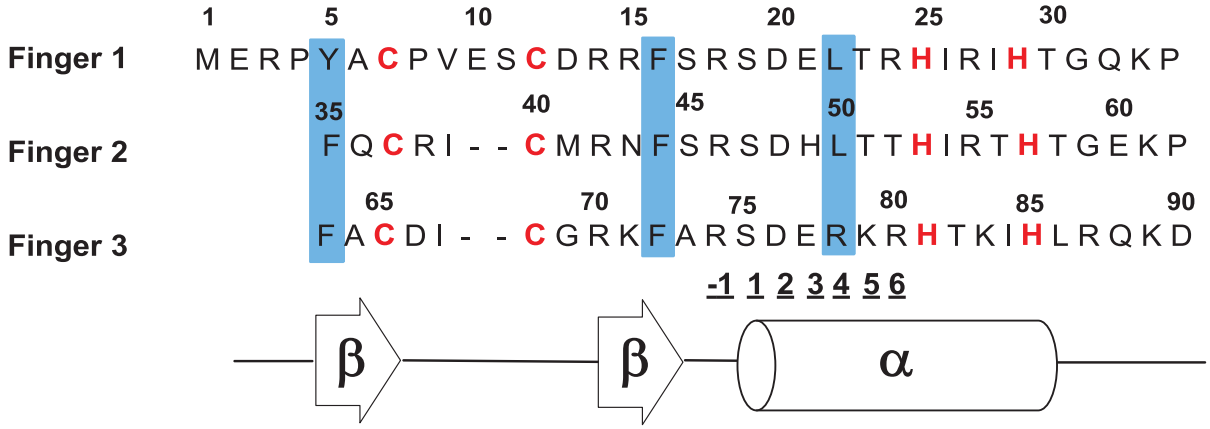



**Figure 2:** Sequence of Zif268 zinc finger domain. Zinc binding amino acids (red) and conserved hydrophobic residues (blue) are highlighted, as well as the positions of the β-sheets and the α-helix.([55], adapted from figures in [52,53,56])

As already mentioned, residues within the α-helix maintain the contact with the DNA bases for DNA binding. Therefore, the sole α-helix of the zinc finger motif is also referred to as the recognition helix.[43] The affinity of a zinc finger domain to DNA rises slightly with the number of fingers participating,[52] while it is known that two or three fingers are sufficient for selective DNA binding.[45] This is why zinc fingers are usually found in tandem repeats of two to thirty fingers connected by short linkers, mostly containing the sequence TGEKP.[57]

To achieve DNA binding, the protein winds around DNA in the major groove (Figure 1a).[58] Binding of Zif268 to DNA leads to a deepening of the major groove ($B_{enlarged\ groove}$-DNA), raising the number of base pairs per helical turn to 11.2, causing a displacement of 1.6 Å of the base pairs when compared to canonical B-DNA.[53]

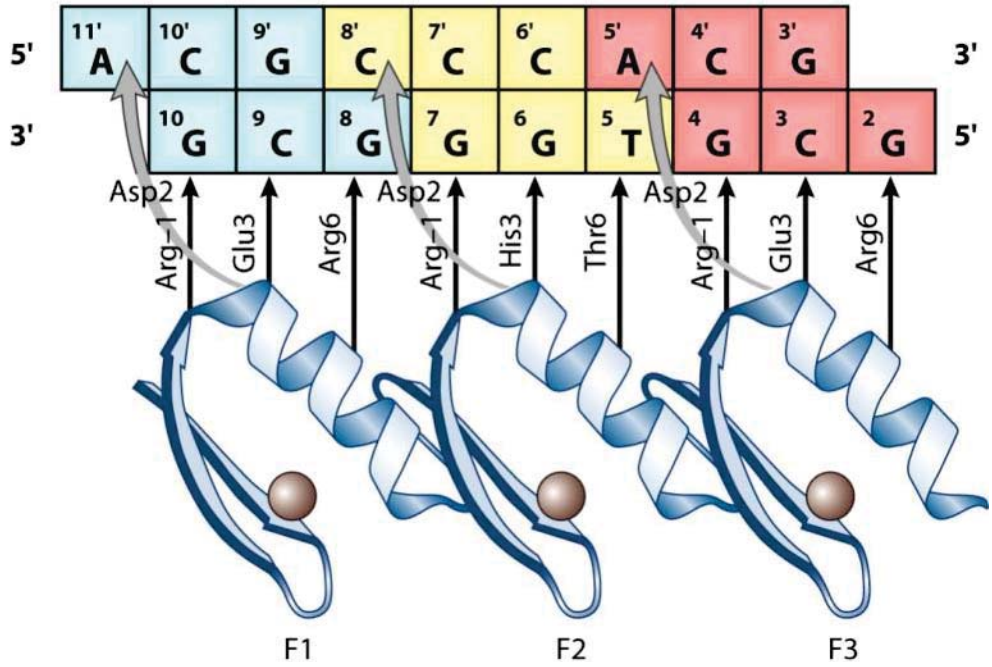



**Figure 3:** Schematic representation of the binding of Zif268 to DNA by specific interaction with the amino acid residues at position -1, 2, 3 and 6 of the α-helix of each zinc finger.[44]

In Zif268 the binding to DNA crucially depends on four amino acid residues located at the N-terminal end of the α-helix of each finger. Thereby, the residues in the positions -1, 3 and 6 of each α-helix are positioned so they can bind to three subsequent bases on the primary DNA strand, while the aspartate at position 2 can bind to the base preceding the triplet on the complimentary strand (Figure 3). The bound DNA region is G/C-rich and most of the contacts are interactions between arginine and guanine.[56] The arginine at position -1 binds to the guanine at the 3'-end of each base triplet by a hydrogen bond.[53] The aspartic acid at position 2 of each α-helix makes a hydrogen bond to the complimentary base of the base directly preceding the triplet, but it also stabilizes the positioning of the arginine at position -1 by hydrogen bonding to this residue.[53] At the same time, the second base of the triplet interacts with the residues at position 3 of the α-helix, in zinc fingers 1 and 2 through a hydrophobic interaction between glutamic acid and cytosine, while the histidine at this position in finger 2 is connected to the guanine in the middle of the respective base triplet by the formation of a hydrogen bond.[52] The 5' base of the triplet is bound by a hydrogen bond to the arginine at position 6 of the α-helix in zinc fingers 1 and 3, while the threonine at this position in the second finger is not known to make any contact to the respective 5'-base of its triplet.[52] Thus, each finger of Zif268 recognizes a four base pair subsite of the DNA sequence, while the recognition triplets on the primary strand are just three base pairs long and directly adjacent.[52] The overlapping of the recognition sites allows for very specific DNA

recognition. By variation of this recognition pattern many DNA subsites can be addressed by this motif and the variation of a few amino acids.

Variation of the amino acids that make contacts to DNA have shown that alternation of these residues results in zinc finger proteins with specificity to other base pair triplets.[59] Based on these modifications and the testing of their specificity to different DNA sequences by *phage display* techniques a recognition code has been proposed.[60,61] This recognition code only applies to Zif268-like classical zinc finger domains, however, it makes the engineering of zinc finger proteins tailored for site-specific DNA binding possible. One of the first examples of a tailored zinc finger protein was a construct targeted at an oncogene formed by chromosomal translocation, which led to repression of the gene's effects *in vitro* and *in vivo*.[62] Several interactions postulated by the recognition code were proven by crystallographic studies of modified zinc fingers and DNA.[63] The scope of the recognition code is limited, as it does not take into account that the amino acid residues also interact with each other and with the DNA backbone.[52] The model also disregards interactions between the individual fingers.[52] Furthermore, most recognized triplets contain the sequence GNN in a guanine-rich overall sequence and thus the concept is not widely applicable to all DNA sequences.[59,64] For this reason, the recognition code is a model for the design of artificial DNA-binding proteins, but it has its limitations.

To target a unique site within the human genome ($3 \times 10^9$ base pairs), a sequence-specific DNA-binding peptide would have to recognize more than 16 base pairs ($4^{16} = 4 \times 10^9$).[45] Naturally occurring tandem zinc finger arrays usually only target shorter DNA subsequences, often only using a few selected fingers.[65,66] Recognition of longer DNA sequences is a rare ability with DNA-binding proteins in general.[67] The formation of artificial polydactyl zinc finger proteins is thus an important feature to allow for high specificity of engineered zinc finger proteins. 6-, 9- and 15-fingered artificial zinc finger proteins have been developed which bind to 18, 27 and 45 base pair DNA regions, respectively.[67–69]

Artificial zinc finger proteins have been produced for a number of purposes, including the inhibition of HIV gene expression,[70] the activation of the expression of vascular endothelial growth factor (VGEF)-A,[71] induction of DNA bending[72] and gene editing utilizing sequence-specific zinc finger nucleases (ZFN).[73] For the latter a double strand break (DSB) is induced by engineered zinc finger proteins, which consist of a zinc finger tandem domain (usually 3 or 4 consecutive fingers) fused to the non-sequence-specific cleavage domain of the restriction enzyme FokI type II (Figure 4).[44] As the FokI domains have to dimerize to induce double strand breaks,[74] the zinc finger hybrids are introduced as pairs, targeted at two DNA sequences on both sites of the cleavage site with a 6 base pair spacing between the two half sites.[44,75] Therefore, an 18 to 36 base pair region can be targeted by the use of engineered zinc finger proteins which renders this system highly specific within mammalian genomes.[73,76] The cleaved DNA sequence can be edited in different ways after the double strand break. It is repaired by non-homologous end joining (NHEJ) in the absence of a donor leading to gene disruption by small deletions or insertions, which can be applied for gene knockout.[73] A simultaneous cleavage at a second site would lead to a large deletion. If a donor DNA is provided, the break will be resolved by the homology-directed repair

mechanism (HDR).[44] In case the donor is a single nucleotide or short sequence, this may lead to gene correction, while the provision of a larger transgene donor leads to gene addition. This has for example been applied in gene therapy of human Severe Combined Immunodeficiency (SCID), a monogenic disorder caused by point mutations.[77,78]

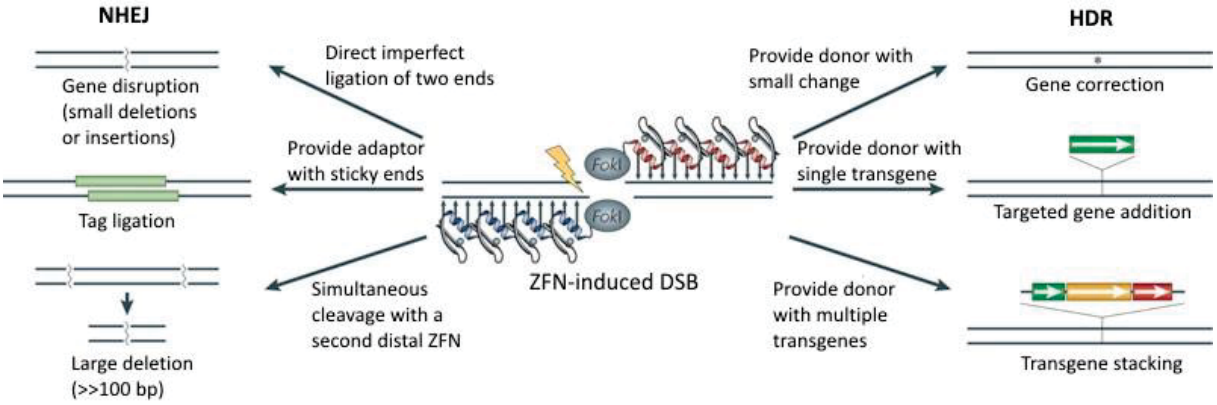



**Figure 4:** Possible ways of gene editing after a zinc finger nuclease induced double strand break in absence (left) of presence (right) of a donor DNA.[73]

## 2.2. DESIGN PRINCIPLES AND APPLICATION OF ZINC SENSORS

Zinc is the second most abundant transition metal in the human body and plays an important role in biological processes.[79] Tightly bound zinc ions serve as structural cofactors in metalloproteins such as zinc finger proteins[80] and are present in the active sites of hydrolytic enzymes such as carbonic anhydrase,[81] while the "free zinc(II)" pool is associated with neurological processes.[82] The understanding of the zinc metabolism is crucial, as a disorder can lead to severe neurological diseases such as Alzheimer's or Parkinson's disease.[83] In contrast to other metal ions present in biological systems, $Zn^{2+}$ is spectroscopically and magnetically silent and cannot be detected by common analytic techniques such as UV-VIS, NMR, Mößbauer or EPR spectroscopy due to its $4s^0$ $3d^{10}$ electronic configuration. The analytical method of choice is visualization by fluorescence spectroscopy utilizing compounds that can bind zinc ions and show a change in fluorescence upon binding.[84] This sensitive method allows for real-time imaging of living cells.[84] Desirable characteristics for a fluorescence sensor besides metal affinity and selectivity are brightness, an excitation wavelength above 400 nm to minimize UV cell damage and an emission wavelength or intensity change upon zinc binding.[85]

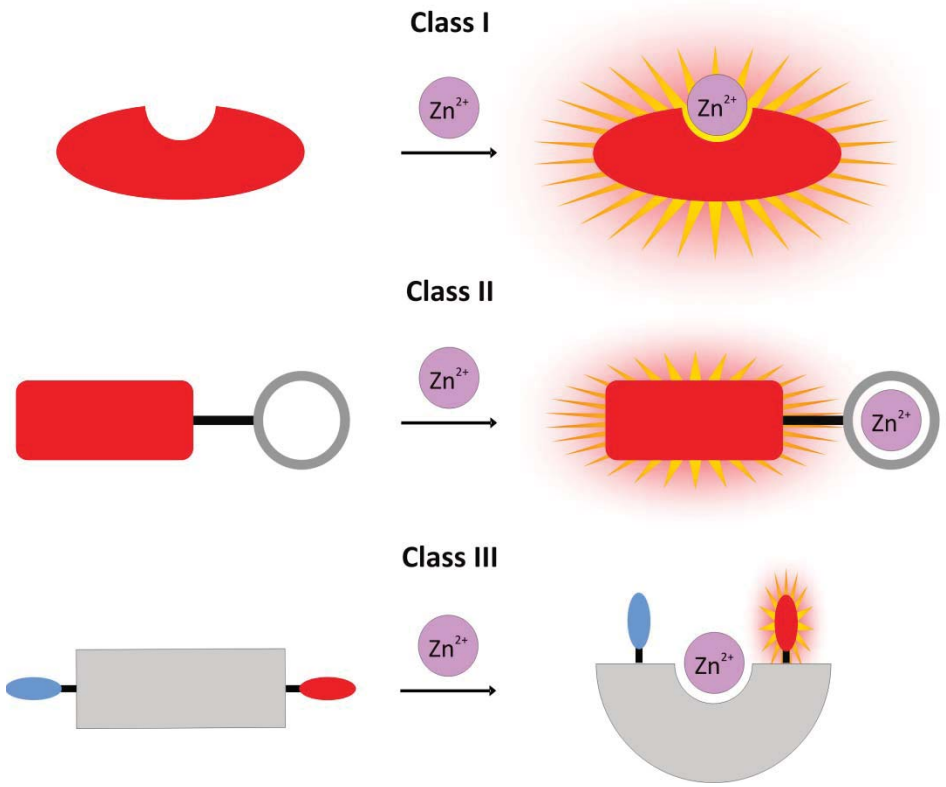



**Figure 5:** Classes of zinc sensors: Fluorogenic chelators (class I), fluoroionophores (class II, fluorophore (red), ligand moiety (grey) and linker (black)) and FRET sensors (class III, donor (blue) and acceptor (red)).(adapted from [84])

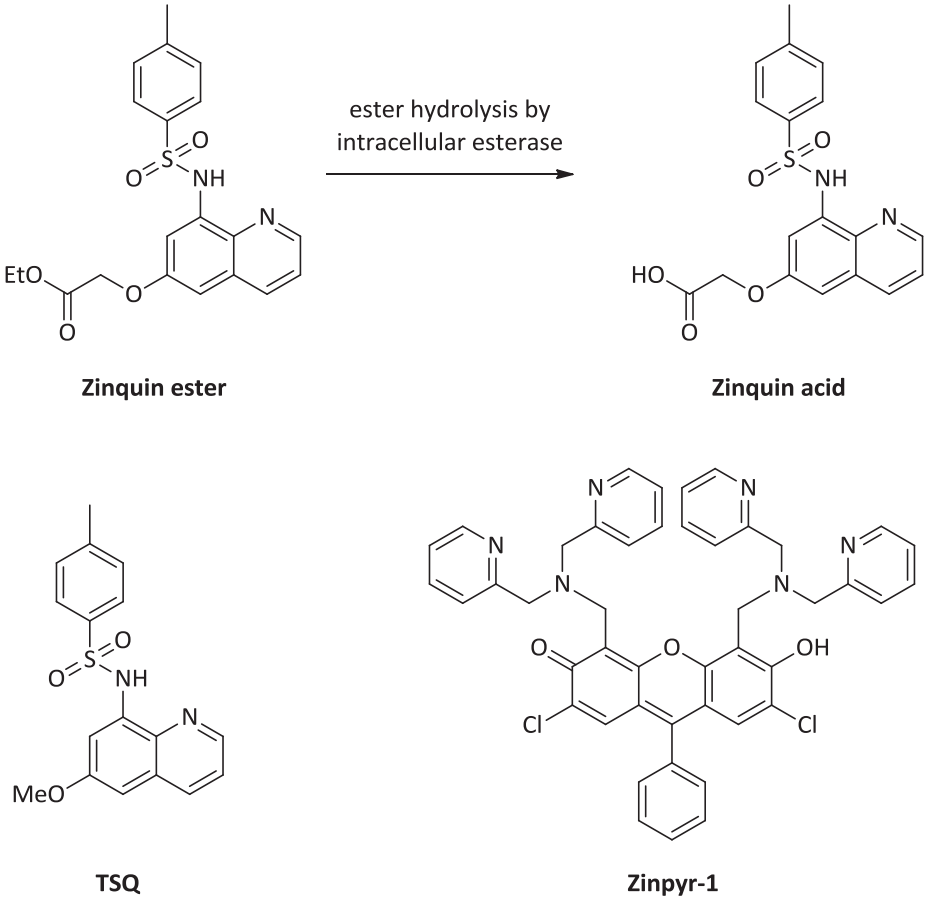
There are three classes of small molecule metal ion fluorescence sensors that can be categorized by their different modes of ion coordination (Figure 5).[84] Class I are fluorogenic ligands, class II sensors consist of an ion chelating moiety connected to a fluorophore by a

linker and class III sensors enable fluorescence resonance energy transfer (FRET) upon zinc binding.[84] Class II sensors are the most prevalent among the three classes, as they allow for easy variation of fluorophore and ligand.[84] Two photo-induced processes are mainly responsible for the fluorescence change upon metal ion coordination: photo-induced electron transfer (PET) and photo-induced charge transfer (PCT).[86] In PET sensors the ligand moiety is comprised of an electron donor and the fluorophore is an acceptor. Upon excitation, an electron in the highest occupied molecular orbital (HOMO) of the fluorophore is transferred to the lowest unoccupied molecular orbital (LUMO), which allows one of the electrons from the HOMO of the ligand electron donor to be transferred to the HOMO of the fluorophore, quenching the fluorescence as the excited electron cannot return to the ground state by a fluorescence pathway. If a metal ion is coordinated by the ligand moiety, the HOMO of the ligand is lowered in energy, inhibiting the electron transfer, thereby enabling fluorescence upon excitation.[84,86] PCT sensors contain an electron-donor and an electron-acceptor within the fluorophore, allowing for intramolecular charge transfer upon excitation. Depending on whether the coordinated cation interacts with the donor or the acceptor functionality of the fluorophore, the fluorescence response is blue or red shifted.[84,86]

Fluorescent zinc sensors can generally be divided into chemosensors and biosensors, depending on the origin of the metal-chelating moiety.[84] One of the first chemosensors developed was the quinoline-based 6-methoxy-8-*p*-toluenesulphonamido-quinoline (TSQ), which was used for *post mortem* histochemical stains of the brain and other tissues (Scheme 1).[85,87] TSQ exhibits strong fluorescence at 495 nm upon zinc coordination and shows high selectivity in presence of $Mg^{2+}$ and $Ca^{2+}$, but has its limitations because it is only sparingly soluble in water.[84,88] While it is not applicable for live cell imaging, TSQ is still widely used for staining tissue sections.[88]

Several sensors based on TSQ have been developed to increase biocompatibility and solubility, of which Zinquin is the most prominent (Scheme 1).[84] Zinquin can be introduced into cells in its ester form and is intracellularly converted into the corresponding carboxylic acid form by esterases,[89] a loading protocol first applied for calcium sensors which are membrane permeable in their acetoxymethyl ester form.[90] Zinquin's initial weak fluorescence is turned on by subnanomolar concentrations of zinc, reaching saturation at 1 µM with a 20-fold fluorescence enhancement.[85] The fluorescence response is unaffected by the presence of other divalent cations[85] and was for example used to monitor the cellular zinc concentration flux during apoptosis.[91]

Another important class of zinc chemosensors is the Zinpyr sensor family. Zinpyr sensors consist of bis(pyrid-2-ylmethyl)amine (bpa) moieties which are known for specific chelation of zinc ions and fluorophores based on fluorescein.[84] Zinpyr-1 is the original sensor of this family and shows a 3-fold fluorescence enhancement upon zinc binding (Scheme 1).[84] A number of Zinpyr-based sensors have been developed and used in a variety of biological systems,[35] e.g. for the visualization of zinc-induced neuronal damage after epileptic seizures in rats.[92]




**Scheme 1:** Different chemosensors for fluorescent zinc detection.

Biosensors are fluorescent metal ion sensors based on biological peptides and proteins known to have a high affinity towards the targeted metal ion.[88] By this approach, lower cellular zinc concentrations can be visualized, as the affinity level for chemical probes is limited. While nanomolar affinity can be reached with chemosensors, wild type zinc proteins with picomolar or even femtomolar affinities are known.[84] The application of zinc finger peptides and proteins for this purpose is a logical consequence. WALKUP and IMPERIALI modified a zinc finger peptide based on the human zinc finger protein ZFY, a classical $C_2H_2$ zinc finger motif,[54] with a microenvironment-sensitive fluorophore.[93] One of the amino acids of the hydrophobic core was exchanged for an artificial amino acid bearing a dansyl amide based fluorophore.[93] Upon coordination of the zinc ion, the previously unfolded peptide assumes the known ββα-fold and the fluorophore residue is embodied in the hydrophobic core, which causes a blue shift and linear intensity increase of the fluorescence response.[93] While this sensor has a nanomolar zinc affinity, it is susceptible to oxidation of the cysteine residues.[85] As one of its applications is the analysis of zinc in sea water, it was tried to eliminate this major drawback by enhancing the oxidative stability through the exchange of one of the cysteines by an additional histidine or aspartate.[94] While those sensors were stable to oxidative conditions, the change in the zinc coordination sphere lead to a decrease of affinity towards $Zn^{2+}$ and a decrease in fluorescence response, even if different fluorophores were applied.[94]