Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by

Ayan Samanta, M.Sc.
born in: Howrah, West Bengal, India
Oral-examination: ……..

RNA functionalization strategies and their application to RNA folding dynamics and experimental RNomics

Referees: **Prof. Dr. Andres Jäschke**

Prof. Dr. Stefan Wölfl

Acknowledgements

I first would like to thank my supervisor Professor Andres Jäschke for giving me the opportunity to work in his laboratory. His constant enthusiasm, availability for scientific discussions and especially the thrilling and demanding ideas contributed substantially in developing my scientific personality and to what I am now. Starting from only being a synthetic organic chemist, I consider myself not the same anymore.

I am thankful to my second supervisor Professor Stefan Wölfl for reviewing my thesis and taking part in my oral exam.

I am especially grateful to my examiners (Professor Andres Jäschke, Professor Stefan Wölfl, Professor Ulrike Müller and Dr. Karsten Rippe) for such an easy and smooth scheduling of my oral defense exam.

I would also like to thank Professor Mark Helm and Professor Yuri Motorin for their constructive criticism regarding my scientific work. Special thanks to Yuri concerning project design and career development in science. I wish I had listened to him earlier.

I would like to express my sincere gratitude to Frau Viola Funk and to Frau Karin Weiß for taking such an outstanding care of all administrative issues during my stay in Germany and my work at IPMB.

I am also thankful to some other members in our department, namely Sandra Suhm for supporting me with large scale back up synthesis and to Heiko Rudy for measuring innumerous mass spectra for me. I should not forget to acknowledge our computer and NMR administrator Tobias Timmermann for helping me out with computer problems.

I would also like to thank Richard for his unique knowledge-transfer to me regarding photoaffinity crosslinking.

My sincere gratitude goes to Sandeep, Juliane, André and Marie for proof-reading my thesis. Special thanks to André for being so patient with me regarding the paper.

Furthermore, I would like to thank all past and present members of Jäschke and Helm research groups for providing me a creative and cheerful work atmosphere.

At the beginning, being a stranger in this country, it was not very easy for me but I had two good friends namely, Salifu and Markus (the deejay). Thank you for all your support and all the good time we spent together drinking beer.

I want to thank Marie and Ece from the deep core of my heart for being such good friends and constantly cheering me up for almost everything (I am already always cheered up for new scientific projects!). ``Ece, I wish you would have joined the group before and thank you Marie for all your support``. Perhaps, only when leaving in a place far away and being completely alone, people will understand what friends mean. I will really miss both of you. I believe, I learned a lot from both of you and in future will try to be a better person. Finally, although it is not a tradition in my home country, I want to thank my father for everything I achieved so far in my life. None of this would have ever been possible without his constant encouragement.

A scientific man ought to have no wishes, no affections – a mere heart of stone.

Charles Darwin

ABSTRACT **Ayan Samanta, M.Sc.**

RNA functionalization strategies and their application to RNA folding dynamics and experimental RNomics

1. Referee: Prof. Dr. Andres Jäschke

2. Referee: Prof. Dr. Stefan Wölfl

The oversimplified notion of RNA being a mere carrier of sequence information from gene to protein has been repeatedly undermined over the decades by yet another newly discovered function performed by certain RNA species. These new species include in particular RNAs which regulate gene expression in response to a metabolite sensing event. These RNAs — known as riboswitches — elegantly couple metabolite recognition with gene regulation in the apparent absence of protein helpers.

Here we first sought to investigate the folding dynamics of the S-adenosyl-L-methionine responsive riboswitch by FRET spectroscopy. This requires the synthesis of full-length riboswitch constructs site-specifically modified with multiple fluorophores. For this challenging task we have established a 5-way splinted-ligation strategy to prepare dual-fluorophore labelled full-length riboswitch constructs in an unprecedented overall yield of 10 %. These constructs have further been subjected to bulk and single molecule FRET spectroscopy for ligand induced folding analysis. We confirmed similar folding dynamics for the aptamer of the complete riboswitch (aptamer + expression platform) as reported earlier for constructs containing the aptamer alone. However, we also observed a few other folding phenomena induced by a chemically slightly different, yet non-cognate metabolite, which cannot be explained by any facts known about this riboswitch to date and require further experiments to reach a final conclusion.

During the course of the aforesaid work, we realized the limitations of existing nucleic acid functionalization strategies. Therefore we decided to use bioorthogonal click reactions as part of our labelling strategy. Among various different click reactions, we first had to find the one which best suits our purpose and to optimize its conditions. Having the optimized click reaction conditions at hand, we developed enzymatic strategies to site-specifically functionalize long RNAs with clickable residues. In our nucleic acid labelling strategy a diverse array of different chemical functionalities can be introduced exploiting the modular nature of click chemistry. This does not demand either *de novo* synthesis or optimizations of enzymatic reaction conditions for each new single compound. Furthermore, we developed a chemical approach using two different mutually orthogonal click reactions for concurrent, site-specific labelling of DNA molecules with multiple fluorophores.

Moreover, we sought to extend this strategy of enzymatic, site-specific transfer of clickable residues to long RNAs towards photochemical transfer of clickable moieties to a target RNA in a mixture of many unrelated sequences. This technique, which we call ``Affinity-based Chemical RNomics`` is a chemical approach in experimental RNomics whereby RNA sequences which bind to a given small-molecule metabolite are to be isolated from a total RNA isolate of any organism just by the virtue of its tight binding to its cognate metabolite and without any prior knowledge of its sequence. This method would therefore allow for the discovery of previously unknown riboswitches, currently the only known kind of natural RNA that binds small-molecule metabolites. Since all currently known riboswitches have been discovered by rational approaches, this will considerably extend the chances of discovering new riboswitches.

ZUSAMMENFASSUNG

Die übervereinfachte Vorstellung, dass RNA lediglich als Träger von Protein-Sequenzinformation dient, wurde über die Jahrzehnte immer wieder durch neu entdeckte Funktionen, die bestimmte RNA-Spezies ausüben können, untergraben. Diese neuen Spezies beinhalten insbesondere RNAs, die Genexpression als Reaktion auf einen Metaboliten-Stimulus regulieren. Diese RNAs – bekannt unter dem Namen *Riboswitches*, wörtlich Ribo-Schalter – koppeln auf elegante Weise Metaboliten-Erkennung an Genregulation, ohne dafür Proteine zu benötigen.

In dieser Arbeit wollten wir zunächst die Faltungs-Dynamik des S-Adenosyl-L-Methionin-*Riboswitches* mittels FRET-Spektroskopie untersuchen. Dazu war die Synthese von Volllängen-*Riboswitch*-Konstrukten nötig, welche ortsspezifisch mit Fluorophoren modifiziert waren. Um dies zu bewerkstelligen, etablierten wir eine Strategie, bei der mittels eines "Splints" fünf Fragmente mit einer noch nie erreichten Effizienz von 10% ligiert wurden. Die doppelt farbstoffmarkierten Konstrukte wurden danach mittels „*Bulk*"- und Einzelmolekül-FRET-Spektroskopie auf ligandeninduzierte Faltung hin untersucht. So konnten wir für das Aptamer des gesamten *Riboswitches* (Aptamer + Expressionsplattform) ähnliche Faltungsdynamiken bestätigen, wie sie bereits für Konstrukte, die nur das Aptamer enthielten publiziert worden waren. Dennoch beobachteten wir auch einige andere Faltungs-Phänomene, welche durch einen chemisch nur wenig abweichenden, jedoch eigentlich *Riboswitch*-fremden Metaboliten induziert wurden, die durch keine der bisher über diesen *Riboswitch* bekannten Fakten erklärt werden können und weitere Experimente zur endgültigen Klärung benötigen.

Während der zuvor beschriebenen Arbeit bemerkten wir die Einschränkungen der bestehenden Nukleinsäure-Funktionalisierungsstrategien. Daher entschlossen wir uns, bioorthogonale Click-Reaktionen als Teil unserer Funktionalisierungsstrategie zu verwenden. Unter den verschiedenen Click-Reaktionen mussten wir zunächst diejenige identifizieren, die für unsere Zwecke am besten geeignet ist, und dann deren Reaktionsbedingungen optimieren. Danach entwickelten wir enzymatische Strategien um lange RNA-Moleküle mit „clickbaren" Resten zu versehen. Mit unserer Nukleinsäure-Markierungsstrategie kann unter Ausnutzung der modularen Natur der Click-Chemie eine Reihe verschiedener funktionaler Gruppen eingeführt werden. Dazu sind weder *de novo* Synthese noch Optimierung enzymatischer Reaktionsbedingungen für jede neue Verbindung vonnöten. Weiterhin entwickelten wir eine chemische Methode zur gleichzeitigen, ortsspezifischen Doppelmarkierung von DNA-Molekülen mit multiplen Fluorophoren unter Benutzung zweier wechselseitig orthogonaler Click-Reaktionen.

Desweiteren versuchten wir, die Strategie des enzymatischen, ortsspezifischen Transfers „clickbarer" Reste auf lange RNAs auf den photochemischen Transfer ebensolcher Gruppen auf eine Ziel-RNA in einer Mischung von vielen nicht-verwandten Sequenzen auszuweiten. Diese Technik, die wir als „affinitätsbasierte chemische RNomik" bezeichnen, ist eine chemische Methode in der experimentellen RNomik. Durch diese sollen RNA-Sequenzen, die an einen gegebenen niedermolekularen Metaboliten binden, ausschließlich aufgrund ihrer festen Bindung an diesen Metaboliten und ohne Vorwissen bezüglich ihrer Sequenz aus einem Total-RNA-Isolat eines beliebigen Organismus' isoliert werden. Diese Methode sollte es daher ermöglichen, zuvor unbekannte *Riboswitches* zu entdecken – die bisher einzige bekannte Sorte natürlich vorkommender RNAs, die niedermolekulare Metaboliten bindet. Da alle bisher bekannten *Riboswitches* über rationale Ansätze entdeckt wurden, wird dies die Möglichkeiten zur Entdeckung neuer *Riboswitches* deutlich erweitern.

Table of contents

1. Introduction

1.1. Small molecule binding natural RNA aptamers - riboswitches

The folding dynamics of structural RNAs are of widespread interest in modern life sciences(1,2). The prevalent notion of RNA as a mere carrier of protein sequence information in biology has been repeatedly undermined over the decades by yet another newly discovered function performed by RNA. These new functions include in particular RNAs which regulate gene expression in response to a metabolite sensing event. This regulation responds to changing levels of the metabolites in the cell, a type of feedback mechanism that was previously known to be mediated only by proteins(3). These RNAs — known as riboswitches — elegantly couple metabolite recognition with gene regulation in the apparent absence of protein helpers(4-12).

Numerous riboswitches that specifically recognize substrates as diverse as nucleotides, amino acids, vitamins and co-enzymes have been discovered in recent years. Accounting for about 4% of genetic control in bacteria, riboswitches have turned out to be quite frequent. After their initial discovery in bacteria, they were also found in archaea, fungi and plants, albeit so far in lesser numbers(13,14). The known regulatory RNA domains are typically located in regions of mRNA that directly precede the protein-coding sequence. A highly interesting common feature is the simple modular architecture — they consist of a metabolite-sensing domain, which binds the substrate, and another region, known as the expression platform, which in response to substrate binding mediates the gene regulation, i.e. usually protein production. Thus, the end result of the metabolite binding event is frequently the suppression of the production of enzymes that are responsible for the biosynthesis of the detected metabolite (Fig. 1.1).

In vitro investigations show, that upon substrate binding to the metabolite-sensing domain, a structural reorganization of the RNA occurs that unveils (or sometimes masks) the gene-expression signal. There is, however an ongoing discussion as to whether the actual mechanism *in vivo* relies on such a rearrangement, or if the metabolite directs the folding of the *de novo* synthesized RNA co-transcriptionally into one or the other conformation(15).

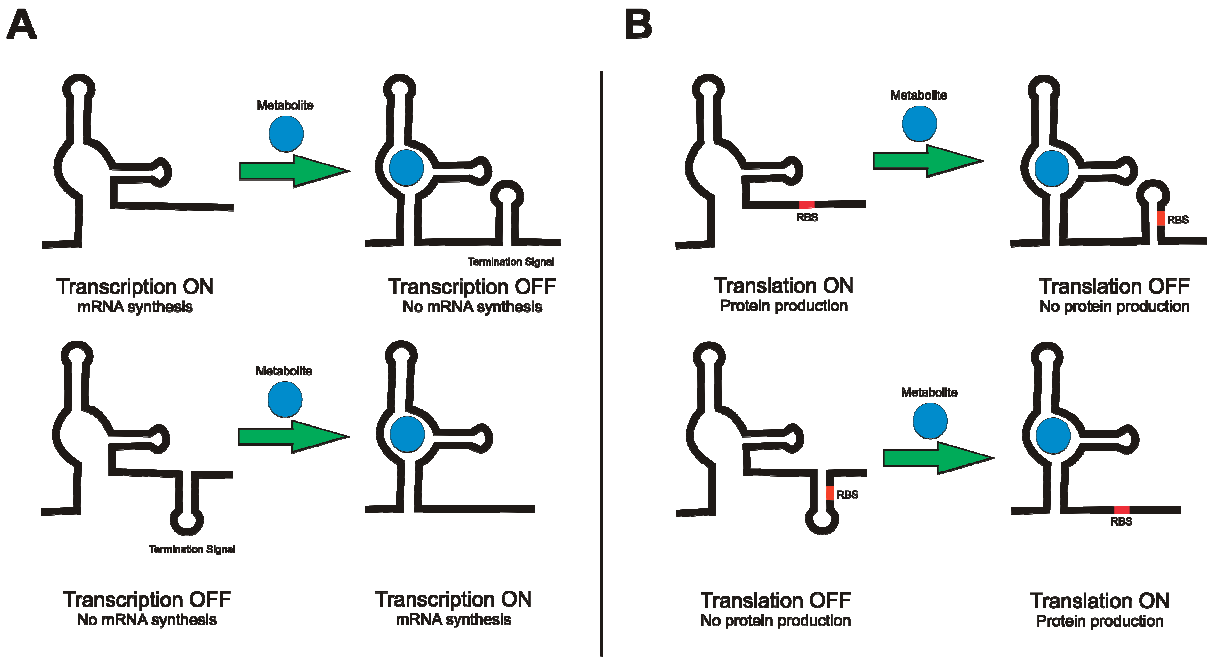


Figure 1.1: Schematic representation of A) transcription and B) translation regulating riboswitches. RBS: Ribosome Binding Site.

The metabolite sensing domain (the aptamer) is the distinguishing feature of each riboswitch class - consisting of unique and highly conserved structural motifs - which accounts for the high specificity for its cognate ligand, rivalling that of the repressor proteins(16). E.g. the guanine-responsive riboswitch shows a preference for guanine over adenine by 20,000-fold similar to that of the purR repressor(17,18). Similarly all known S-adenosyl-L-methionine (SAM) -responsive riboswitches discriminate between SAM and S-adenosyl-L-homocysteine (SAH), which differ from each other by only a single methyl group and a positive charge on sulphur (Fig. 1.2)(19), an ability that competes the metJ repressor in gram-negative bacteria(20). This high specificity and selectivity of the riboswitch aptamers for their cognate metabolites can be accounted from the fact that almost every single functional group present in the metabolite molecule is recognized by the riboswitch aptamer during the binding event.

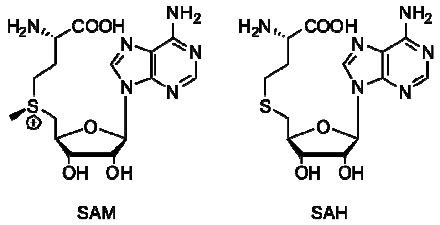


Figure 1.2: Chemical structures of SAM and SAH.

In past years, the availability of a number of different high resolution crystal structures for various riboswitch aptamers with or without their cognate metabolites enabled us to

categorize these aptamers in two distinct types(6). The type-I riboswitches consist of a single, localized binding pocket supported by a largely pre-formed global fold. Therefore the conformational changes induced upon ligand binding are mainly confined to a small region of the aptamer. A representative example of this type is the purine riboswitch. In contrast, the type-II riboswitches contain ligand binding pockets that are split into spatially distinct sites and therefore metabolite binding induces severe changes in both local and global architecture of the aptamer. The thiamine pyrophosphate (TPP) responsive riboswitch is a representative example of this class (Fig. 1.3).

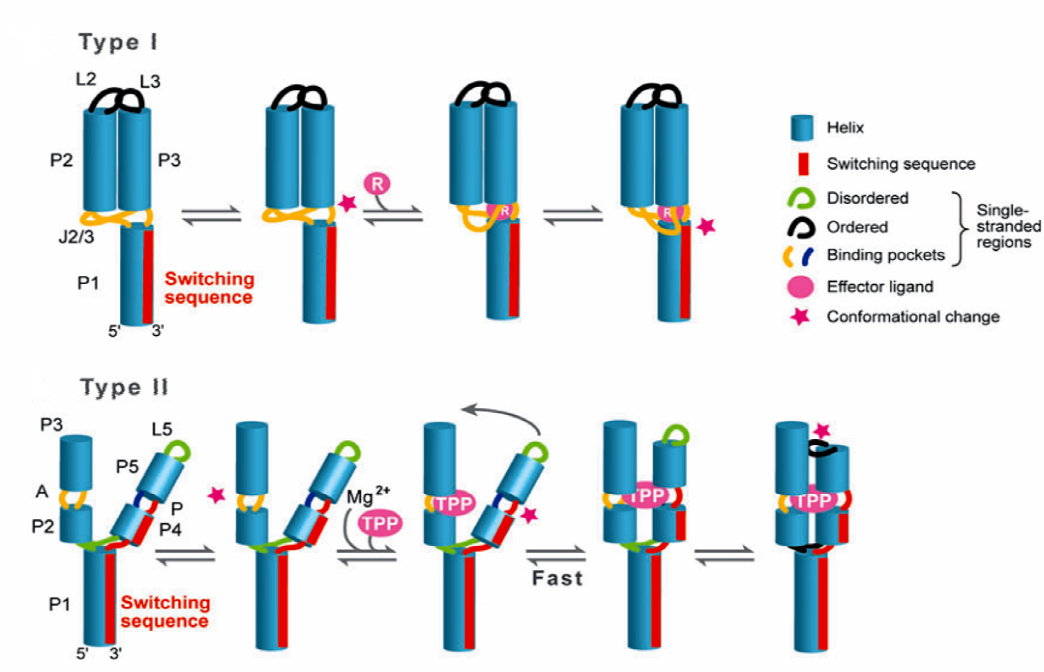


Figure 1.3: Cartoon representation of ligand-induced tertiary conformational changes in Type I and Type II riboswitches. Helices are represented as cylinders, single-stranded regions are drawn as lines (black is ordered, green is disordered). The arrows indicate large motions. The purine riboswitch is a representative example of Type I class and the TPP riboswitch is a representative example of Type II class [adapted from ref. 6].

1.1.1. The SAM riboswitch

S-adenosyl-L-methionine (SAM) is the most common methylating reagent in cells and is exclusively involved in nucleic acid, protein and sugar methylation. The SAM-responsive riboswitches have proven to be one of the most widespread among various kingdoms of life(7). Although it is not clear why some riboswitches are widespread, while others are exceedingly rare, it can be expected that rare riboswitch classes might be the result of either a lack of need by a cell to sense that particular metabolite or the existence of other protein-based sensing machineries for that metabolite.

To date, five distinct classes (named as SAM-I to V) of SAM-responsive riboswitches have been discovered (Fig. 1.4)(19,21), and representatives of these classes have been

found to strongly discriminate between SAM and S-adenosyl-L-homocysteine (SAH) which is the metabolic side-product when SAM is used as a cofactor in methylation reactions, although SAH lacks only a single methyl group and a positive charge on the sulphur compared to SAM (Fig. 1.2). However, some of these SAM-sensing riboswitch classes are only narrowly distributed in bacterial species, and they rarely coexist in the same organisms.

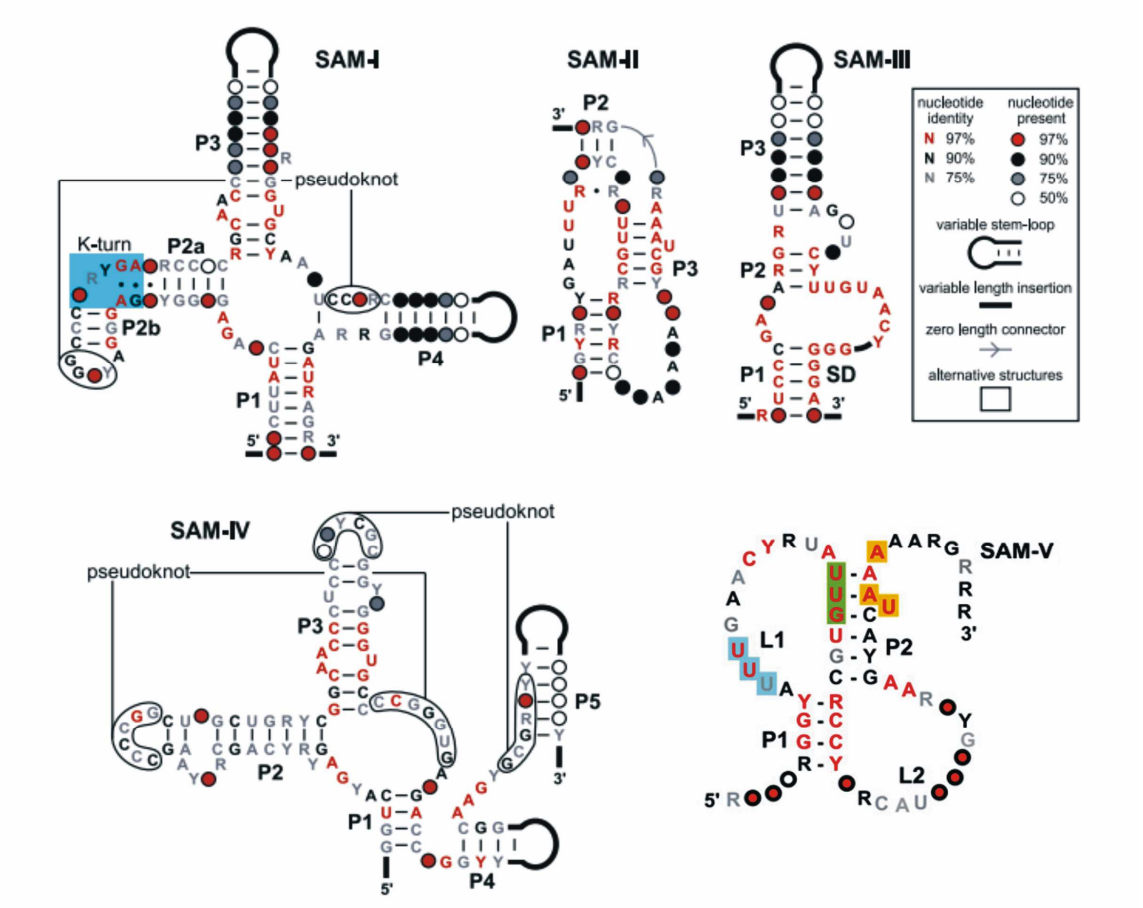


Figure 1.4: Consensus sequence and structural motifs of five different SAM-riboswitch aptamers [adapted from ref. 19, 21].

The SAM-I motif is one of the most widespread among all different SAM sensing riboswitches (Tab. 1.1)(19). The aptamer of the SAM-I riboswitch belongs to the structural class of type-II. Therefore upon binding to SAM it undergoes global as well as local changes in its conformation. This riboswitch regulates genes that are involved in sulphur metabolism, which includes biosynthesis of cysteine, methionine and SAM itself. SAM is synthesized in cell from ATP by adenosyl methionine transferase. The SAM-I riboswitch motif regulates the expression of this enzyme, thereby providing a direct feedback regulation for SAM production. The SAM-I riboswitch aptamer is able to distinguish between SAM and SAH like all other SAM-responsive riboswitches. In chapter 2 we will address the folding dynamics of

this riboswitch upon binding to its cognate metabolite SAM by fluorescence resonance energy transfer (FRET)-based methods.

Table 1.1: Phylogenetic distribution of SAM riboswitches.

	SAM-I	SAM-II	SAM-III	SAM-IV
Acidobacteria	X			
Actinobacteria	X			X
Bacteroidetes	X	X		
Chlorobi	X			
Chloroflexi	X			
Cyanobacteria	X			
Deinococcus-Thermus	X			
Fusobacteria	X			
Firmicutes				
Bacillales	X			
Clostridia	X			
Lactobacillales	X		X	
Mollicutes	X			
Proteobacteria				
α-proteobacteria	X	X		?
β-proteobacteria		X		
δ/ε-proteobacteria	X			
γ-proteobacteria	X	X		

1.2. FRET and folding studies of biomolecules

Fluorescence resonance energy transfer (FRET) was first quantitatively described by Theodor Förster in 1948. FRET between donor and acceptor fluorophore takes place, *via* dipole-dipole interaction when they are within 10 nm of each other(22). The energy transfer efficiency is given by the equation

$$E = \frac{1}{1 + (R/R_0)^6},$$

where R is the distance between the dyes and R_0 is the characteristic distance between that particular fluorophore-pair, also termed as Förster radius at which FRET efficiency, E = 0.5.

As obvious from the above equation the FRET efficiency is directly dependent on the inter-fluorophore distance (Fig. 1.5). This information can be used to investigate various

biological systems for both intramolecular interactions, such as conformational dynamics of DNA, RNA and proteins, and intermolecular interactions of proteins with DNA, RNA and other ligands. Single molecule FRET (SM-FRET) is arguably the most general and adaptable among many other fluorescence techniques for biology(22,23). SM-FRET has been successfully applied to address various questions regarding the folding dynamics of many biologically important RNAs as well as *in vitro* selected ribozymes(24-29). Although a traditional two-colour FRET is useful in most instances, this theory can be extended towards three- (or multi-) colour FRET involving three or more fluorophores. In this case it is possible to find out the inter-fluorophore distances between three dyes (therefore between three different sections of a biomolecule) during the folding event. However, finding out three fluorophores which fulfil all spectroscopic requirements might not be a straight-forward task.

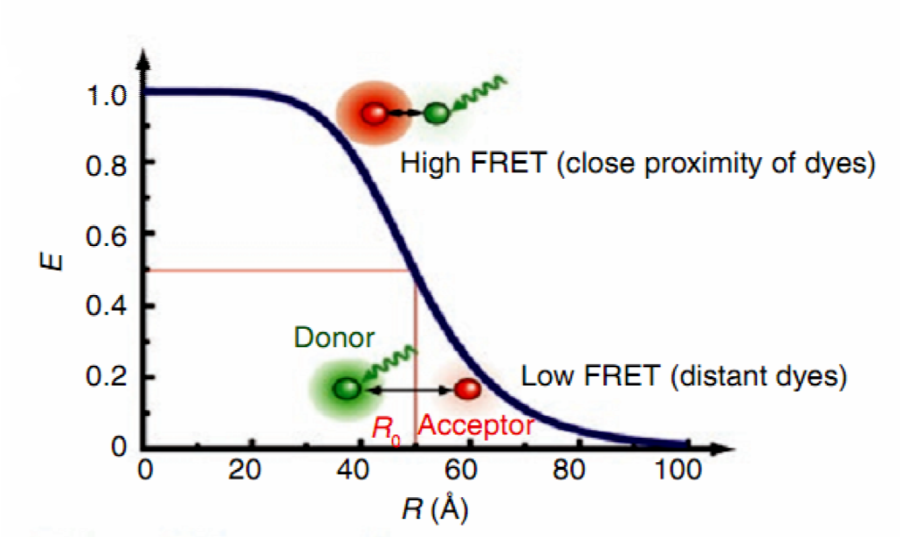


Figure 1.5: FRET efficiency, E, as a function of the inter-dye distance (R) for Förster radius, R_0 = 50 Å. Excitation of the donor dye with a laser will cause the donor either to fluoresce or to transfer energy to the acceptor dye, depending upon its proximity. The Förster radius (R_0) indicates the distance (R) between the dyes, at which E = 0.5. At a smaller distance E > 0.5 and vice versa [adapted from ref. 22].

1.3. Preparation of large site-specifically modified RNAs by splinted-ligation

It is obvious that SM-FRET will require RNA molecules that are site-specifically labelled with donor and acceptor fluorophores for studying RNA conformational dynamics. Although phosphoramidite chemistry allows the site-specific introduction of a range of different modifications into short RNA, the same task remains a major challenge for larger RNAs(30-34). Furthermore, having multiple fluorophores even in short RNAs can be troublesome during solid-phase synthesis. A straight-forward solution to this problem would be the ligation of multiple short fragments each carrying a single fluorophore – thereby generating the full-length functional RNA modified site-specifically with donor and acceptor dye**.**

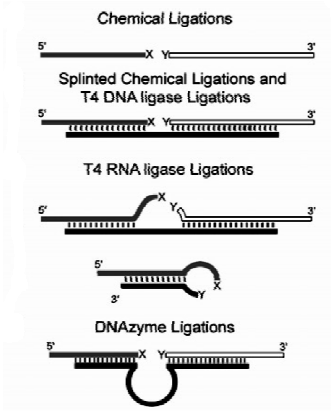
The ligation of RNAs can be achieved by chemical and enzymatic methods. The specificity during joining multiple fragments together can be achieved by bringing the reactive ends in close proximity to each other in correct order by hybridizing to a complementary splint oligonucleotide (Fig. 1.6). Among many different chemistries used for nucleic acid ligation the cyanogen bromide (BrCN) or the water soluble carbodiimide (EDC) method are of particular interest(35). Chemical ligation has also been successfully applied to prepare lariat DNA and RNA molecules(36,37). Although this sounds attractive, the reaction rate of chemical ligation often depends on the local environment of the nick (38). To improve the rate of chemical ligation, often non-natural, more reactive functional groups are introduced at the ligation joint(39).

Contrary to chemical ligation, enzymatic ligation is more efficient and arguably the method of choice for preparing long RNA by ligation. Both proteins and nucleic acid enzymes can be used to ligate RNA strands(40,41). Although various different ligases have been isolated from many different organisms, those originating from the bacteriophage T4 are the best characterized and most highly used ligases in molecular biology, in particular for RNA ligation. T4 DNA ligase (T4 Dnl) catalyzes the repair of nicks in double stranded DNA *in vivo*. However, its activity is not restricted towards joining only DNA nicks in a double stranded environment. T4 Dnl is also reported to join nicks between two RNAs when hybridized over a DNA splint, albeit with a lower efficiency(42,43). This is due to the slow release of the product after ligation. Due to the higher ligation efficiencies, milder reaction conditions and the lack of need for unnatural functional groups at the ligation site for efficient ligation, enzymatic ligation is more popular than chemical ligation and surely the method of choice in modern molecular biology.

A crucial point in ligation is the formation of a ligation competent complex (LCC) which consists of all ligation fragments correctly hybridized on a DNA splint. A previous study(42) indicated that the isolation of LCC from other intermediates (complexes resulting from incorrect annealing of the fragments on a splint) can increase the ligation efficiency to >75 % per nick. However, it is obvious that for statistical reasons, the higher the number of ligation fragments involved, the lower would be the probability for correct LCC formation. Taking into account that a higher number of ligation fragments would proportionally increase the number of nicks to be sealed, the preparation of large RNAs modified site-specifically with fluorophores from many smaller modified and unmodified fragments by ligation can be highly challenging.

The ligation of RNA fragments can also be achieved by T4 RNA ligase 1 (T4 Rnl1). In contrast to T4 Dnl, T4 Rnl1 is a single stranded ligase. *In vivo* this ligase is responsible for repairing the nicks in RNA hairpin loops, e.g. anticodon loops in tRNA. Many groups have successfully used this ligase for ligating RNA. Since T4 Rnl1 ligates only single stranded

RNAs it is not possible to achieve specificity in ligation involving multiple fragments unless a complementary DNA splint is used for which the enzymatic efficiency is extremely low due to the double stranded nature of the nick. In absence of a splint, circularization and concatenation are the most prominent reactions. Therefore a DNA splint, which forces the ligation joint to be locally single stranded has been used to achieve successful ligations (Fig. 1.6)(44). Alternatively, the ligation fragments can be partially self templating and thereby adopting a hairpin like conformation where the ligation joint is situated in the loop region (also known as Y-ligation) (Fig. 1.6)(45). Albeit in a different context, T4 Rnl1 has been used to prepare circular ribozymes with increased activity, decreased divalent metal ion requirements and improved exonuclease stability(46). In addition, T4 Rnl1 can ligate single-stranded RNA or DNA to double stranded RNA. This is particularly useful to amplify double stranded RNA(47,48).



Ligation	Chemistry	X	Y	Splint
Chemical	BrCN	3'-phosphate	5'-OH	DNA, peptide
	BrCN (Lariat)	2'-phosphate	5'-OH	DNA, RNA
	EDC (Lariat)	2'-phosphate	5'-OH	DNA, RNA
	Periodate	2',3'- diol	5'-NH_2	No
	Thiol	3'-disulfide	5'-SH	No
	Photoligation	3'-pyrimidine	5'-^{CV}U	DNA
	Aldehyde	2'-NH_2	5'-aldehyde	No
	Click	3'-O-propargyl	5'-azide	No, RNA
		3'-azide	5'-phosphohexenyl	No, RNA
T4 DNA ligase	Phosphate	3'-OH	5'-phosphate	DNA
T4 RNA ligase	Phosphate	3'-OH	5'-phosphate	DNA, RNA
DNAzyme	Phosphate	2',3'-cyclic Phosphate	5'-OH	DNA
		2',3-diol	5'-triphosphate	

Figure 1.6: Various RNA ligation schemes with reactive termini X and Y (left). Examples of different chemical and enzymatic RNA ligations (right) [adapted from ref. 40].

Yet another RNA ligase isolated from the phage T4 is T4 RNA ligase 2 (T4 Rnl2). Unlike T4 Rnl1, T4 Rnl2 is a double stranded ligase and its RNA nick joining activity has been found to be a lot higher for double stranded substrates compared to T4 Rnl1 or T4 Dnl(43). In our experiments for preparing full-length dual fluorophore labelled riboswitches, we found this enzyme to be particularly useful (Chapter 2).

1.4. Click chemistry – application towards nucleic acid functionalization

1.4.1. Present challenges in site-specific nucleic acid functionalization

Apart from the FRET-based biophysical studies described before, many other areas of research also make extensive use of site-specifically modified nucleic acids(2,49-52). Most of the chemical modifications have so far been introduced during solid-phase synthesis of nucleic acids by the phosphoramidite approach. Although it is possible to introduce a large number of modifications into short oligos(31,32) by this approach, the same task remains a major challenge for longer nucleic acids. This is because of the degenerative nature of the overall process which involves a large number of individual coupling, protection and deprotection steps, therefore the longer the nucleic acid, the higher the number of steps involved in synthesis, and the poorer the overall yield of the final product. The reason lies rather at the core of organic synthesis where most of the reaction trajectories cannot be fully controlled since the majority of chemical functionalities show a promiscuous reactivity towards a wide range of other functional groups, and the chemical reactivity for one particular functional group towards all others varies rather gradually than in discrete steps. This therefore leads to an uncontrolled reaction trajectory leading to not only many products but also insufficient yield for the desired compound.

To circumvent these problems, synthetic chemists developed two different strategies – i) the use of protecting groups(53) which allows selective functionalization of a particular chemical entity in a pool of other functional groups and ii) the use of domino reactions(54). The former strategy is more popular among synthetic chemists since it offers a simpler reaction design and also gives the opportunity to deal with a wide range of functional groups. However this strategy exponentially increases the labor involved in synthesis due to the higher number of reaction steps and multiple intermediate purifications involved and ultimately leads to poor overall yield of the desired product. The latter strategy relies on sequential reactions, all happening in one reaction vessel where the product from the first reaction is used as the starting material for the successive reaction. But this strategy neither allows a straight-forward design of reaction cascades, nor manipulation of a diverse range of functional groups. However, both of these approaches suffer from the same degenerative nature as mentioned earlier, therefore, the higher the number of steps (domino or discreet) involved in a synthesis, the lower the overall yield of the product.

A plausible solution to this problem would therefore involve the *de novo* chemical synthesis of only a small oligonucleotide carrying the required modification followed by the ligation of that fragment to a larger enzymatically prepared non-modified oligonucleotide to finally yield the full-length product carrying the site-specific modification (Section 1.3).

In contrary, Nature has evolved a wide range of highly specialized enzymes capable of modifying nucleic acids site-specifically(55). However, most of these modifications, although immensely important for biological function, do not have the required chemical functionalities to facilitate our understanding of biological mechanism at molecular detail. Therefore it would be extremely useful to manipulate these enzymes or their substrates in a way that will allow us to choose the chemical entity of our interest at the advent of nucleic acid modification.

However there are two different issues to be addressed during site-specific enzymatic introduction of chemical modifications to nucleic acids – i) the modified substrate has to be accepted by the enzyme preferably with a comparable efficiency with respect to its natural substrate to ensure reasonable modification yield and ii) the inserted modification should cause only minimal changes to the nucleic acid molecule so that its function is preserved. While the second task can be achieved rather easily when bio-chemical and -physical data are available for that particular nucleic acid, so that one can easily find out certain regions in the sequence where modifications can be inserted without hampering its function, the first task poses a serious problem. Therefore, to expand the substrate tolerance of various enzymes often either the enzyme is mutated(56,57) or enzymatic reaction conditions (buffers, presence of various metal ions etc.) are severely manipulated(58-62). However, none of these approaches are quite desirable since the first involves laborious biochemical techniques required to produce these mutated enzymes which is applicable though for only one subset of modifications, thereby limiting their use as a universal enzyme capable of introducing every kind of modification one wishes to have, and in the second approach severe manipulations of an enzyme´s preferred reaction condition often lead to decreased, unpredictable or promiscuous enzymatic activity.

One solution to the above mentioned problem is a two-step approach, where, at the first step only a small, but still chemically modifiable functional group causing only minimal structural changes with respect to the natural substrate is introduced and in the second step this functional group is converted to the desired modification *via* a selective chemical reaction. This two-step approach neither demands production of new mutated enzymes nor alteration of preferred reaction conditions, thereby both decreasing the labor involved in enzyme mutation and increasing the enzymatic reaction efficiency. Several different biocompatible reactions are developed for such purposes(63); notable examples include NHS esters, acyl azides and isothiocyanates for labeling with amines; maleimides, aziridines and halo-acetyl derivatives for labeling with thiols; hydrazine derivatives and Schiff-Base chemistry for labeling with aldehydes and ketones. However, this approach can potentially be degenerative since it requires multiple steps (in fact many of these reactions often do not lead to quantitative reaction yield); additionally none of the above mentioned conjugation

reactions are completely orthogonal to all other functional groups present in the biomolecule, therefore leading to many side products and poor labeling efficiency.

1.4.2. Click chemistry as a reliable tool in biomolecule labeling – bioorthogonal click reactions

These problems of chemical promiscuity can be surmounted by the use of ``Click Chemistry`` as mentioned by Sharpless and coworkers in 2001(64). Originally ``click reactions`` are defined as a ``*set of powerful, highly reliable, and selective reactions for the rapid synthesis of useful new compounds and combinatorial libraries*`` (64). Click chemistry is not limited to a specific type of reaction, but stands for a synthetic philosophy that comprises a range of reactions, with different reaction mechanisms leading to different products but with common reaction trajectories. Click reactions are driven by a high thermodynamic driving force (>20 kcal mol^{-1}) therefore ensuring quantitative reaction yields in most cases, which might improve the inherent degeneracy of a multi-step chemical synthesis. A sub-group of click reactions, known as bio-orthogonal click reactions, has to meet an even more rigorous set of requirements, namely high reactivity, orthogonality (therefore selectivity) to other functional groups present in a biological system, compatibility with water and other protic solvents, quantitative reaction yields leading to stable and non-toxic products. Moreover, these reactions should preferably involve non- (less) - toxic reagents and small reactive groups, thereby causing only minimal structural changes to the modified substrate compared to its natural analog and therefore increasing the likelihood of these modified substrates being accepted by the enzyme. To achieve most of these criteria, the majority of click reactions involve reactive groups which are absolutely abiotic. Once all (or majority) of these requirements are fulfilled, bio-orthogonal click reactions can be used as a universal ligation/conjugation strategy in molecular biology.

These bio-orthogonal click reactions can further be sub-divided into two categories – i) reactions involving functional groups which are completely stable towards each other and can react only in presence of a certain external chemical or photochemical trigger, ``Triggered Click Reaction`` and ii) the functional groups involved have high mutual reactivity and do not require any further external trigger for efficient conjugations, ``Non-triggered Click Reaction``. Both of these two classes of reactions have their advantages and disadvantages, and therefore the right choice of a click reaction depends on the specific applications. Non-triggered click reactions can be converted to their triggered versions by chemical manipulations(65) and multiple levels of triggers can be implemented in an existing triggered click reaction(66).

1.4.3. Copper catalyzed azide alkyne cycloaddition (or stepwise ligation) (CuAAC) for labeling oligonucleotides

The 1,3-dipolar cycloaddition reaction between terminal alkynes and azides without the need of any transition metals to form regioisomeric triazoles, as described by Huisgen *et al.* (67), is a concerted π^4s-π^2s cycloaddition. This reaction has a very slow reaction rate and needs elevated temperature and prolonged reaction time. Formation of the regioisomeric products is indeed in accordance to a concerted cycloaddition since in case of alkyne and azide as dipolarophile and dipole, energetically the reaction can be both dipole-HOMO or dipole-LUMO controlled(68). The rate of such cycloadditions is determined by the energy to distort the reactants to the transition state geometry rather than FMO interactions or reaction thermodynamics. However, the same reaction, when performed in presence of Cu(I) reaches a tremendous rate acceleration leading to a single 1,4-disubstituted triazole. This copper catalyzed version does not need elevated temperature or prolonged reaction time and is tolerant towards a wide range of pH and solvents including water (69,70).

Mechanistic studies as well as theoretical calculations favor a stepwise reaction (therefore ligation) rather than a concerted cycloaddition(70). This reaction has been optimized at different levels by Sharpless and coworkers(64,71-78) and in the past decade CuAAC has become the most widely used among click reactions for synthesizing combinatorial libraries of drug-like small molecules(71,79) as well as in diverse biotechnological applications involving protein, nucleic acid and carbohydrate molecules(80-84). CuAAC offers a range of advantages, namely i) the triazole product is very stable, ii) the molecular dimension of 1,4-disubstituted triazoles is somewhat similar to amide bonds in peptides in terms of distance and planarity(85), iii) the reactive alkyne and azide functional groups are completely abiotic, therefore minimizing the probability of unwanted side reactions with other functional groups present in the biomolecule, iv) in spite of their high mutual reactivity in presence of Cu(I), alkyne and azide are very stable towards a wide range of functional groups and reaction conditions, v) both alkyne and azide functional groups are very small, therefore substrates carrying any one of these two modifications have only minute structural differences compared to their natural analogs which in turn increases their likelihood of being accepted as substrates by various enzymes.

However CuAAC also has certain limitations – i) it involves copper which is toxic towards cells since copper can generate radicals in aqueous media in presence of oxygen which can lead to nucleic acid backbone scission. In fact, a copper-phenanthroline complex had been used as artificial nuclease for footprinting DNA-protein interactions(86,87). ii) Since Cu(I) salts are poorly soluble in water the majority of the CuAAC protocols involve *in situ* generation of Cu(I) by reducing Cu(II) with sodium ascorbate or ascorbic acid. However,

ascorbate derivatives are toxic as well, since they are capable of initiating Fenton´s chemistry in presence of oxygen and transition metals, thereby leading to nucleic acid scission(88-90). Moreover Cu(II) is capable of being strongly coordinated by histidine which in fact has been used for affinity precipitation as an initial step in certain protein purifications(91). Additionally, Cu(II) is capable of oxidizing imidazole to 2-imidazolone(92) which can cause protein-protein histidine-lysine cross-link, since the 2-imidazolone moiety from oxidized histidine is prone to nucleophilic attack by the ε-amino of lysine(93).

For reducing Cu(II) to Cu(I) other phosphine- or disulfide-based reducing agents have been used in few cases, however these reducing agents are not quite desirable since phosphine-based reagents can also reduce the reactive azide (similar to Staudinger reaction)(94,95) and disulfide-based reagents lead to the generation of thiols which are good copper-coordinating ligands therefore the Cu(I) formed *in situ* would not be available for catalysis.

To circumvent many of these problems, various copper stabilizing ligands have been introduced(74,96-99), therefore converting this ``ligand-free`` reaction(100) to a ``ligand-accelerated metal-catalyzed`` reaction(101). These ligands are meant to stabilize Cu(II) and Cu(I), as well as to keep the harmful ascorbate oxidation products low, thereby providing more biomolecule friendly reaction conditions(102). However Cu(I) should not be too well stabilized, otherwise it would not be available for coordination with alkyne and azide and therefore catalysis (Chapter 3).

With these advancements in CuAAC, this reaction has been successfully applied in various studies involving DNA(83). A range of different pyrimidine(103) phosphoramidites bearing alkynes were synthesized and incorporated in solid-phase DNA synthesis, thereby enabling site-specific modification of short DNA. Also, various triphosphates bearing either alkyne or azide were synthesized and incorporated in primer extension studies, thereby enabling random modification of long DNA(104,105). Moreover, these statistically modified DNAs can be used to prepare uniformly metallized DNA(106), gold nanoparticle DNA conjugates(107) and to immobilize DNA on a glass-surface in a pre-defined micropattern(108).

1.4.4. Copper-free click reactions

Despite of all aforementioned successes with CuAAC, the cytotoxicity of copper has somewhat restricted the use of this reaction to *in vitro* applications. However, there is another group of bioorthogonal reactions that are cycloadditions lacking exogenous metal catalysis - the so-called copper-free click reactions. Exogenous metals can have severe cytotoxic effects and can thus disturb the delicate metabolic balance of the systems being studied(109). The azide functional group is of notable interest in this regard since it can also participate in many other metal-free click reactions, e.g., Bertozzi-Staudinger ligation(110) and strain promoted azide alkyne cycloadditions (SPAAC)(111). However, the relatively slow kinetics of the two abovementioned click reactions limited their use to probe fast biological processes.

The rate of a cycloaddition can be improved either by modulating the reactant orbital energies or by imposing ring-strain to the reactants. Ring-strain is indeed the reason for an efficient reaction between cyclooctynes and azides compared to traditional Huisgen cycloaddition between terminal alkynes and azides. To further improve the reactivity of the cyclooctyne, electron withdrawing groups (e.g., fluorine) have been introduced into the cyclooctyne moiety, thereby lowering the LUMO energies(112,113). And to improve the water-solubility of these cyclooctynes, various heteroatom substituted cyclooctynes have also been synthesized(114).

Similarly, the reactivity of cyclooctynes can also be improved by imposing further ring strain. For this purpose various (di)benzocyclooctynes have been synthesized and applied for biomolecule labelling(115-117).

A relatively new reaction in the field of metal-free click reactions is based on the inverse electron demand [4+2] cycloaddition between 1,2,4,5-tetrazines and strained dienophiles, such as norbornene, cyclooctyne and *trans*-cyclooctene(118-124). Generally, these cycloadditions are faster than the previously described strain promoted cycloadditions. Moreover, these reactions work efficiently in aqueous solutions and in serum. Although the initial studies for these reactions were performed in organic solvents like dioxane, the rates of these reactions were reported to be enhanced by the addition of water(125). A driving force for this reaction is the liberation of nitrogen from the tetrazine residue during the reaction. Additionally, the tetrazine reaction with dienophiles can be followed spectroscopically by the disappearance of an absorption band between 510 and 550 nm(126). This UV-property of tetrazines has been exploited to design fluorogenic probes(120). It is noteworthy here that fluorogenic probes are far more attractive choices for intracellular labelling due to their low background signal. Due to its unprecedented reaction rate, yet high selectivity in a biological environment, its metal-free nature and fluorogenic properties, this reaction has been widely

used in various applications(127-131). We used this reaction in combination with CuAAC for simultaneous double labelling of DNA (Chapter 4).

1.5. Experimental RNomics

Riboswitches, as described before, a solely RNA-based metabolite sensing gene regulatory system, currently account for about 3-4 % of the total gene regulation for certain prokaryotes. Apart from their aforesaid abundance, the scope of these regulatory mechanisms is also of equal interest(132,133). Although there is much effort in elucidating the structures and mechanisms by which riboswitches work, only little work has been performed to develop experimental strategies to isolate these riboswitches. All known riboswitches have been discovered by rational approaches, followed by experimental verification of candidates, and not by a random experimental screening strategy. These discoveries were made by either investigating mRNAs in pathways where biochemical evidence showed regulation, but no protein regulator had been found, or by genetic screens. Biocomputational approaches were mainly applied to find known aptamer motifs in genomes, to find known riboswitch types in different organisms, or to detect conserved structural elements in the untranslated regions (UTRs) of mRNAs(134-136). Given the fact that the *de novo* prediction of RNA three-dimensional structure from sequence is quite inaccurate in the absence of reference structures, it can be assumed that many more riboswitches exist that have not been discovered so far, and that may not be identified at all by the aforementioned approaches.

The investigation of metabolite-sensing RNAs in prokaryotes might enable us to design new antibiotics because such RNA systems appear to be absent in humans. One example of an antibiotic acting on riboswitches is PTPP (pyrithiamine pyrophosphate), a TPP analogue that functions by blocking the binding site of the TPP-sensing riboswitch, thereby shutting down thiamine metabolism in bacteria(13,137,138).

As mentioned earlier, all riboswitches contain a simple modular architecture – a metabolite sensing aptamer domain and a gene regulatory expression platform. Generally, the aptamer domain binds its cognate metabolite with a very high affinity, often multiple orders of magnitude higher than their *in vitro* selected counterparts. Therefore, a straight-forward approach would involve an affinity purification of cellular RNA extracts, exploiting the strong binding of the metabolite by its cognate aptamer. To this end, genomic SELEX could be an attractive strategy.

1.5.1. Genomic SELEX

Genomic systematic evolution of ligands by exponential enrichment (Genomic SELEX) is an experimental tool for expression of condition-independent genomic RNA aptamers. This approach has been successfully applied for the isolation of protein-binding RNA (139-142). The principle of genomic SELEX has been elaborated in figure 1.7. RNA libraries derived from genomic DNA are generated *via* random priming, PCR amplification and *in vitro* transcription. The DNA library consists of genomic sequences of selected size flanked by constant primer binding sequences required for amplification and transcription. The transcribed RNA pool is subjected to several rounds of selection and amplification to enrich for RNA sequences with desired small molecule binding activity. Since in this approach the RNA sequences are transcribed *in vitro* from genomic DNA, the expression of such small molecule binding RNAs is independent from *in vivo* conditions.

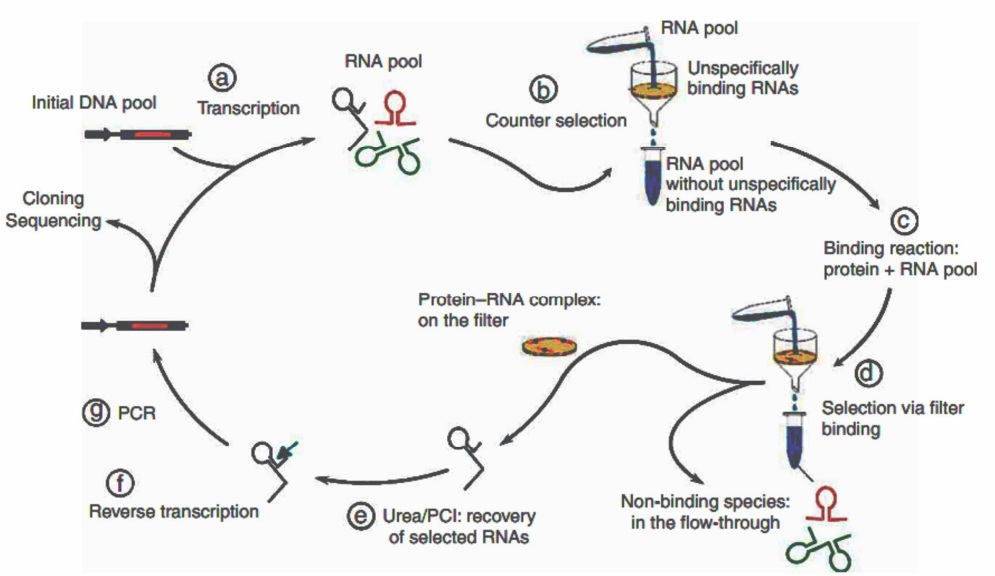


Figure 1.7: Overview of the genomic SELEX process for isolating protein-binding genomic RNA aptamers. a) Transcription from the genomic DNA pool; b) counter selection to remove RNAs with unspecific binding to the membrane; c) incubation with bait-protein; d) selection of sequences *via* filter binding; e) selected RNAs are recovered by denaturing the bait-protein; f) reverse transcription and g) PCR followed by either characterization of the selected sequences or continuation to the next round [adapted from ref. 129]

However, there are two major obstacles in using genomic SELEX for riboswitch discovery. i) Although for riboswitches the metabolite-aptamer binding has been reported to be very strong (often with a low nanomolar K_d), these measurements are exclusively performed *in vitro* and under thermodynamic equilibrium conditions. Therefore *in vitro* SELEX binding conditions designed from these data might not reflect the actual scenario *in vivo.* ii) Genomic SELEX requires immobilization of the bait (in this case the metabolite) on a solid-support, whereas in almost every example of riboswitch-metabolite recognition the

metabolite has been found to be completely encapsulated by the riboswitch. Immobilization of the metabolite to a solid-support even through a long linker might severely hamper the metabolite binding of the riboswitch.

The aforesaid disadvantages prevented us from taking a genomic SELEX approach for fishing out riboswitches. Therefore we have adapted to a photoaffinity tagging approach as described below.

1.5.2. Affinity tagging, photocrosslinking and affinity-based chemical RNomics

Affinity tagging is one of the major workhorses of preparative biochemistry and is routinely used in every laboratory to facilitate the isolation of over-expressed proteins. Selected examples are the His-tag and the Strep-tag(143). In nucleic acid biochemistry, similar tags have been developed(144), but these have found only limited use, since methods based on hybridizing complementary oligonucleotides are simple and reliable.

Photocrosslinking is a powerful methodology for probing structural features of biological systems in aqueous solution(145,146). Several studies report the use of such techniques for mapping RNA structural neighbors in three-dimensional space, and for identifying important tertiary interactions. In combination with an affinity tag, photocrosslinking was also used to study mRNA-protein interactions(147-150).

One of the more commonly used methods in protein enzymology is photoaffinity crosslinking, in which photoreactive compounds with high affinity to the active site or a binding pocket are used to explore the direct vicinity of the catalytic center. In RNA enzymology, however, photoaffinity crosslinking has so far only been used for RNA-cleaving ribozymes, applying a photo-reactive RNA substrate strand(151).

Activity(affinity)-based protein profiling has been developed rather recently and is basically the protein equivalent of what we hope to establish here with RNA: Metabolites are derivatized with a (photo-)reactive group and an affinity tag, allowed to bind to their cellular targets (in this case riboswitches), purified, and then identified by sequencing. In chapter 5 we will describe our convergent synthetic approach for synthesizing a library of different photoaffinity probes followed by affinity tagging by UV-irradiation. After this step, the target RNA sequences are affinity purified from other unrelated sequences, and then subjected to a ligation protocol to attach constant sequences at each end required for amplification. The amplified RNA sequences are to be tested *in vitro* for their biological properties.

1.6. Goals and current work

In this work we first sought to investigate the folding dynamics of riboswitches by bulk and SM-FRET (section 1.2 and chapter 2). This requires the synthesis of large site-specifically functionalized structured RNAs for which we have established ligation strategies (section 1.3 and chapter 2). At this point we realized the limitations of existing nucleic acid functionalization strategies (section 1.4.1). Therefore we decided to use bioorthogonal click reactions as part of our functionalization strategy. We first had to find the best click reaction for our purpose among various different click reactions and to optimize its conditions (section 1.4 and chapter 3). Having the optimized click reaction conditions at hand, we developed chemo-enzymatic strategies to site-specifically functionalize long RNAs (chapter 4). In addition to this, we developed a chemical approach using two different mutually orthogonal click reactions for concurrent, site-specific labelling of DNA molecules with multiple fluorophores which might be suitable for FRET-based biophysical studies involving DNA in the future (chapter 4). In summary, we developed enzymatic strategies for site-specific transfer of clickable residues to long RNAs. We extended this strategy of enzymatic, site-specific transfer of clickable residues to long RNAs towards photochemical transfer of clickable moieties to a target RNA in a mixture of many unrelated sequences (section 1.5.2 and chapter 5). This technique, which we call ``Affinity-based Chemical RNomics`` is a chemical approach in experimental RNomics whereby riboswitch sequences for a given metabolite are to be isolated from a total RNA isolate of any organism just by the virtue of its tight binding to its cognate metabolite and without any prior knowledge of its sequence.

2. Conformational dynamics of riboswitches analyzed by FRET

2.1. Riboswitch mediated gene regulation strictly requires the expression platform

The dynamics of RNA structure are of widespread interest in all domains of life science(1,2). As mentioned previously (Section 1.1) the simple modular architecture of riboswitches made them ideal candidates for folding study.

In vitro investigations show, that upon substrate binding to the metabolite-sensing domain, a structural reorganization of the RNA occurs that unveils (or sometimes masks) the gene-expression signal. There is, however an ongoing discussion as to whether the actual mechanism *in vivo* relies on such a rearrangement, or if the metabolite directs the folding of the *de novo* synthesized transcript into one or the other conformation(15).

From the above, two questions can be formulated, which have been the central interest ever since riboswitches were first discovered(137,152), namely: how is the metabolite recognized, and how does binding of the substrate trigger gene regulation. Although the answer of the first question can be obtained from various high-resolution X-ray structures(138,153-161), the answer to the second question is still missing since majority of the aptamer crystal structures are resolved without the expression platform and in most cases with their cognate metabolite. Therefore very little is known for the ligand-unbound state of a riboswitch and especially the dynamics of metabolite-induced rearrangements which is responsible for the gene regulation *in vivo*. Ideally the determination of rate-constants of certain folding events might finally give us the true picture of riboswitch mediated gene regulation. Therefore we decided to study the folding dynamics of a complete riboswitch (aptamer + expression platform) upon binding to its cognate metabolite and not only with the aptamer, as has been pursued in many previous studies(162-171).

2.2. The SAM-I riboswitch

The SAM riboswitches are the most common among all riboswitches(7,19). Additionally, the widespread distribution of the SAM-I motif (Section 1.1.1) among all different SAM sensing riboswitches arguably makes this RNA an ideal candidate for further folding study. Since our laboratory has successfully applied SM-FRET for studying RNA folding dynamics(24-29), we decided to take a similar approach in this case. Additionally, FRET-based methods are most efficient for studying global rather than local conformational changes and SAM-I aptamer motif belongs to the structural class of type-II (Section 1.1) which undergoes a global change

in its architecture upon metabolite binding. These two points together led us to assume that the SAM-I riboswitch is indeed a very good candidate for such analysis.

The SAM-I riboswitch regulates various genes that are involved in sulfur metabolism and is able to distinguish between SAM and its closely related analog SAH (Section 1.1.1). The SAM-I riboswitch from *Bacillus subtilis* regulates gene expression by pre-mature transcription termination(172-174). The regulating element (i.e. the expression platform) is a stem-loop structure followed by a stretch of five to nine uridine residues and is stabilized upon ligand binding(175,176).

The aptamer depicts a single four-way junction, in which two pairs of coaxially stacked helices (P1-P4 and P2-P3) pack closely against each another (Fig. 2.1).

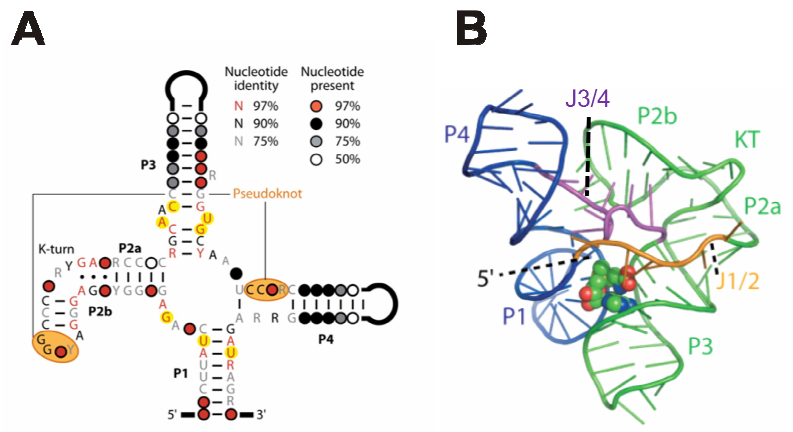


Figure 2.1: SAM-I riboswitch aptamer. A: Consensus sequence and secondary structure model of the SAM-I riboswitch. The nucleotides highlighted in yellow were observed to contact SAM ligand directly. Nucleotides that participate in pseudoknot formation are shaded in orange. Solid round lines at the ends of P3 and P4 stems indicate variable stem loops**. B:** Atomic resolution structure for the SAM-I riboswitch aptamer of *Thermoanaerobacter tengcongensis*. The structure models are shown as ribbon diagrams and the ligand is present as spheres. Pairing elements and joining regions are indicated. Legend: *KT = kink-turn motif* [adapted from ref. 6 and 7].

In the crystal structure of the SAM-I riboswitch it was observed that its global architecture is achieved *via* tertiary interactions between L2, J3/4 and J1/4(177). A conserved kink-turn motif (also called the GA-motif)(178) within the P2 stem enables the corresponding L2 to dock with J3/4 via a pseudoknot interaction thereby stabilizing the global fold(166,179). SAM is embedded in the grooves of P1 and P3 with additional contacts contributed by junction residues (Fig. 2.1 B).

Earlier it was proposed that the P1 stem is only formed in the presence of SAM (Fig. 2.2)(166). In absence of the ligand, the 3´-end of P1 stem forms an anti-terminator resulting in transcription of the downstream mRNA(179). However, the crystallographic structure is known only for the aptamer in its ligand-bound state(179). The complete riboswitch as well as the aptamer in their ligand unbound state are likely to be too flexible for crystallographic structure determination. Therefore, analyzing ligand induced folding dynamics of the SAM-I

riboswitch might be a challenging task, because of the unknown structure of the ON state of this riboswitch (Fig. 2.2).

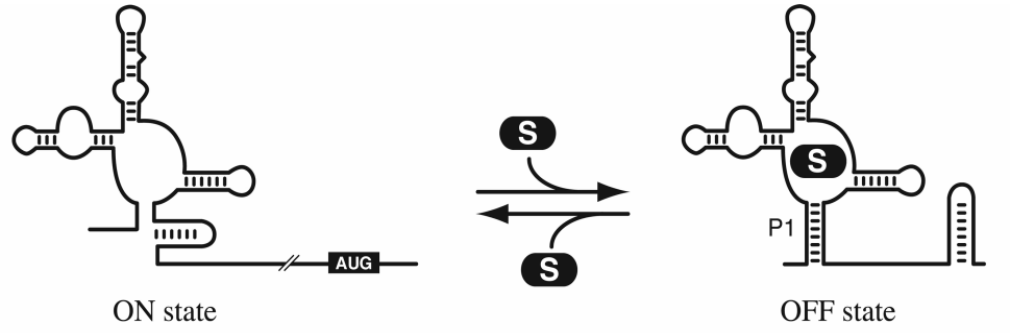


Figure 2.2: Schematic representation of the ligand-induced SAM-I riboswitch premature transcription termination. The structure in the ON state (left) and the structure in the ligand bound or OFF state (right) are depicted [adapted from ref. 166].

2.3. Choice of suitable dye positions - combinatorial construct design

We have chosen the C-5 position of uridine for attachment of the respective fluorophores *via* a C_6 linker. On the contrary to some of our previous studies(29), the uridine residues containing the dye labels bear a 2´-OH to ensure minimal structural changes compared to the wild-type riboswitch. The nucleotide positions in the riboswitch have been chosen after careful analysis of the previously published X-ray structure(179) and biochemical data(175,176) on this RNA motif. However the crystal structure was published for SAM-I motif from *Thermoanaerobacter tengcongensis* and biochemical experiments were performed for the same motif but from *Bacillus subtilis*. Since these two sequences slightly differ from each other, we have created a translation map between these two sequences for estimating the inter-fluorophore distances in our designed constructs which is based mainly on the sequence from *Bacillus subtilis* (Fig. 2.3).

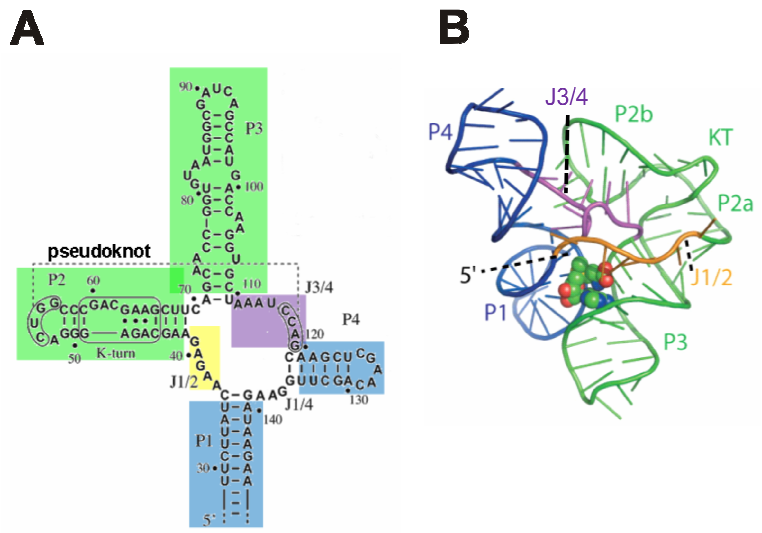


Figure 2.3: Translation map from B) the X-ray structure [sequence from *Thermoanaerobacter tengcongensis*] to A) the secondary structure [sequence from *Bacillus subtilis*] of the SAM-I riboswitch. Same color code depicts similar regions in the structure.

We have taken Cy3 and Cy5 as a FRET pair for which the Förster radius was reported to be 56 Å (180). The dye positions are chosen in a manner that the estimated inter-fluorophore distances are close to the Förster radius of the dyes (Fig. 1.5). A combinatorial map of all six different constructs and estimated inter-dye distances were shown in figure 2.4. Since the crystal structure was resolved only for the aptamer, it was not possible to estimate the inter-fluorophore distances in case of constructs carrying dyes in the expression platform (Fig. 2.4). As can be seen from the X-ray structure, the aptamer in its ligand bound state consists of two pairs of coaxially stacked helices (P1-P4 and P2-P3)(179). To our rationale, the stacking of P1-P4 and P2-P3 can happen even in the absence of SAM and ligand binding induces these two pairs of coaxially stacked helices to pack closely against each other around a single four-way junction. Therefore placing the FRET-pair in such preformed coaxially stacked helices (i.e. either on P1-P4 or on P2-P3) might not be beneficial since we wanted to monitor ligand induced conformational changes by monitoring the changes in FRET (which strictly requires distance changes between the donor and acceptor fluorophores) while ligand binding should not alter the pre-formation of such a pair of coaxially stacked helices.

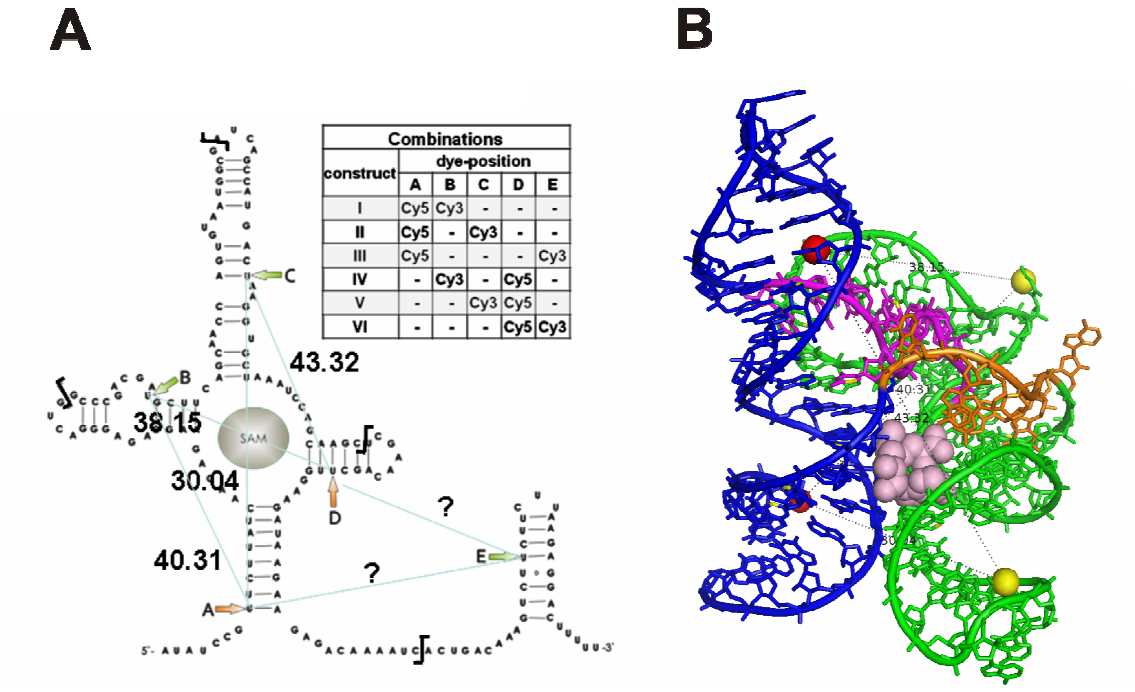


Figure 2.4: A: Combinatorial map of all SAM-I riboswitch constructs from *Bacillus subtilis*. B: Crystal structure of the SAM-I riboswitch aptamer of *Thermoanaerobacter tengcongensis*. Distances are depicted in Å. SAM is represented as a bluster of grey spheres and the Cy5 dye as red and the Cy3 dye as yellow sphere **(B)**.The ligation joints are indicated with black lines in the combinatorial map **(A)**.

2.4. Preparation of dye-labeled riboswitches by splinted ligation

For assembling various dye-labeled SAM-I riboswitch constructs a splinted ligation scheme consisting of five different ligation fragments and one splint covering the complete riboswitch (Fig. 2.4) was used.

To obtain sufficient amount of ligated products for all downstream biophysical experiments the splinted ligation scheme was optimized at several levels. The RNA ligase reaction mechanism has been briefly shown in figure 2.5.

Figure 2.5: Pathway of nucleotidyl transfer to polynucleotide 5´-ends catalyzed by RNA ligase. In the first step the enzyme (E) attacks the alpha-phosphate (p) of ATP (pppA) to form a covalent Enzyme-AMP intermediate (EpA) thereby releasing a pyrophosphate moiety (PPi). In the second step EpA transfers the AMP moiety to a 5´-phosphate of the donor RNA forming the AppRNA intermediate. Finally the 3´OH of the acceptor RNA attacks the alpha-phosphate of AppRNA to form the RNApRNA ligation product thereby releasing an AMP moiety.

The ligation reaction was optimized at five different levels, as summarized in table 2.1. At the first optimization step T4 RNA ligase 2 (T4 Rnl2) was used instead of T4 DNA ligase (T4 Dnl). Although T4 Dnl was used previously in our lab for ligating highly structured small RNAs over a DNA splint (42), its nick-joining activity was found to be far less compared to T4 Rnl2 (43). As a second optimization, the concentration of ATP in the ligation reaction mixture was varied together with an optimization for T4 Rnl2 incubation time. At the third optimization stage the DNase I digestion period was adjusted followed by the variation of T4 Rnl2 concentration in the ligation reaction mixture as fourth level of optimization. Finally, an increase in ligation yield could be achieved by the omission of phenol-ether extraction after enzymatic phosphorylation of the dye-labeled donor RNAs.

Table 2.1: Optimizations for 5-way splinted ligation of SAM-riboswitch constructs. Changes in reaction conditions are highlighted in yellow. Each optimization step (except for Opt. 1) involved at least one change in the depicted reaction conditions.[1]

Optimization step	DNase I digestion time	RNA ligase 2 concentration	Incubation time	ATP concentration	Extraction*
Opt. 1 (Fig. 2.6)	30 min	3.6 µM	0.5 h	1.2 - 1.7 mM	+
Opt. 2 (Fig. 2.7)	30 min	3.6 µM	0.5 h 4 h	50 µM	+
Opt. 3 (Fig. 2.8)	15 min	3.6 µM	1 h	50 µM	+
Opt. 4 (Fig. 2.9)	15 min	0.5 µM	1 h	50 µM	+
Opt. 5 (Fig. 2.10)	15 min	0.5 µM	1 h	50 µM	-

*phenol-ether extraction after enzymatic phosphorylation of the dye-labeled RNAs

It was observed that T4 Rnl2 has a 1000-fold higher rate of ligation [mol nicks joined/ mol protein/ min] for ligating RNA fragments over a DNA splint at 37 °C compared to T4 DNA ligase (43). Thus, we exchanged T4 Rnl2 against T4 Dnl in our ligation protocol (Fig. 2.6). In the first ligation reaction the yield for the full-length riboswitch was 1 % and that for the 4-way ligated product was 16 % (Fig. 2.6). Indeed, these yields are substantially higher compared to previous experiments with T4 DNA ligase.

[1] This part of the work has been performed together with Marika Ziesack

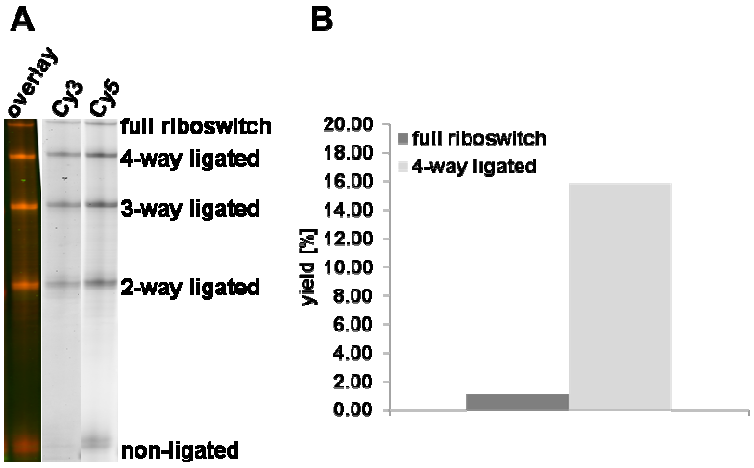


Figure 2.6: Five-way splinted ligation of dye-labeled SAM-I riboswitch using T4 RNA ligase 2. The reaction conditions are depicted in table 2.1, Opt. 1. **A:** PAGE analysis of the crude ligation mixture after digesting the DNA splint. The slowest moving band corresponds to the full riboswitch whereas all other faster moving bands correspond to incompletely ligated intermediate products as indicated. **B:** Yields of the full riboswitch (dark grey) as well as the 4-way ligated product (light grey) were calculated by using the Cy5 scan in **A**.

To further improve the yield of ligation product the ATP concentration was varied. Previous biochemical studies (181) involving T4 Rnl2 revealed a negative effect of ATP on the T4 Rnl2 ligation efficiency. During the ligation reaction, T4 Rnl2 catalyzes the formation of adenylated[2] RNA (AppRNA) as an intermediate product (Fig. 2.5). T4 Rnl2 is prone to dissociate from the AppRNA. An immediate binding of ATP to the enzyme precludes the ligase from rebinding the AppRNA to perform the subsequent nick joining. Consequently, at higher ATP concentration the AppRNA accumulates, which in turn leads to a decrease in the ligation yield. In order to avoid this phenomenon, the ATP concentration in the ligation reaction mixture was reduced from 1.2 – 1.7 mM to 50 µM (Fig. 2.7). Moreover previous studies (182) revealed that 15 min after the commencement of the ligation reaction two third of the available RNA ends were sealed by T4 Rnl2, which suggests a high catalytic efficiency. Therefore to test the time dependence of T4 Rnl2 ligation, the reaction mixture was incubated either for 0.5 or 4 h (Fig. 2.7). The yield for the full-length riboswitch after 0.5 or 4 h of incubation was 5 % in both cases and that of 4-way ligated product was 12 % (Fig. 2.7).

[2] ``Adenylylated`` has been abbreviated throughout this document as ``adenylated`` for easier reading.

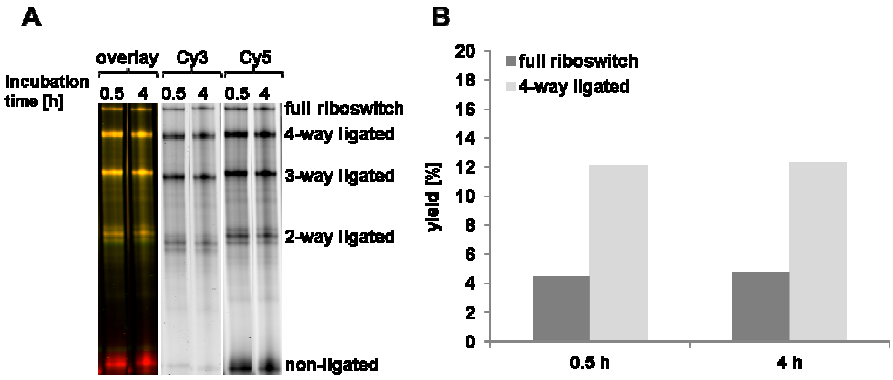


Figure 2.7: Varying ATP concentration and ligation time. ATP concentration was reduced from 1.2 – 1.7 mM to 50 µM. Ligation time is either 0.5 h or 4 h. The reaction conditions are depicted as in table 2.1, Opt. 2. **A:** PAGE analysis of the crude ligation mixture after digesting the DNA splint. The slowest moving band corresponds to the full riboswitch whereas all other faster moving bands correspond to incompletely ligated intermediate products as indicated **B:** Yields of the full riboswitch (dark grey) as well as the 4-way ligated product (light grey) were calculated by using the Cy5 scan in **A**.

Since there was no detectable difference between the 0.5 h and 4 h of ligation time, all following ligation reactions were performed for 1h. Additionally since the overall yield of both full-length riboswitch as well as 4-way ligated product was increased compared to the first optimization step (Fig. 2.6), all following ligations were performed at a ATP concentration of 50 µM in the ligation mixture.

We observed degradation of ligated riboswitch constructs upon prolonged incubation with DNase I which has also been hinted in a previous study (42) involving T4 Dnl, plausibly due to a slight contamination of RNases in the preparation of DNase I. We observed a digestion period of 15 min (in contrast to previous experiments with 30 min DNase I digestion time) is sufficient for complete removal of the DNA splint (as revealed from the fluorescein-channel scan in figure 2.8). Therefore in all following experiments a 15 min of DNase I digestion period was employed.

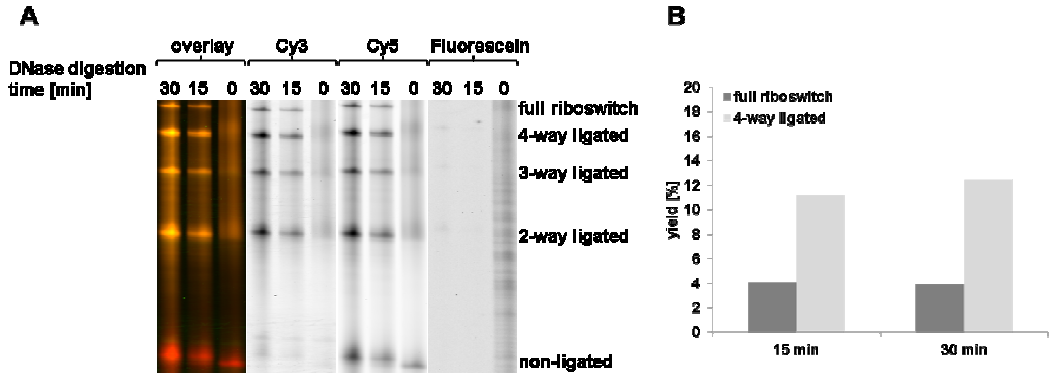


Figure 2.8: Variation of DNase I digestion time. The reaction conditions are depicted in table 2.1, Opt. 3. Samples of the same ligation reaction mixture were taken after 0, 15 and 30 min of DNase I digestion. **A:** PAGE analysis of the crude ligation mixture. The slowest moving band corresponds to the full riboswitch whereas all other faster moving bands correspond to incompletely ligated intermediate products as indicated. A smear over complete lane of 0 min in all three channels indicates the presence of the DNA splint. **B:** Yields of the full riboswitch (dark grey) as well as the 4-way ligated product (light grey) after either 15 min or 30 min DNase I digestion time were calculated by using the Cy5 scan in **A**.

Previous studies of T4 Rnl2 revealed that a relatively low RNA to enzyme ratio led to quantitative ligation yield for overlapping conglomerate ligation(183). Therefore, in our assay the T4 Rnl2 concentration was reduced from 3.6 µM to 0.5 µM resulting in RNA to enzyme ratio of 1:0.05 (Fig. 2.9). Although this change in enzyme concentration did not improve the ligation efficiency, it can be concluded from this experiment that similar ligation efficiency could be achieved by using seven fold less enzyme. Therefore in all following experiments T4 Rnl2 concentration was maintained at 0.5 µM in the ligation reaction mixture.

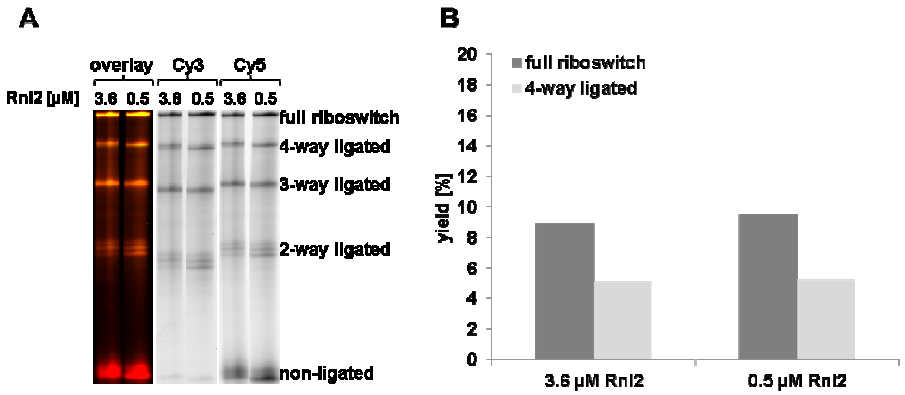


Figure 2.9: Reduction of T4 Rnl2 concentration. T4 RNA ligase 2 concentration was reduced from 3.6 µM to 0.5 µM. Reaction conditions are depicted in table 2.1, Opt. 4. **A:** PAGE analysis of the crude ligation mixture after digesting the DNA splint. The slowest moving band corresponds to the full riboswitch whereas all other faster moving bands correspond to incompletely ligated intermediate products as indicated. **B:** Yields of the full riboswitch (dark grey) as well as the 4-way ligated product (light grey) were calculated by using the Cy5 scan in **A**.

As observed earlier in our experiments, high ATP concentration led to a decrease in ligation efficiency by T4 Rnl2 (Fig. 2.7) - therefore it is mandatory to purify the dye labelled RNAs after enzymatic phosphorylations since the kinasing reaction mixture contains 2 mM

ATP to ensure quantitative phosphorylation [the presence of a phosphate group is also compulsory for successful ligation (Fig. 2.5) thereby demanding the quantitative phosphorylation of the dye labeled RNAs as an absolute requirement]. However we envisioned that the removal of polynucleotide kinase (PNK) might not be necessary, therefore the omission of phenol-ether extraction to remove the PNK prior to ethanol precipitation of the phosphorylated RNAs might lead to an increase in ligation yield. Indeed in a comparative assay we observed a higher ligation yield (Fig. 2.10) for both full-length riboswitch as well as 4-way ligated product for non-extracted samples.

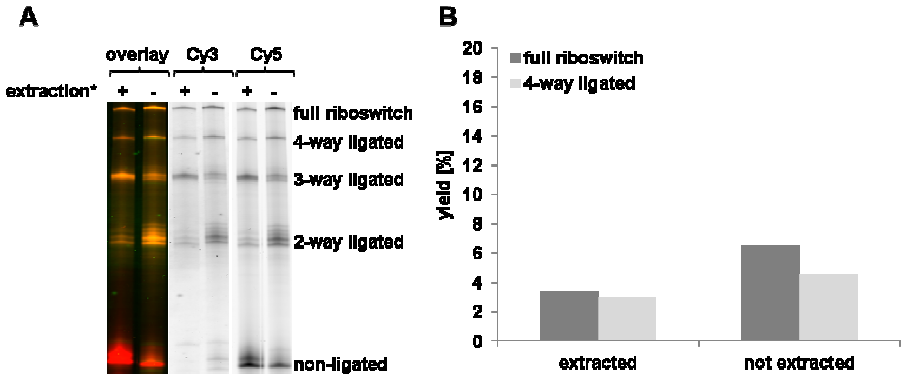


Figure 2.10: Omission of phenol-ether extraction after phosphorylation of the dye-labeled RNAs. For the ligation reaction either extracted or non-extracted dye-labeled RNA was used. The reaction conditions are depicted in table 2.1, Opt. 5. **A:** PAGE analysis of the crude ligation mixture after digesting the DNA splint. The slowest moving band corresponds to the full riboswitch whereas all other faster moving bands correspond to incompletely ligated intermediate products as indicated. **B:** Yields of the full riboswitch (dark grey) as well as the 4-way ligated product (light grey) were calculated by using the Cy5 scan in **A**.

By pursuing the above mentioned optimizations (Tab. 2.1) the overall ligation yield of a 169 nucleotide long riboswitch was increased from 1 % (Fig. 2.6) to 10 % (Fig. 2.9). This optimized splinted-ligation toolbox, consisting of five 33-34 nucleotide long ligation fragments and one splint covering the complete ligation region, was used to synthesize various dual-labelled SAM-I riboswitch constructs in high yields for bulk as well as single molecule FRET measurements.

All optimization ligations were performed in a volume of 20 µl whereas preparative ligations were performed in 300 µl. The technical difficulties e.g., precipitation efficiency and losses during various extractions became negligible during preparative ligations because of the presence of higher amounts of RNA. Therefore even higher yields were obtained for preparative ligations suggesting that this splinted-ligation toolbox is amenable for scale-up even to higher reaction volume.

2.5. Pitfalls in splinted ligation – use of T4 RNA ligase 2

The number of different commercially available DNA ligases is a lot higher than RNA ligases. In addition, the majority of the DNA ligases (T4 Dnl in this case) are double stranded ligases, all of which might account for their popularity in splint-mediated RNA ligation since the ligation competent complex in this case is also double stranded in nature. However many studies indicate inefficient product release (or product inhibition of the enzyme) by the DNA ligases in case of DNA-splint mediated RNA ligation where the ligation product is a RNA-DNA heteroduplex (41). Similarly, a lot of effort has been put for characterization of single stranded phage RNA ligase, T4 RNA ligase 1 (T4 Rnl1). Because of its very low nick joining efficiency on double stranded substrates, complicated ligation schemes involving DNA splint which forces the ligation site to be locally single stranded, were established to prepare long RNA by ligation (Section 1.3)(44,45). On the other hand, T4 Rnl2 is a double-strand specific RNA ligase and its nick joining efficiency was reported to be much higher than that of T4 Dnl or T4 Rnl1 for double stranded substrates(43). However despites its discovery a long time ago, its extremely high RNA nick joining activity and its commercial availability, the use of this enzyme in splinted-ligation was relatively rare for unknown reasons(181,182,184). In our study, we have successfully developed a splinted-ligation toolbox for the production of dual-labelled SAM-I riboswitch constructs in high yields using this enzyme. This ligation protocol can further be scaled-up to prepare even higher amounts of dual-labelled RNA constructs for biophysical studies.

While setting up a splinted-ligation scheme it should be kept in mind that the higher the number of ligation fragments involved to construct the full-length product, the lower would be the statistical probability for the formation of ligation competent complex. On the other hand, for preparing long site-specifically dye-labelled RNA from smaller synthetic RNA fragments by ligation, it is beneficial to design the ligation scheme involving only short dye-labelled RNA sequences because of i) higher yields and reliable sequence homogeneity (therefore avoiding error-correction steps during enzymatic ligation) in chemical RNA synthesis for short RNAs compared to very large RNAs and ii) higher number of dye-labelled constructs can be prepared out of only a very few dye-labelled ligation fragments in a combinatorial manner, thereby facilitating the finding of the labelling positions and the best suited constructs easier as well as making the complete ligation protocol more cost-effective. Despite the difficulties in the formation of ligation competent complex in a five-way ligation and the possibility of having stable secondary structures for large RNA, an overall yield of 10 % in splinted-ligation during preparation of a 169 nucleotide long complete riboswitch is a substantial advancement in the field of splint-mediated RNA ligation.

2.6. Bulk- and SM-FRET measurements for folding analysis of the SAM-I riboswitch[3]

Ligand induced folding of the SAM-I riboswitch was addressed by measuring changes in FRET efficiencies of the dye-labels upon SAH or SAM addition at either 10 °C or 25 °C (Fig. 2.11 and 2.12). It was found that the FRET changes are similar at both temperatures, which is consistent with the previously postulated thermal stability of the SAM-I riboswitch aptamer(185). The bulk-FRET efficiency values for all six different constructs in their native states as well as upon incubation with SAM and SAH are given in figure 2.11.The differences in the FRET efficiencies (E_{FRET}) from the native, ligand unbound state to the metabolite-bound state upon addition of either SAM or SAH were calculated (Fig. 2.12).

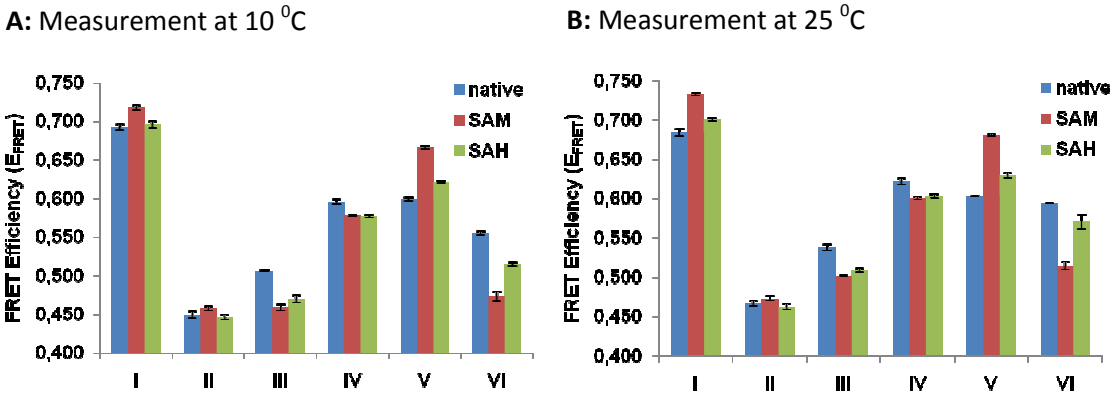


Figure 2.11: Bulk FRET efficiencies for all six SAM riboswitch constructs in native state as well as upon addition of SAM and SAH. All measurements were performed both at 10 °C **(A)** and 25 °C **(B)** respectively.

Irrespective of the ligand used, SAM or SAH, we always observed same numerical signs for the ΔE_{FRET}. This indicates that either SAM or SAH induces the folding of the riboswitch towards one single (or a particular group of) conformation(s), albeit to different extents.

In figure 2.12 bars within the positive area indicate that the FRET efficiency increased upon SAH or SAM addition meaning that the dyes in these respective constructs converged. This observation is true for constructs I and V. Bars in the negative area in figure 2.12 indicate that the FRET efficiency decreased upon SAH or SAM addition meaning that the dyes in the respective constructs diverged from each other. This behavior was observed for constructs III, IV and VI. For construct II the bars in figure 2.12 stayed close to zero indicating that there was no significant change in FRET efficiency.

[3] This part of the work has been performed together with Marika Ziesack

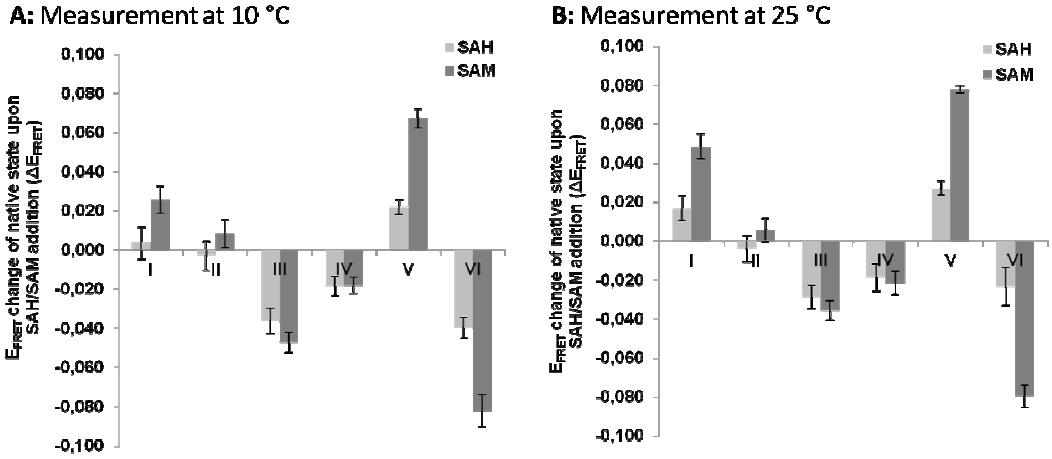


Figure 2.12: Changes in FRET efficiency (ΔE_{FRET}) of the native state upon SAH or SAM addition for all six SAM riboswitch constructs. E_{FRET} of the native state of all constructs as well as E_{FRET} after addition of SAH or SAM was measured three times. The means as well as the standard deviations were calculated and values for the native state were subtracted from respective SAH/SAM values to obtain ΔE_{FRET}. The errors were calculated using Gauß error Propagation. All measurements were performed both at 10 °C **(A)** and 25 °C **(B)** respectively.

The structure of the SAM-I riboswitch depends on tertiary interactions between the regions L2, J3/4 and J1/4(179). These interactions result in a RNA structural organization, in which two sets of coaxially stacked helices (P1-P4 and P2-P3) orient relative to each other with an angle of ~ 70°(166). L2 and J3/4 form a pseudoknot that stabilizes the global architecture of the SAM-I riboswitch(7). According to the crystal structure of the SAM-I riboswitch an adenine within the J1/4 junction, a uridine within the J3/4 junction and another adenine within the P2b stem are important for tying the P1-P4 stack to the pseudoknot(179)(Fig. 2.1). Furthermore, *via* aminopurine-quenching experiments it was found, that SAM binding changes the environment of these conserved nucleobases(166). It was suggested that this alteration is important for stacking of P1-P4 stem which we were able to confirm in this study. Changes in E_{FRET} upon addition of the cognate ligand were found in the constructs I and V (positive bars, Fig. 2.12) and construct IV (negative bars, Fig. 2.12). The changes in constructs I and V suggest that the P2 and P1 stem as well as the P3 and P4 stem converge upon SAM addition (Fig. 2.13). The change in construct IV indicates that P2 and P4 diverge (Fig. 2.13). Summarizing these results, the global architecture was found to change upon ligand addition because of the SAM induced interactions between J1/4, J3/4 and P2b(166). Thus it can be concluded that the pseudoknot was formed in the absence of the ligand.

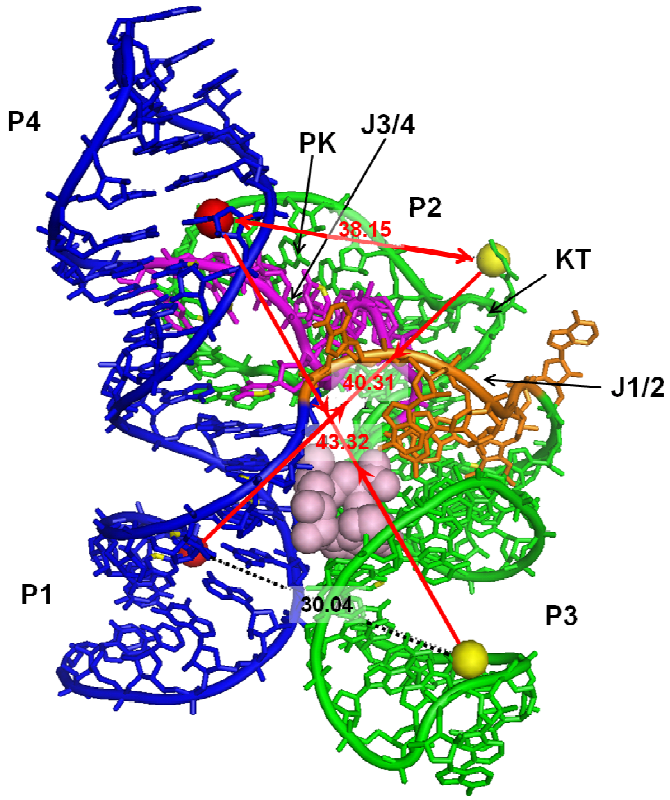


Figure 2.13: Tertiary structure of the SAM-I riboswitch aptamer (*Thermoanaerobacter tengcongensis*) according to ref 174. The P1/P4 helix stack is highlighted in blue and the P3/P2 helix stack is highlighted in green. The J1/2 is shown in bronze and the J3/4 is shown in violet. SAM is represented as a bluster of grey spheres and the Cy5 dye as red and the Cy3 dye as yellow sphere. The distance between the dye positions is given in Å. Dotted line shows distance between mutual dye positioins found to be unchanged in this study. Filled lines represent the same with red arrows indicating whether the dyes are converging or diverging upon SAM or SAH addition. *Legend: P = paired, J = joining, KT = kink-turn, PK = pseudoknot.*

In a previously proposed model(185), the SAM-I riboswitch adopts a pre-binding conformation that is similar to the SAM-bound state. Upon ligand binding the SAM-bound conformation is 'locked' down through an induced-fit mechanism. The structural dynamics observed in this work (Fig. 2.13) are consistent with this model.

In most of the previous studies only the aptamer of the SAM-I riboswitch was used(166,179). However, in this work we examined the complete SAM-I riboswitch including its expression platform. In the SAM-unbound state the P1 stem forming residues and the 3´-end of the riboswitch form an anti-terminator(166) resulting in transcription of the downstream mRNA. Upon ligand recognition, the formation of an anti-anti-terminator (P1) is induced and a terminator is formed within the 3´-end of the riboswitch (Fig. 2.2). This ligand-induced reorganization could be observed as changes in E_{FRET} between dyes located within the terminator and the P4 stem as well as within the terminator and the P3 stem (Fig. 2.12, CIII and CVI). In both cases the change in E_{FRET} is negative indicating that these two dyes diverge.

As observed before (Fig. 2.11 and Fig. 2.12) the conformational changes induced by SAM are more pronounced than those induced by SAH. This can be expected since the riboswitch has been shown to strongly discriminate between these two ligands(175,176). However, there were two different observations which was not expected, namely having the same numerical signs for the ΔE_{FRET} in case of all constructs upon incubation with SAM or SAH (Fig. 2.12) indicative of a folding event which is convergent to one group of conformations and secondly these effects are more pronounced in case of C-III and C-VI compared to all other constructs (Fig. 2.11 and Fig. 2.12). These constructs bear one fluorophore which is directly situated at the expression platform. Therefore these constructs are likely to provide more information about the conformational dynamics of the riboswitch which is responsible for its gene regulatory properties. Does the riboswitch aptamer response differently towards ligand-recognition in the presence or the absence of the expression platform? And does this riboswitch, during co-transcriptional folding *in vivo*, capable of making use of either SAM or SAH to direct the correct folding of its aptamer but makes the decision for the gene regulation only upon binding to its cognate metabolite SAM? Although it might not be possible to address these questions definitively by only *in vitro* experiments, these constructs bearing fluorophores at their expression platform are surely one of the best candidates for further investigations.

Thus we performed further SM-FRET[4] studies with C-III and C-VI. At first the riboswitch molecules were titrated with magnesium ions to observe the structural compaction. The results are depicted in figure 2.14. The data fit best with a two-state model for C-III and a three-state model for C-VI.

[4] SM-FRET measurements have been performed by Andrei Kobitski at Karlsruhe Institute of Technology (KIT)

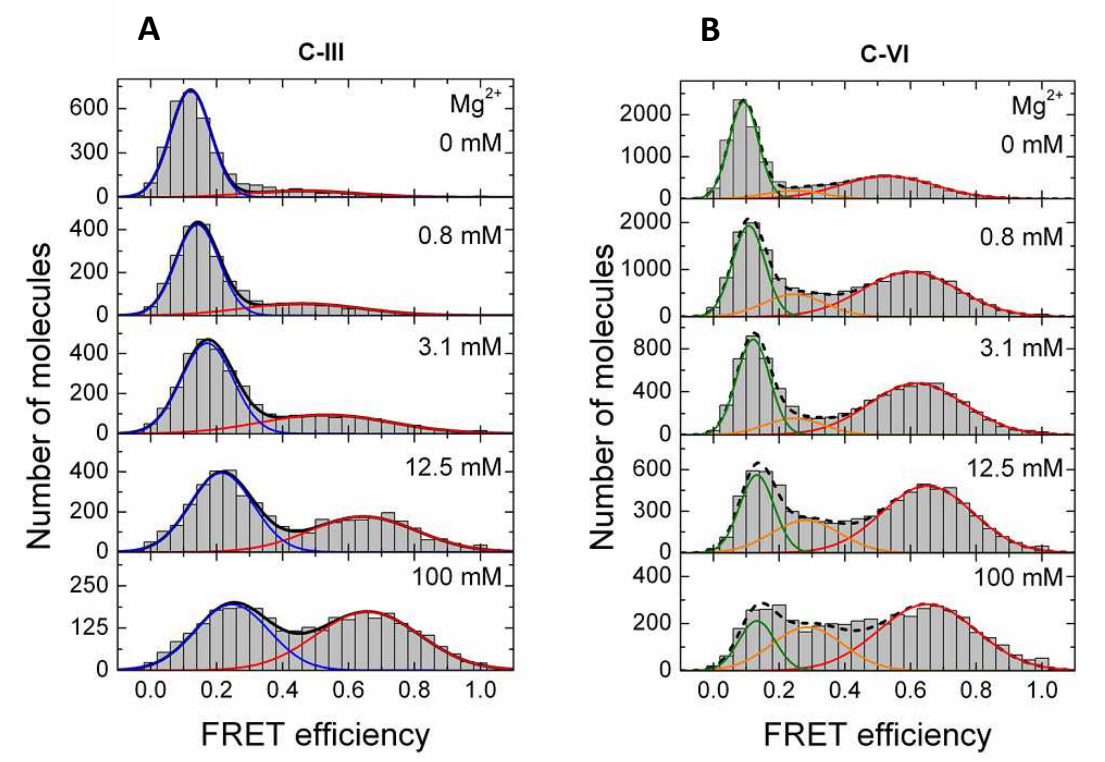


Figure 2.14: FRET efficiency histograms of freely diffusing C-III (A) and C-VI (B) riboswitch molecules at different concentrations of Mg^{2+} ions. Lines represent the best-fit two-state (for the C-III construct) and the three-state (for the C-VI construct) model distributions.

Furthermore, to validate the ligand induced folding we titrated the SAM-I riboswitch with SAM and SAH (Fig. 2.15).

As a summary, we have observed similar folding phenomena for the complete riboswitch in response to its cognate metabolite SAM as reported in previous studies involving constructs carrying only the aptamer. Additionally we have observed certain conformational changes induced by a closely related yet non-cognate metabolite SAH. These conformational changes all together might be useful in dissecting the mechanism of riboswitch mediated gene-regulation *in vivo* by premature transcription termination at a molecular detail.

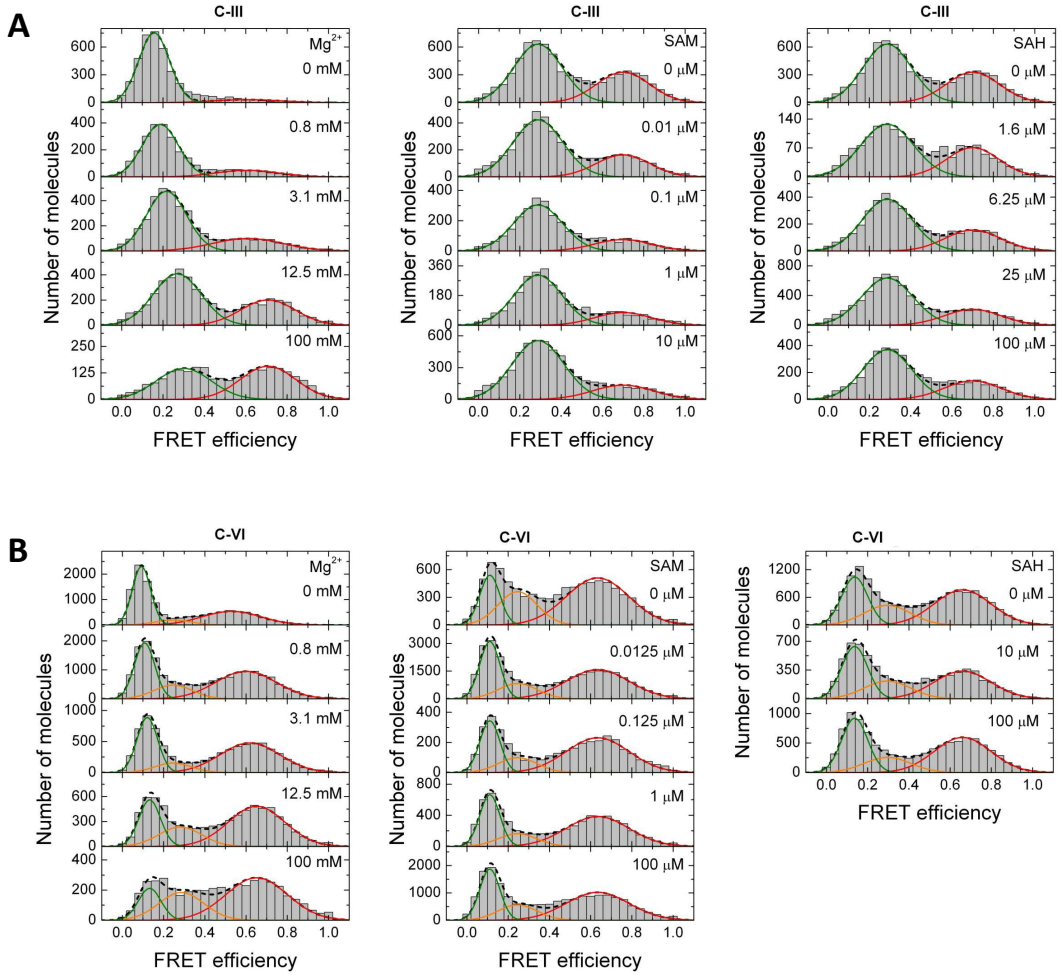


Figure 2.15: (A) FRET efficiency histograms of freely diffusing C-III riboswitch construct at different concentrations of Mg^{2+} ions (left), SAM (middle), and SAH (right); SAM and SAH titration measurements were carried out at 20 mM Mg^{2+} concentration. Lines represent the best-fit two-state model distributions. (B) FRET efficiency histograms of freely diffusing C-VI riboswitch construct at different concentrations of Mg^{2+} ions (left), SAM (middle), and SAH (right); SAM and SAH titration measurements were carried out at 20 mM Mg^{2+} concentration. Lines represent the best-fit three-state model distributions.

2.7. Conclusion and future work

Here we have successfully established a splinted-ligation strategy to prepare dual fluorophore-labelled full-length riboswitch constructs for bulk- and SM-FRET measurements for studying ligand induced folding dynamics of the SAM-I riboswitch. To our conclusion, T4 RNA ligase 2 is much more efficient in joining nicks of RNA in a double stranded environment compared to T4 DNA ligase. We have optimized the ligation scheme to improve the overall yield of the full-length ligation product from 1 % to 10 %. However, we believe that further optimizations for the pH of the ligation buffer might improve the ligation yield. It is noteworthy here that we have observed incorrect annealing of some of the ligation fragments when only four fragments are used together with the DNA-splint previously used for the 5-

way ligation. This indicates that the correct formation of ligation competent complex (LCC) can be prevented, ultimately leading to decreased overall yield of the product. This seems to be a general problem for all transcriptionally acting riboswitches. This is due to the fact that for all transcriptionally acting riboswitches, the 3´-end of the P1 stem has certain sequence complementarity to the 5´-end of the terminator hairpin. Therefore even though longer ligation fragments are used (in our case each fragment consists of 34 nucleotides which is almost as large as gene-specific primers used in our laboratory to amplify specific riboswitch transcription templates from isolated genomic DNA) there can always be some extent of incorrect annealing. This can be prevented by using multiple smaller DNA splints, each capable of hybridizing to a particular RNA fragment and afterwards leaving a sticky end for overlap with its neighbouring fragment. Thus, all RNA fragments can be separately annealed to its specific DNA splint, then brought together into one reaction vessel and allowed for the formation of the complete LCC by the help of the sticky ends present in each heteroduplex. This ligation strategy is currently under investigation in our laboratory.

In case of our aim to address ligand induced folding dynamics of riboswitches, we have observed similar folding phenomena for the complete riboswitch with its expression platform as reported earlier for constructs carrying only the aptamer. However we made a couple of observations in studies involving a slightly different yet non-cognate metabolite which can not be explained by any known facts about this riboswitch to date. We believe that these data might further shed light on the gene regulation of riboswitches by premature transcription termination and hopefully settle down the long standing discussion about the ``switching`` behaviour of these regulatory RNA elements *in vivo*.

3. Optimizations of copper-catalyzed click reaction for biomolecule labeling

3.1 Ligand-accelerated copper-catalyzed click reaction - scientific background

As mentioned in section 1.4, the use of click chemistry for nucleic acid functionalization by a two-step chemo-enzymatic strategy has multiple advantages over one-step direct labeling. The most heavily used among click reactions is the copper-catalyzed click reaction (CuAAC) (section 1.4.3). The small size of the reactive alkyne and azide made this reaction particularly useful since substrates carrying any one of these two modifications have only minor structural differences compared to their unmodified analogues which increase the likelihood of these compounds to be accepted as substrates by the enzyme. The main bottle-neck of this reaction is however the cytotoxicity copper.

To circumvent many of these problems associated with copper, various copper stabilizing ligands have been introduced(74,96-99), therefore converting this ``ligand-free`` reaction(100) to a ``ligand-accelerated metal-catalyzed`` reaction(101). These ligands are meant to stabilize Cu(II) and Cu(I), as well as to keep the harmful ascorbate oxidation products low, thereby providing more biomolecule compatible reaction conditions(102) (section 1.4.3). However Cu(I) should not be too well stabilized, otherwise it would not be available for coordination with alkyne and azide and therefore catalysis.

With these advancements in CuAAC, this reaction has been successfully applied in various studies involving DNA(83). A range of different pyrimidine(103) phosphoramidites bearing alkynes were synthesized and incorporated in solid-phase DNA synthesis, thereby enabling site-specific modification of short DNA. Also, various triphosphates bearing either alkyne or azide were synthesized and incorporated in primer extension studies, thereby enabling random modification of long DNA(104,105). Moreover, these statistically modified DNAs can be used to prepare uniformly metallized DNA(106), gold nanoparticle DNA conjugates(107) and to immobilize DNA on a glass-surface in a pre-defined micropattern(108).

Despite its tremendous success as a click reaction for biomolecule labeling, examples of CuAAC-labeling on RNA are rather limited(186), which is plausibly due to the inherent instability of RNA towards metals ions, pH and elevated temperature. Careful examination of current literatures revealed that majority of CuAAC reactions performed on DNA require mM concentration of DNA for efficient labeling and often has been performed on fully protected, support-bound, short DNA at elevated temperature under neat or anhydrous conditions requiring organic solvents with rigorous exclusion of oxygen from the reaction

mixtures(96,103-105). Furthermore, the majority of these protocols require presence of multiple convertible residues per biomolecule for efficient labeling(187). Therefore we sought to establish a CuAAC protocol which will be completely compatible with standard molecular biology buffers and does not require any technically complicated synthetic methodologies like anhydrous conditions or rigorous exclusion of oxygen, as well as will allow us to achieve fast and quantitative labeling of nucleic acids bearing single alkyne/azide modification per biomolecule at room temperature at sub-micro molar concentration without any nucleic acid degradation.

3.2 Different classes of Cu-ligands

Among various different Cu-stabilizing ligands used for CuAAC, the most heavily used ones are built around a central structural core of tris(heterocyclemethyl)amines. Different classes of Cu-stabilizing ligands are shown in figure 3.1. The class I ligands are highly active when used in a 2- to 4-fold molar excess relative to Cu in aqueous reactions having no donor solvents, e.g. DMSO. On the contrary, the class II ligands are highly active when used in stoichiometric amounts relative to Cu but lose most of the activity when used in higher excess. The class III consists of relatively inactive systems over all ligand-metal ratios. The commercially available Cu-coordinating ligand TBTA (Fig. 3.1 and 3.2) belongs to the class I. We have used a water soluble version (THPTA, Fig. 3.2) of the commercially available class I ligand TBTA for our CuAAC optimizations for the purpose of biomolecule labeling.

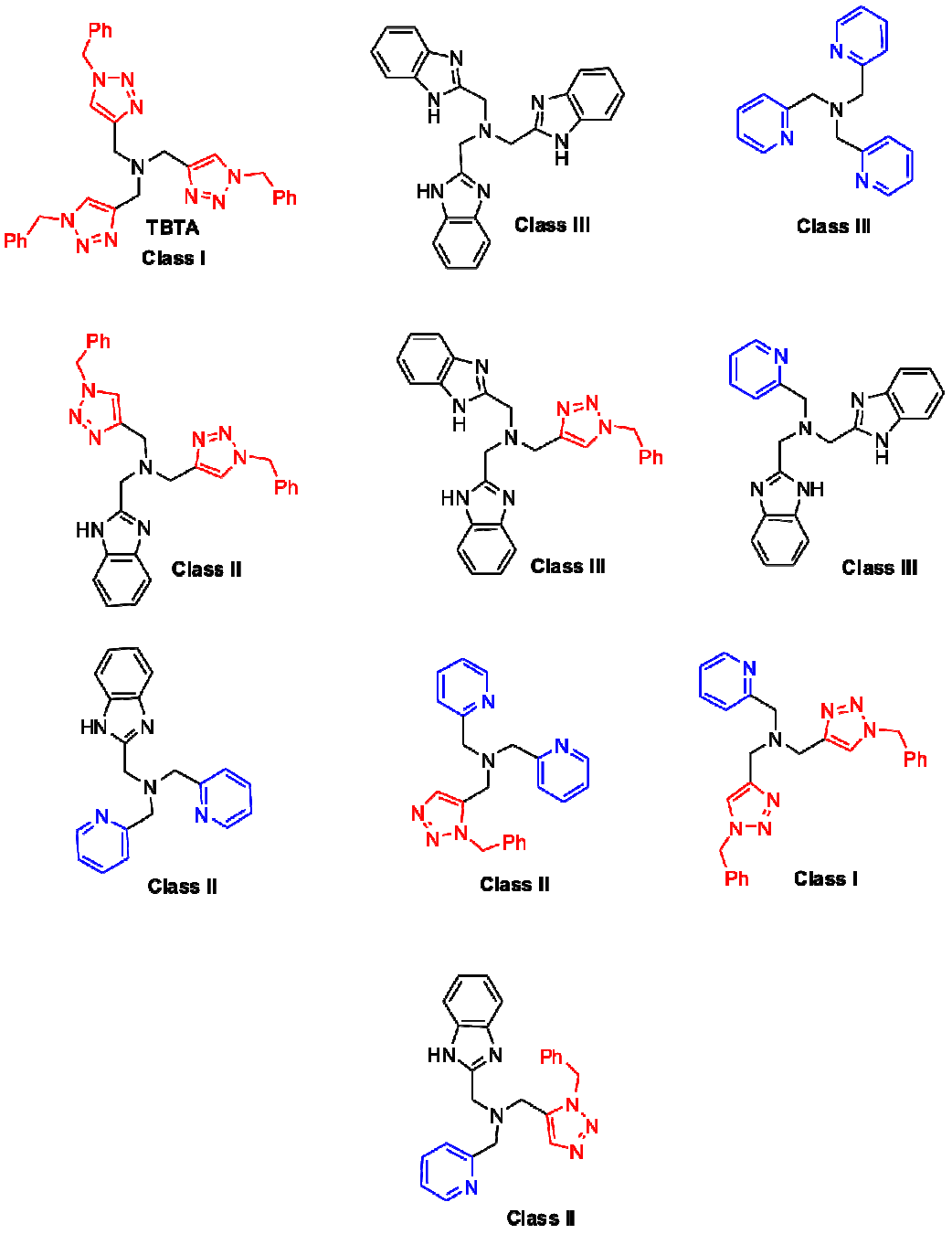
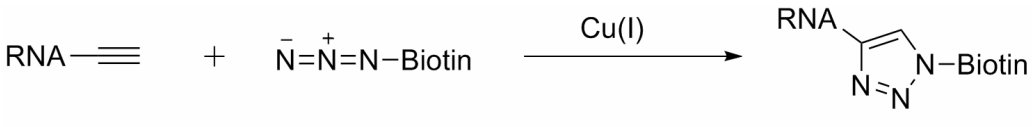


Figure 3.1: Chemical structures of different classes of Cu-ligands.

3.3 Optimizations of click chemistry for biomolecule labeling

We performed our CuAAC optimizations on a 34 nucleotide long internally alkyne modified RNA with a biotin azide as shown below.



Due to the relatively short length of our reacting RNA molecule attachment of a biotinyl residue resulted in a marked difference in its electrophoretic mobility on a denaturing polyacrylamide gel which therefore allowed us for an unambiguous assessment of the CuAAC conjugation efficiency. As mentioned earlier, the copper should not be too well coordinated otherwise it will not be available for catalysis. Therefore we sought to find out first the best buffer for biomolecule labeling by CuAAC. Although a vast number of molecular biology protocols require Tris-based buffers, to our rationale this buffer should rather be avoided because of the presence of a primary amine in Tris, since primary amines are excellent copper chelators. Furthermore, we envisioned that since the Cu(I) has to be coordinated by the reactive alkyne and terminal alkyne co-ordination is severely improved by the ionization of the acetylinic proton, thus use of a slightly alkaline buffer might facilitate the reaction. However, care should be taken about the general instability of RNA in alkaline buffers. Therefore in our experiment we made a direct comparison between pH 8.5 Tris-HCl buffer, pH 7.0 sodium-phosphate buffer and water (Fig. 3.2 panel A). We observed slightly better conjugation efficiency in phosphate buffer compared to Tris buffer while reaction in water is least efficient. Moreover, we have tested the effect of ionic strength of the medium on the CuAAC conjugation efficiency by pursuing the reaction in buffers containing high concentrations of potassium ions (Fig. 3.2 panel B) and observed no effect on the conjugation efficiency. Next we made a direct comparison between two different copper coordinating ligands (TBTA and THPTA, Fig. 3.2 panel B and E) both belonging to the same class but one being highly soluble in water and the other is only slightly soluble. We observed low reaction yield and some extent of RNA degradation for the commercially available water-insoluble ligand TBTA. Therefore, it was mandatory to establish the synthesis of the THPTA for further use of CuAAC in our laboratory. Additionally the CuAAC conjugation can be severely improved when high concentrations of the other reaction partner is used and lead to quantitative reaction yield within 1 hr of reaction time (Fig. 3.2 panel C and D). It is worth mentioning that due to the existence of one previous report(186) claiming CuAAC labeling of RNA in the absence of any copper stabilizing ligand, we have actually tested the importance of THPTA in CuAAC reaction involving RNA. We indeed observed severe RNA-degradation (often to even single nucleotide level) which is in accordance with previous

studies in the field of artificial nucleases(86,87). The reason that in the aforesaid report the RNA labeling was successful is due to the fact that the labeling has been performed on fixed cells. Finally we performed CuAAC conjugation in presence of various concentrations of $CuSO_4$.

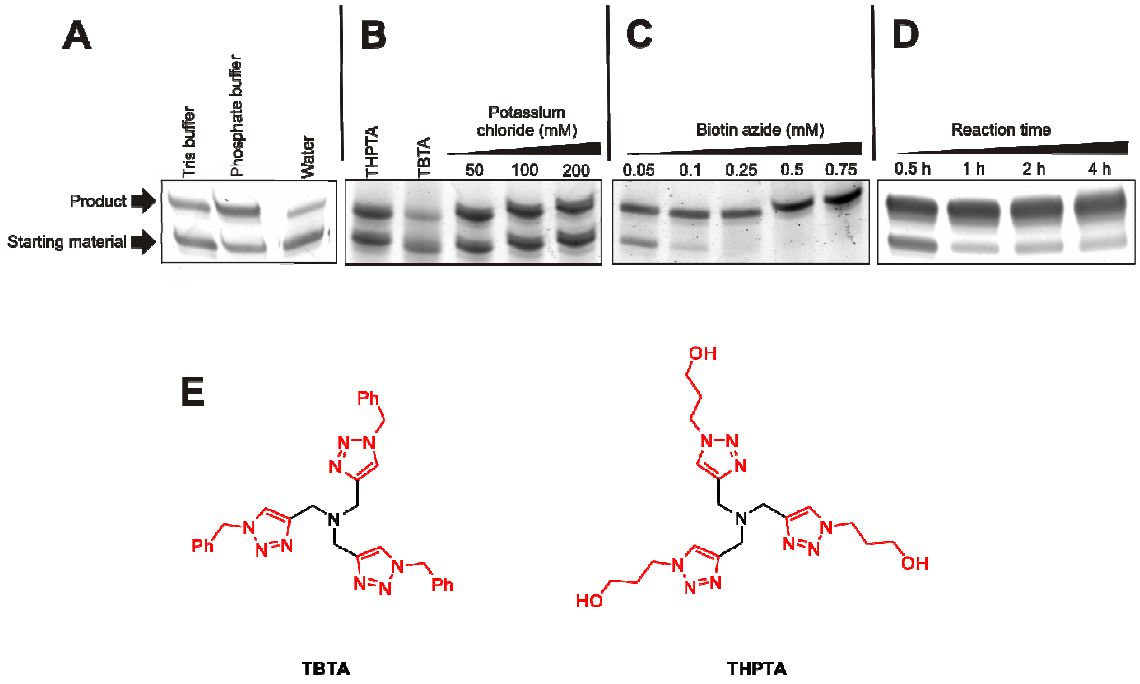


Figure 3.2: A-D) CuAAC optimizations. E) Chemical structures of TBTA and THPTA. Cu(II), ligand and sodium ascorbate end concentrations were maintained at 0.5, 2.5 and 5 mM respectively through out all experiments depicted here.

All of these taken together, we reached the conclusion that i)CuAAC should be performed in neutral, amine-free buffers, ii)the copper-stabilizing ligand is absolutely necessary to prevent degradation of nucleic acids during the reaction, iii)the ligand should be used in 5 fold excess over the copper ion, iv)if Cu(I) is generated by *in situ* reduction of Cu(II) salts with sodium ascorbate, the reducing agent should be used in 10 fold excess over Cu(II). To our rationale, 500 µM of Cu(II) is sufficient to achieve a quantitative labeling of alkyne bearing nucleic acids at 10 µM end concentration with the azide counterpart at 50 µM end concentration.

3.4 Current limitations of click chemistry

Despite of all advantages mentioned for click chemistry in general, one potential drawback of this approach is that it often requires newly synthesized compounds, whereas many people working in these areas are non-chemists and therefore synthesizing a complete new

compound is beyond the intellectual and technical capabilities of many laboratories. Although a lot of materials required for click chemistry are nowadays available from different commercial providers at reasonable prices, these cover only the mostly used, top-ranked click reactions. For new or less used click reactions one is still dependent on synthetic chemists. Therefore the development of new click reactions is somewhat restricted among chemists. However non-chemists from various other disciplines can contribute to the applications of these novel reactions towards unraveling biological mechanisms at molecular level. In the following chapter we will use this reaction for chemo-enzymatic nucleic acid functionalization.

4 Site-specific functionalization of nucleic acids

4.1 Chemo-enzymatic labeling of the 5´-end of RNA

4.1.1 Scientific background

Although statistical RNA modifications might be useful in studying certain global phenomena related to RNA(186,188-190), the majority of the applications demand site-specific conjugation of modified chemical entities to nucleic acids. In this chapter we will focus on different strategies for modifying the 5´-end of RNA. Although the number of different modifications introduced so far into RNA is as large as the number of various different applications requiring such chemically modified RNAs, the number of various different tools available to introduce such modifications is rather limited(191).

5´-end modified RNAs have widely been used for SELEX(192,193), RNA conformation and dynamics(2), *in vitro* translation(194) as well as in affinity chromatography(195). Among many different RNA 5´-end labeling techniques the following two strategies are of notable interest.

ATP-γ-S and polynucleotide kinase based approach for labeling the RNA 5´-end

This technique allows us a for a non-*de novo* labeling of RNA, of either synthetic or biological origin, in a two-step protocol where at the first step a phosphorothioate residue has been transferred from the ATP-γ-S to the 5´-OH of the RNA of interest by polynucleotide kinase (PNK) followed by reaction of that phosphorothioate functionality with a maleimide group to further attach the label of interest(196,197). The underlying principle of this reaction is that the nucleophilicity of phosphorothioate is slightly higher than that of water. Therefore compounds carrying a maleimide moiety will selectively react to the 5´-end of the RNA carrying the phosphorothioate group. However, most efficient PNK mediated phosphoryl transfer reaction requires a 5´-OH on the RNA – therefore samples isolated from biological origin needs to be dephosphorylated first by the treatment with alkaline phosphatase. Although ATP-γ-S is accepted as a substrate as efficiently as ATP, the overall yield of this two-step labeling approach is very low. This is due to the low nucleophilicity of phosphorothioate groups, thereby requiring a very high concentration of the maleimide reaction partner and prolonged reaction time ultimately leading to a poor labeling efficiency.

Labeling by transcription priming – initiator nucleotide based approach

Contrary to the previously described polynucleotide kinase based approach, this labeling strategy is applicable for *de novo* labeling of RNA. This technique allows labeling of short as well as very large RNAs. Large unmodified RNAs can be conveniently prepared in big amounts by *in vitro* run-off transcription from DNA templates, either synthetic or biological, using different bacteriophage RNA polymerases (RNAP) under the control of their respective promoters(198-201). During run-off transcription RNA can be selectively 5´-modified, employing the relaxed substrate tolerance of the polymerase at the initiation step(202). This approach uses modified guanosine or adenosine (depending on the type of promoter used) analogues, functionalized though the phosphate, which can only be incorporated at the transcription start because of a lacking triphosphate(203-206) - are therefore termed as initiator nucleotides. The most commonly used polymerase is derived from the T7 phage(203,204,207), while T3 and SP6 RNA polymerases (RNAPs) are also reported for similar purposes in a few studies(205,208-210). This strategy for RNA 5´-end functionalization is particularly attractive due to the commercial availability of these enzymes. A vast number of modifications have been introduced using this strategy by T7 RNAP. These include biotins(195,211), aldol(212,213) and Michael(214,215) reactants, aromatic hydrocarbons(216,217), allylic ethers and carbamates(217), dienes(218), dienophiles(124), small molecule coenzymes (CoA, NAD, FAD)(219) and various fluorophores(220,221). Furthermore initiator nucleotides carrying convertible residues like aliphatic amine(217,222), thiol(222-225), phosphorothioate(226), azide(227) or protected aldehyde(228) have been synthesized and successfully incorporated by T7 RNAP at the 5´-end of RNA. Moreover, examples of initiator nucleotides carrying orthogonally cleavable (photo(213), DTT(229) or enzymatic(228)) linkers between the RNA transcript and the functional tag also exist for efficient release of the conjugated moiety.

4.1.2 Universal initiator dinucleotides – current work[5]

Despite its generality and versatility, the transcription initiation approach for RNA modification suffers from several drawbacks: i) it requires the *de novo* synthesis (often over many steps) of the complete initiator molecule as well as optimizations of enzymatic reaction conditions for its incorporation for each single organic moiety of interest. This completely eliminates the scope of testing a variety of modifications in a high through put manner, which might be necessary in certain applications, e.g. studies involving fluorescence resonance energy transfer (FRET), where it is mandatory to first screen for a good donor-acceptor pair. ii) All currently known initiators generate RNA transcripts whose 5´-ends are blocked and therefore

[5] This part of the work has been performed together with A. Krause.

can not further participate in enzymatic ligations involving that modified end, restricting this approach strictly to 5´-end modification.

Therefore most valuable would be a universal initiator nucleotide that can be incorporated into transcripts under conditions that are optimized once, and that contains a convertible functional group for further functionalization with a variety of different labels. This two-step labelling approach is certainly advantageous over one-step direct labelling especially in case of bulky modifications where enzymatic activity is severely hampered due to the sterical bulk imposed by the modification.

To this end, we envisioned that copper-catalyzed click chemistry might be the right choice for the second conjugation step in our strategy since the reactive moieties, azides and alkynes, in copper catalyzed click reaction are particularly small. Furthermore, due to the bioorthogonal nature of the reaction, this should allow us to selectively functionalize our modified transcript in a milieu of various other functional groups as present in most molecular biology applications (Section 1.4.3). Although the cellular toxicity of copper somewhat restricted the use of this reaction to *in vitro* applications, this is not a concern here since the RNA is synthesized *in vitro* by run-off transcription. To date copper catalyzed click reaction has been successfully used in various applications of DNA(81,83,103-108,187,189,230-232) and RNA(186,190,227,233,234) functionalization *in vitro* or in fixed cells. From synthetic viewpoint it might be easier to incorporate alkyne rather than the azide due to the chemical instability of azides in standard phosphoramidite coupling conditions(235).

Previous studies with di-, tri- or short oligonucleotides for transcription initiation by T7 RNAP revealed highly efficient transcriptional priming with these oligonucleotides and improved transcription yield(236). This has led us to assume that in addition to a more efficient enzymatic incorporation, di- or oligonucleotide initiators would likely allow the use of alternative attachment sites for the non-nucleosidic residue, potentially leaving the 5´-terminus of the transcript available for subsequent enzymatic manipulations, most notably ligations.

Based on this idea, here we report the synthesis and enzymatic incorporation of a novel class of nucleobase-modified clickable dinucleotide initiators (EUpG, EdUpG and OdUpG) (Fig. 4.1) for site-specific RNA functionalization at the 5´-end employing T3, SP6, and T7 bacteriophage RNA polymerases under the control of their cognate class III promoters. Due to the small size of the non-nucleosidic residues (in these cases alkynes) and their attachment through the non-Watson-Crick face of the 5´-nucleoside of our dinucleotide initiators, these compounds bear high structural similarities with their unmodified analogues. This design in turn led to near-quantitative initiator incorporation in many cases and also allowed efficient transcription initiation by three different phage polymerases,

namely T3, SP6 and T7 RNAP. Although T7 RNAP has been widely used for initiator-based RNA 5´-end functionalization, we are not aware of similar approaches using SP6 and T3 RNAP. Given the generality of transcription initiation mechanism by phage polymerases we assumed that SP6 and T3 RNAP can also be used for the same purpose. Noteworthy here that the transcriptional activity of these polymerases *in vitro* differ from each other depending on the transcript length, NTP and magnesium ion concentrations, salt compositions in the buffer and temperature(237). Therefore it is beneficial to establish a labelling protocol which is compatible with all three different polymerases. Albeit in a different context, SP6 and T7 promoters have been placed in opposed direction together with suitable termination sequences to allow the synthesis of either sense or antisense unmodified RNA from a single recombinant plasmid(238).

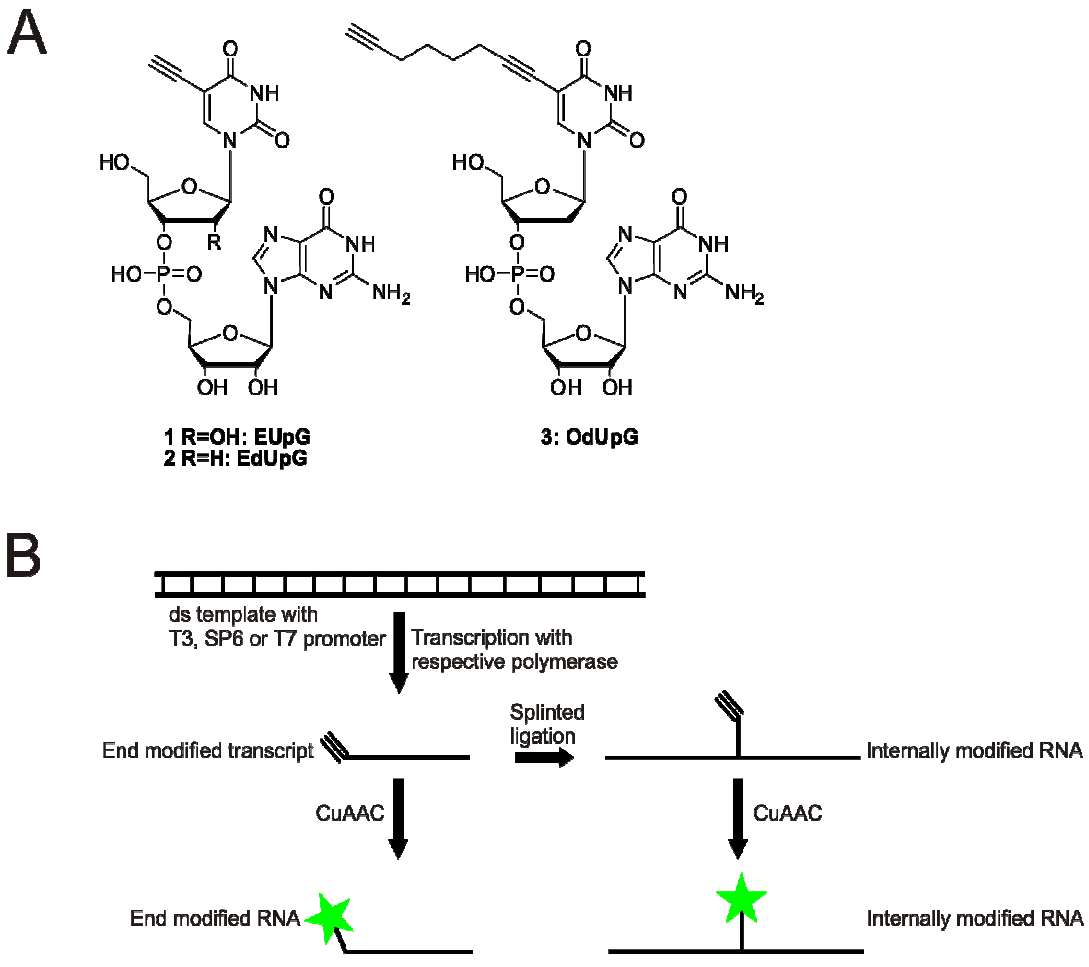


Figure 4.1: A) Chemical structures of three different initiator dinucleotides used in this study. B) Schematic representation of our initiator-based site-specific RNA functionalization approach.

The initiated transcripts have further been functionalized with a diverse array of organic moieties by CuAAC. In this way it is possible to functionalize the target RNA at its 5´-end with a diverse array of biologically relevant functional tags without the need for *de*

novo synthesis of new initiator nucleotides or optimization of enzymatic incorporation for each new compound. Therefore these dinucleotides can be regarded as ``universal initiator dinucleotides``. To investigate the robustness of our labelling approach we have synthesized the 209 nucleotide long complete glycine-responsive tandem riboswitch aptamer from *Bacillus subtilis* by *in vitro* run-off transcription and functionalized the transcript at its 5´-end with a fluorophore. As previously mentioned due to the degenerative nature of chemical oligonucleotide synthesis (Section 1.4.1), RNA sequences of this length carrying a modification at its 5´-end is impossible to prepare by any existing chemical RNA synthesis protocol to date. Furthermore, using this strategy we have functionalized a 233 nucleotide long random pool of RNA from a 252 nucleotide PCR amplicon with a fluorophore. This also demonstrates the sequence independency and generality of our approach for labelling large functional RNAs.

Finally RNA transcripts primed with our dinucleotide initiator has been subjected to enzymatic phosphorylation followed by splinted ligation to a second RNA strand, thereby converting a terminal modification to an internal one which further broadens the scope of our approach for click-type modification of RNA.

4.1.3 Design of dinucleotide initiators EUpG, EdUpG and OdUpG

The design of dinucleotides **1**, **2** and **3** (Fig. 4.1) aimed to combine a chemically variable "carrier-nucleoside" at the 5´-position with a constant guanosine building block at its 3´-position (Fig. 4.1). While the 3´-guanosine nucleoside would be indispensable for transcriptional initiation from bacteriophage class III promoters, the 5´-nucleoside constituted a platform for the attachment of the non-nucleosidic, convertible (clickable) modification to our initiator nucleotide (Fig. 4.1). Although the present work is mainly focused on initiators which allow transcription initiation from bacteriophage class-III promoters, our ongoing research also involves dinucleotides capable of transcription initiation from bacteriophage class-II promoters and bearing carrier nucleoside where the convertible (clickable) residue is connected through a different attachment site.

Few previous studies(239-242) reported successful statistical incorporation of certain 5-functionalized fluorescent uridine triphosphates into RNA by *in vitro* run-off transcription with T7 RNAP and 5-ethynynluridine has been used(186) in mammalian cell culture studies for labeling RNA in fixed cells. From these results two facts can be formulated, i) RNAPs are tolerant towards 5-modified uridines and ii) the sterically little demanding ethynyl moiety when present at the C-5 position of uridine, induces such a little structural change compared to its unmodified analogue uridine that it not only gets accepted as substrate by eukaryotic kinasing machineries and converted to the corresponding triphosphates but also by the

corresponding RNA polymerases and incorporated statistically into RNA *in vivo*. Thus to our rationale, dinucleotides which are modified at the C-5 position of the carrier nucleoside might be a good starting point for our modular synthetic approach to construct initiator dinucleotides. Furthermore, this design might be more suitable for ligation since the modification is not situated very close to the ligation joint.

Synthetically, 3´-phosphoramidites of ribo and deoxyribonucleosides are the most frequently applied building units for chemical RNA and DNA synthesis. Of all four standard nucleosides in RNA, the construction of derivatized uridine phosphoramidites is typically associated with the most efficient synthetic routs, since protective groups on the nucleobase are not required. Furthermore, since previous studies with di-, tri- or short oligo(deoxy)nucleotides for transcription initiation by T7 RNAP revealed highly efficient transcriptional priming and improved transcription yield(236) with these oligo(deoxy)nucleotides, we ideated that the presence of the 2´-OH functional group on the carrier nucleoside might not be mandatory for an efficient transcription initiation. This has led us to design the deoxy-based initiator dinucleotides which also have certain synthetic advantages over their ribo-based counterparts namely i) fewer number of steps involved in the synthesis of deoxy-based phosphoramidites compared to their ribo- analog and ii) higher coupling efficiency of deoxy-phosphoramidites due to the lack of a bulky 2´-O-protective group as also regularly observed in solid phase synthesis of RNA compared to DNA. Consequently, we decided for a modular approach to combine three different alkyne-bearing 5-modified (deoxy)-uridine phosphoramidites with a universal triisobutyryl-guanosine (iBu_3G) unit.

The three different dinucleotide derivatives were then tested for their acceptance by different phage RNA polymerases.

4.1.4 EUpG priming by T3, SP6 and T7 RNA polymerase

First, ethynyl-functionalized ribodinucleotide EUpG was assayed. The dinucleotide was added to the reaction mixtures containing all four ribonucleoside triphosphates (NTPs), a double stranded synthetic DNA template and an RNA polymerase (RNAP) in a suitable buffer. The resulting transcript has a length of 25 nucleotides. As the dinucleotide competes at the initiation step with GTP, the GTP/initiator ratio is an important parameter. A too low GTP concentration, however, will have negative effect on the total transcription yield, since GTP is required during transcript elongation too. Therefore, a two-step optimization was carried out, in which first the GTP/initiator ratio was varied (Fig. 4.2, panel A, C, E) and then the concentrations of the other three NTPs (Fig. 4.2, panel B, D, F) were screened.

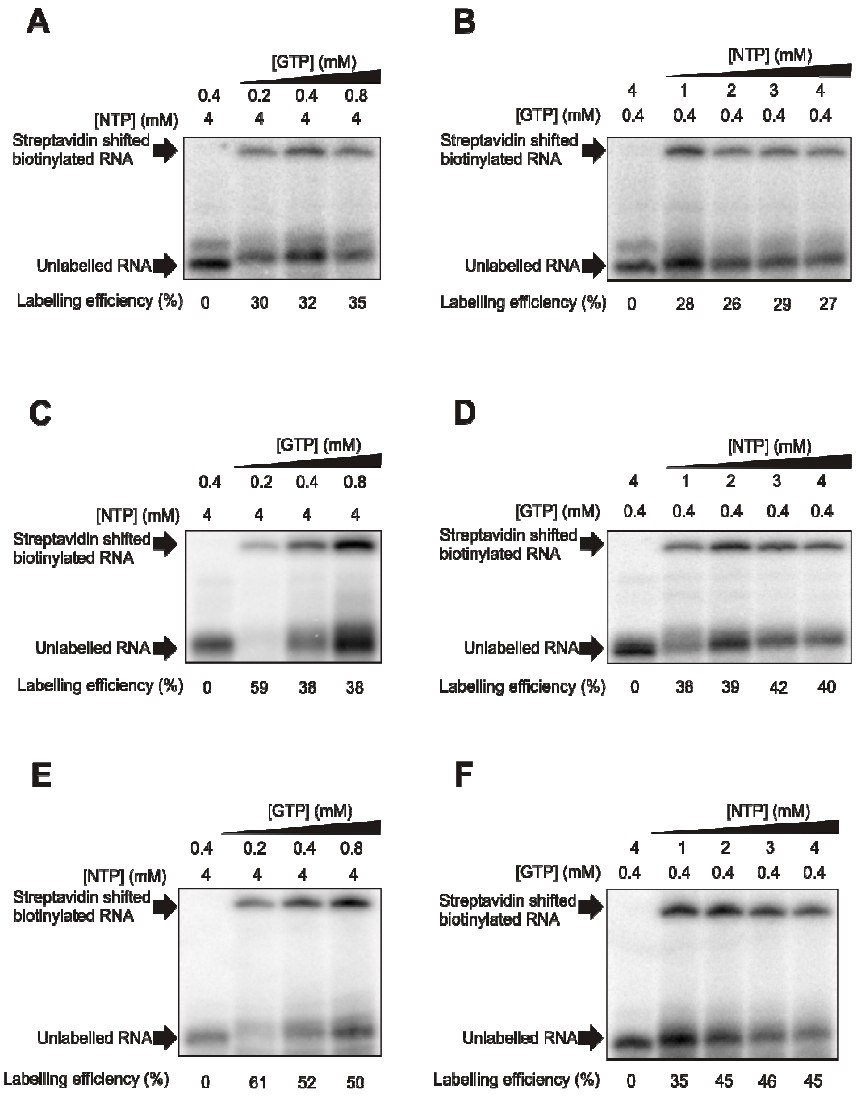


Figure 4.2: Optimization of transcriptional labeling with EUpG by T3 (panel A, B), SP6 (panel C, D) and T7 (panel E, F) RNA polymerases. The extreme left lane in every gel denotes transcription initiation with GMP.

Since the incorporation of the dinucleotide itself induced only a slight shift in the electrophoretic mobility (Fig. 4.2), the appended ethynyl moiety was subsequently employed to quantitatively introduce a biotinyl residue by copper-catalyzed azide-alkyne cycloaddition (CuAAC) using biotin azide. In addition to assessing the general utility of the approach for click-type modification of RNA, the incorporation of biotin allowed for an unambiguous assessment of initiation efficiency and transcript quantification by a Streptavidin electrophoretic mobility shift assay (Strep-EMSA). The results shown in figure 4.2 demonstrate that the initiator nucleotide EUpG is indeed incorporated by all three polymerases, although with different efficiencies. The labelling efficiencies and the corresponding transcription reaction conditions are shown in figure 4.2. A control experiment was introduced where GMP, a well characterized initiator mononucleotide at least with

respect to T7 RNAP, was exchanged against the modified dinucleotide keeping all other transcription conditions the same to directly asses the effect of our synthesized molecule on enzymatic transcriptional activity. Overall transcription yields are found to be comparable.

4.1.5 EdUpG priming by T3, SP6 and T7 RNA polymerase

Encouraged by the fact that a wide number of initiator mononucleotides, where the non-nucleosidic modifications are attached through the 5´-terminal phosphate of a guanosine, have been successfully used to functionalize RNA by transcription priming approach and previous studies(236) indicated highly efficient transcription priming with unmodified dinucleotides lacking a 2´-hydroxyl group on the 5´-nucleotide, we next assayed our deoxy-based initiator dinucleotide EdUpG. Strep-EMSA was used to calculate the labeling efficiencies as before and the results are summarized in figure 4.3. In general, compared to EUpG, we observed a slightly poorer labelling with EdUpG in case of all three polymerases, which might be attributable to the lack of the 2´-OH functionality on the carrier nucleoside of our dinucleotide initiator. However under optimized concentrations of all other three NTPs we observed similar labelling efficiencies in comparison to EUpG. Therefore we concluded that the presence of 2´-OH is not an absolute requirement. This observation led us to design yet another dinucleotide initiator, namely OdUpG (Fig. 4.1), containing a deoxy-nucleotide at its 5´-end and bearing a rather extended alkyne moiety which should not only facilitate the CuAAC reaction(103,232) but also helps making this dinucleotide initiator more non-polar, thereby improving its enzymatic incorporation.

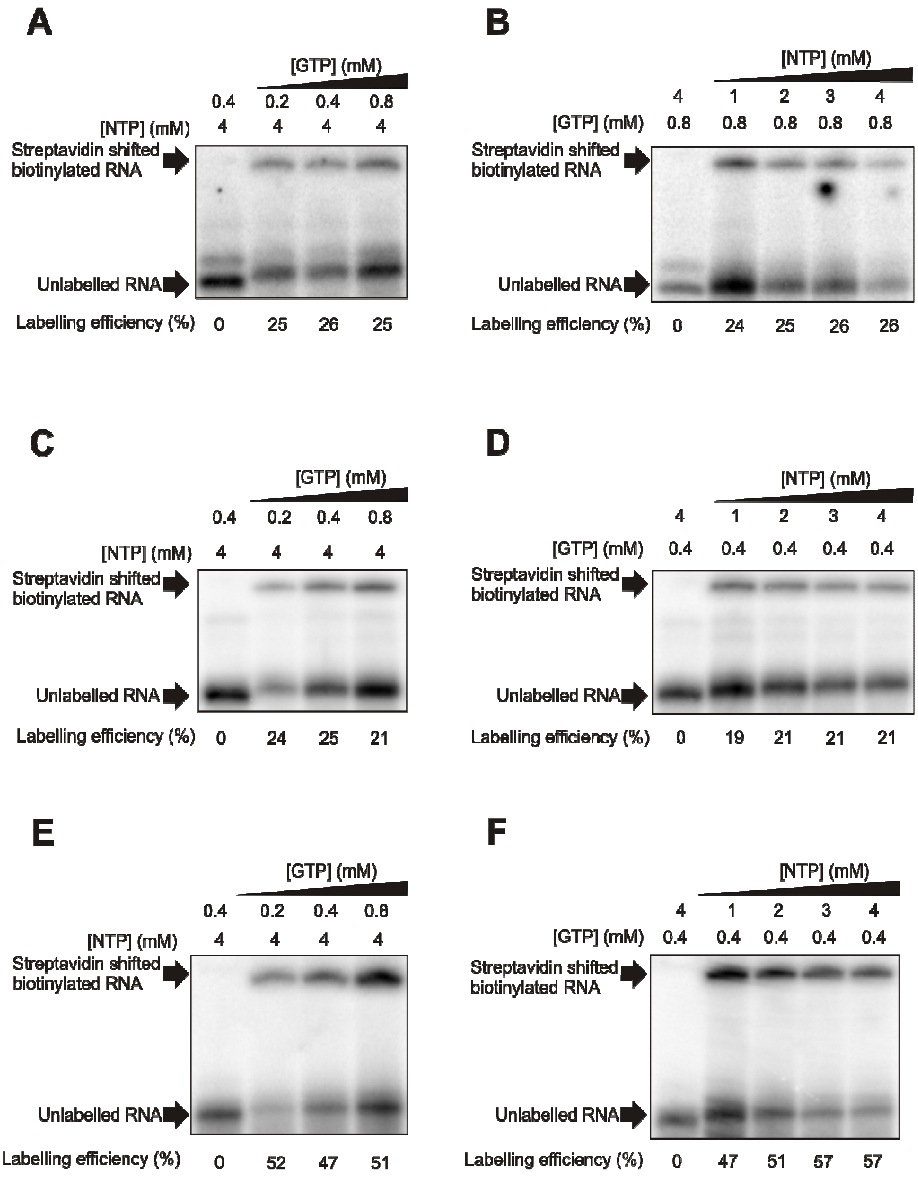


Figure 4.3: Optimization of transcriptional labeling with EdUpG by T3 (panel A, B), SP6 (panel C, D) and T7 (panel E, F) RNA polymerases. The extreme left lane in every gel denotes transcription initiation with GMP.

4.1.6 OdUpG priming by T3, SP6 and T7 RNA polymerase

A similar set of experiments has been performed with our newly designed, less polar dinucleotide initiator OdUpG. The longer alkyne handle should facilitate the follow-up click reaction as described before as well as will make this initiator less polar, which might be beneficial for its preference during transcription initiation by RNA polymerases. The labelling efficiencies are described in figure 4.4. To our conclusion, this dinucleotide initiator leads to the most efficient transcript labelling compared to the other two previously described initiators for all three polymerases. The best labelling conditions are 0.2 mM GTP, 1 mM NTP and 4

mM OdUpG for T3 RNAP with 72 % labelling; 0.8 mM GTP, 4 mM NTP and 4 mM OdUpG for SP6 RNAP with 82 % labelling; and 0.8 mM GTP, 1 mM NTP and 4 mM OdUpG for T7 RNAP with 80 % transcript labelling.

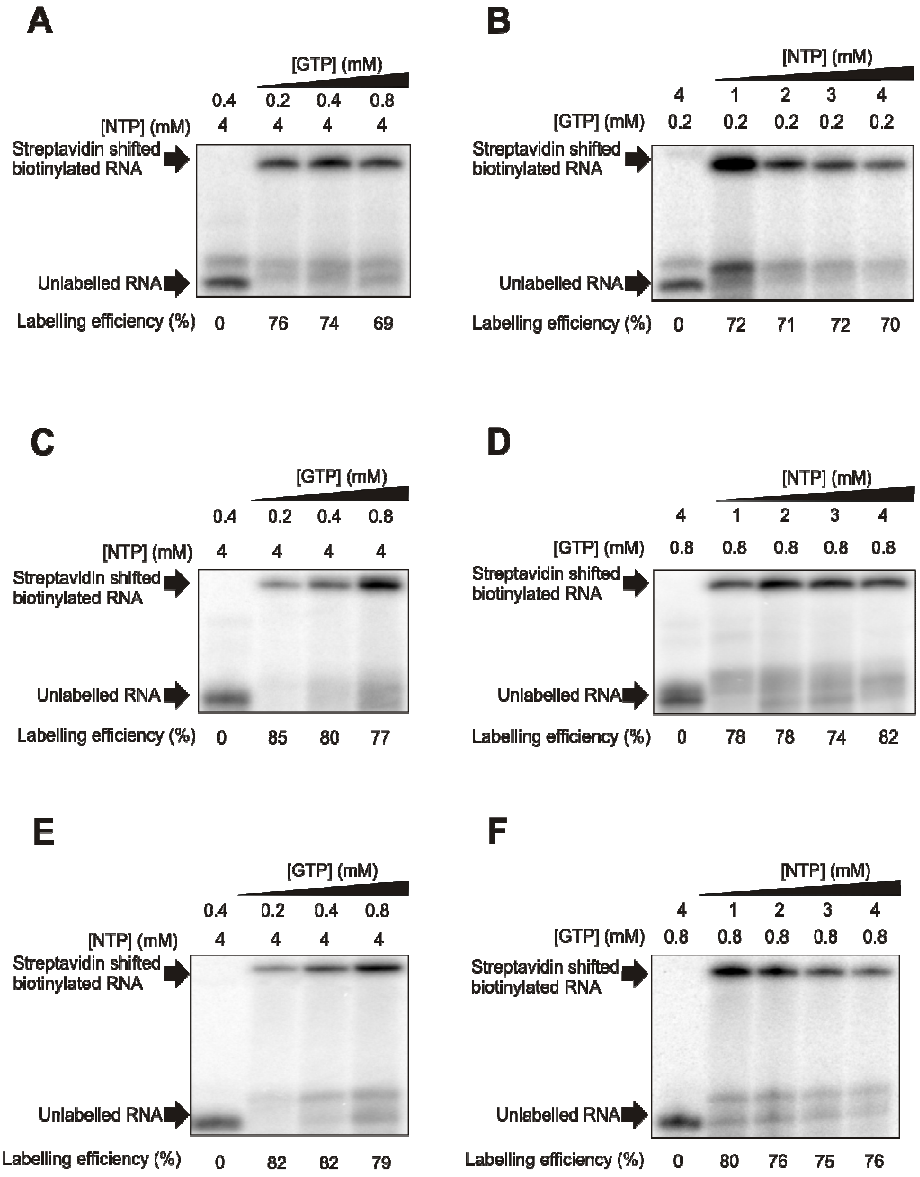


Figure 4.4: Optimization of transcriptional labeling with OdUpG by T3 (panel A, B), SP6 (panel C, D) and T7 (panel E, F) RNA polymerases. The extreme left lane in every gel denotes transcription initiation with GMP.

Although in case of transcriptional labeling with EUpG or EdUpG it was not possible to distinguish our dinucleotide initiated transcripts prior to CuAAC with biotin azide from the unlabelled n+1 transcript due to the very small size of the ethynyl moiety, this is not the case for transcriptional labeling with OdUpG by T7 RNAP. The longer alkyne chain not only facilitates the CuAAC because of sterical reason(103-105,232) but also makes the initiated transcript less polar compared to the unmodified n+1 transcript leading to a clear separation of these species on gel (Fig. 4.5). However, this was not possible in case of transcriptional

labeling by SP6 or T3 RNAPs. We observed a higher sequence heterogeneity in the transcripts produced by these polymerases which has also been hinted in a few previous studies(203,210).

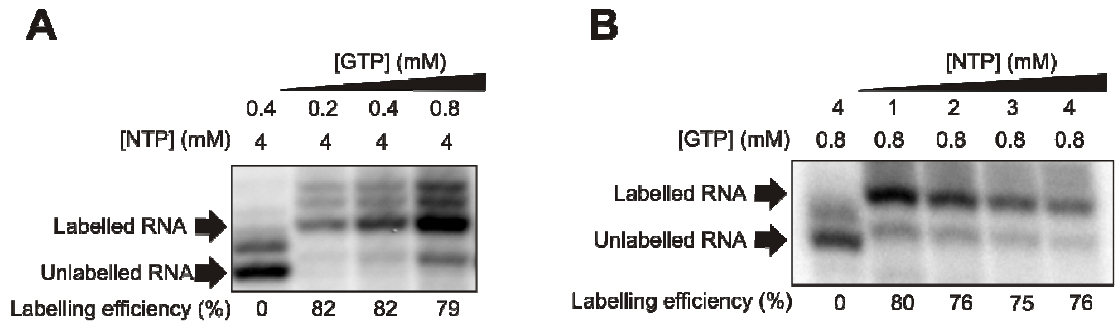


Figure 4.5: Sequencing gel analysis of transcriptional labeling optimizations with OdUpG by T7 RNA polymerase. The extreme left lane in every gel denotes transcription initiation with GMP.

4.1.7 Derivatization and enzymatic manipulation of primed RNA transcripts

Sine we obtained most efficient transcript labelling with OdUpG and our laboratory has substantial expertise in RNA functionalization by T7 RNAP(124,216,217,222,228,229), all following experiments have been performed with this combination. To investigate the robustness of our labelling approach we have synthesized the 209 nucleotide long complete glycine-responsive tandem riboswitch aptamer from *Bacillus subtilis* by *in vitro* run-off transcription and functionalized the transcript at its 5´-end with a fluorophore (Fig. 4.6). As previously mentioned, due to the degenerative nature of chemical oligonucleotide synthesis (Section 1.4.1), RNA sequences of this length carrying a modification at its 5´-end are impossible to prepare by any existing chemical RNA synthesis protocol to date. Furthermore, using this strategy we have functionalized a 233 nucleotide long random pool of RNA from a 252 nucleotide PCR amplicon with a fluorophore (Fig. 4.6). This also demonstrates the sequence independency and generality of our approach for labelling large functional RNAs.

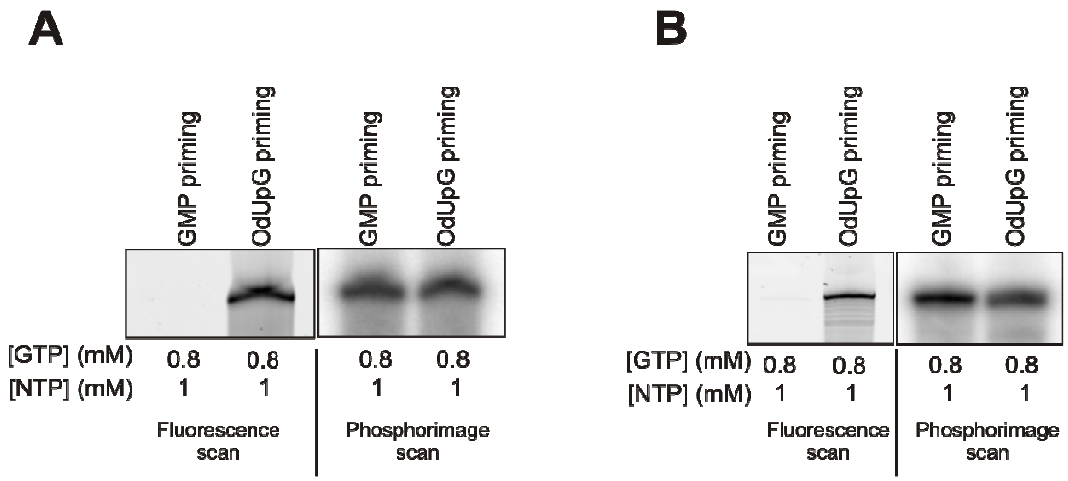


Figure 4.6: Transcriptional labeling with OdUpG by T7 RNA polymerase for A) random RNA pool and B) tandem glycine riboswitch aptamer. A Cy5 azide was used for CuAAC conjugation.

To further demonstrate the utility of our dinucleotide initiator for preparing a variety of RNA conjugates, we prepared a larger batch of a transcript using OdUpG, T7 RNAP, and the optimized conditions describe above. To reduce transcriptional 3´-end heterogeneity, a template with two 2´-OMe substitutions at the 5´-end of the antisense strand was used for these experiments(243,244). The primed transcript was then reacted in CuAAC with a variety of commercial and self-synthesized compounds, ranging from various different affinity handles, sensitive fluorophores with widely different spectral properties, polyethylene glycol moieties carrying other functional groups at one end, protected and fully unprotected carbohydrate molecules to organometallic complexes. Attachment of these various tags resulted in a highly differential electrophoretic mobility of the products compared to the starting material thereby enabling us an unambiguous assessment of the click conjugation efficiency, and therefore utility of this approach for functionalizing RNAs (Fig. 4.7).

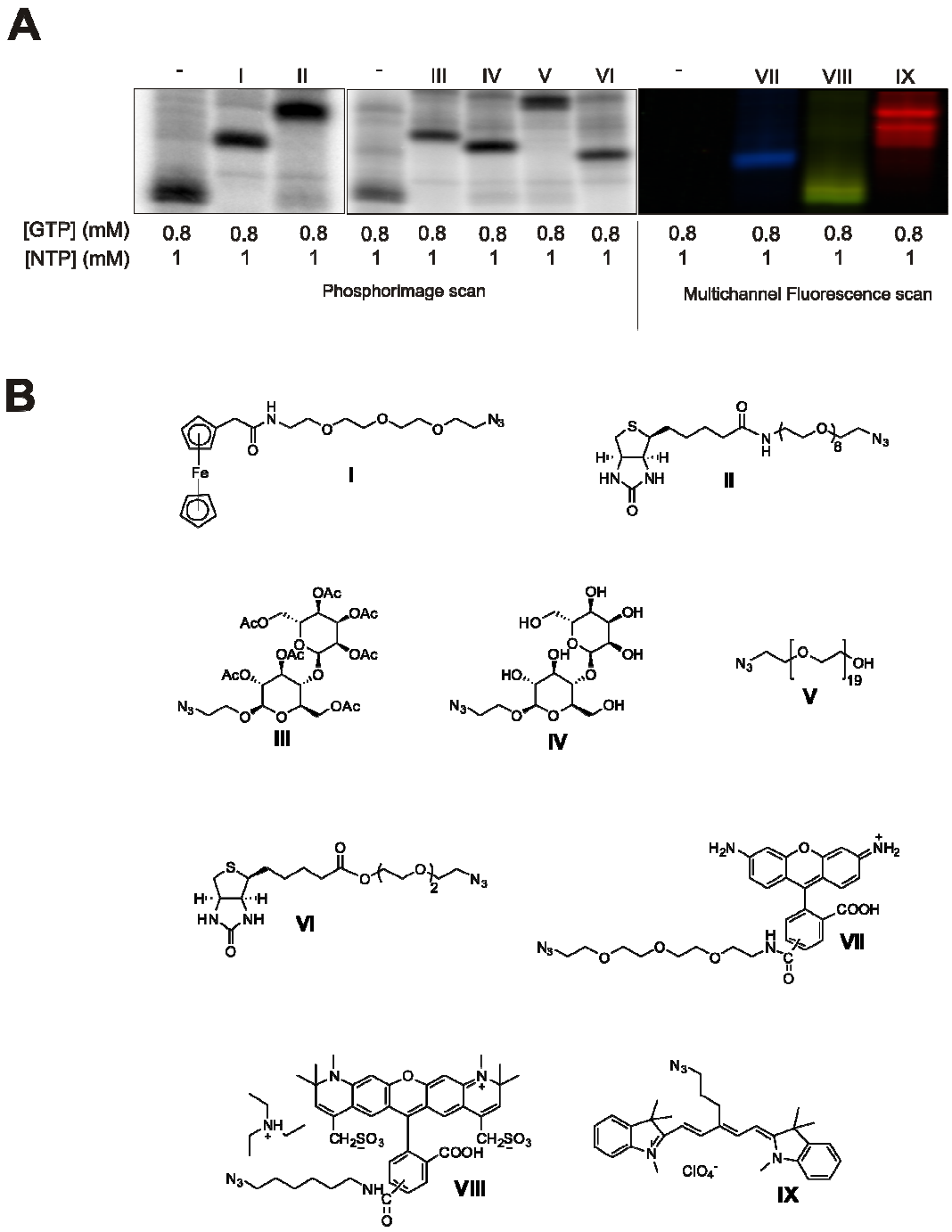


Figure 4.7: A) Gel analysis of the CuAAC conjugation efficiencies for OdUpG primed transcripts with various functional tags. B) Chemical structures of various different compounds used for this purpose.

These data indicate near-quantitative conversion in all cases, irrespective of the nature of the coupling partner. This represents a major advantage in comparison to all other previously described initiator, as one central intermediate is sufficient to synthesize a wide variety of different conjugates without the need to synthesize dozens of initiator nucleotides and to optimize their incorporation.

Another significant advantage of the building blocks described here is their free 5´-OH group, which renders them amenable to further enzymatic manipulations, most notably to ligations. This feature will allow the conversion of a terminal modification into a site-specific internal one which further broadens the scope of our approach for click-type internal modification of RNA. Internal modification could be particularly useful, e.g. in the preparation of dye-labelled long RNAs for intermolecular as well as intramolecular biophysical studies involving RNA.

However our primed RNAs bear a 5´-OH group which has to be converted to the corresponding phosphate moiety prior to ligation. The presence of many 2´-OH functional groups in the transcript and the general instability of RNA towards various chemical reagents forbear us from using chemical phosphorylation reagents for kinasing the 5´-end of the transcript. However this problem can be solved in two different ways, i) the phosphorylation could be achieved at the dinucleotide level, however this is extremely challenging (unless protecting group chemistry has been used which exponentially increases the labour involved in the synthesis of the dinucleotide and finally leads to a poor overall yield) because of the presence of other hydroxyl functional groups at the 3´-guanosine residue of our dinucleotide initiators. Accidental phosphorylation of any of these hydroxyl group will lead to the complete loss of enzymatic activity on these substrates; ii) in contrast, this phosphorylation can be achieved by polynucleotide kinase (PNK). However, this requires the modified 5´-end to be a substrate of PNK. Encouraged by the fact that T4 PNK was used in a few previous studies(245-248) to phosphorylate non-nucleosidic moieties, we tested the feasibility of PNK mediated phosphorylation followed by the ligation of a 25 nucleotide transcript modified at its 5´-end with OdUpG to another 40 nucleotide long RNA fragment. We indeed observed efficient ligation using a mixture of T4 Rnl2, T4 Rnl1 and T4 Dnl (Fig. 4.8). A side-by-side ligation comparison with completely unmodified UpG primed transcript let us to conclude that our dinucleotide initiator OdUpG is indeed tolerated by three different sets of enzymes, namely RNA polymerases, polynucleotide kinase and ligases. To our conclusion, this is possible only due to the very small size of the CuAAC reaction partners. The resulting ligated RNA can further be derivatized with a fluorescent dye or affinity handle by CuAAC, therefore enabling site-specific internal modification, which remarkably broadens the scope of our universal-initiator based RNA functionalization approach. It should be noted that the 3´-end of the ligated product is unmodified and could be easily functionalized with a third modification by a nucleotidyl transferase mediated approach as described in section 4.2.

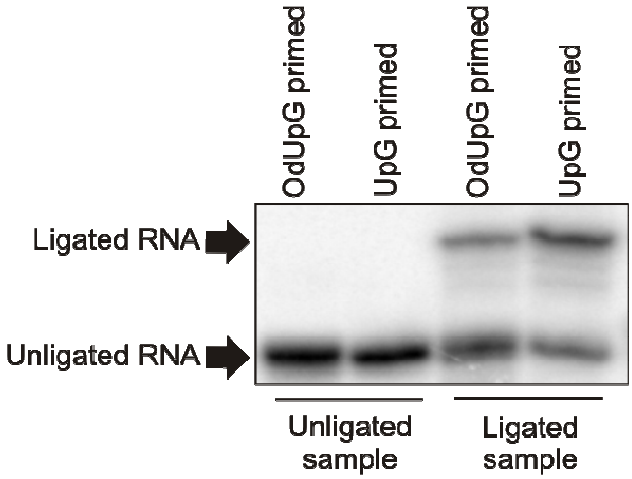


Figure 4.8: Gel analysis for splinted-ligation of OdUpG primed transcripts.

4.1.8 Discussion

The demand for millimolar concentrations of dinucleotide initiators in the transcription mixture promoted us to forbear from a solid-phase approach, as we expected insufficient yield and wasteful consumption of precious phosphoramidites by using classical solid-phase synthetic oligonucleotide protocols. Instead, a solution-phase strategy was established that was adapted to the advantages of classical phosphoramidite chemistry on solid support. Particularly, the high excess of reagents in each synthetic step and extensive washing protocols had to be compensated by the design of the synthetic strategy. Generally, all reagents were applied in stoichiometric amounts, thus reducing the complexity of the final purification. Reagents, solvents and protective groups were chosen to be separable from the target molecule by fast and easy procedures, such as evaporation, lyophilization, liquid-liquid-extraction or precipitation.

Examination of the initiator dinucleotides **1**, **2** and **3** (Fig. 4.1) for their incorporation efficiencies indicated that both 5-ethynyl and 5-octadiynyl substitutions are well tolerated by all three different phage polymerases.

A slight preference of EUpG over EdUpG for all three polymerases during transcriptional labelling can be explained due to the lack of 2´-OH functionality at the 5´-nucleoside for the latter one. However chemical synthesis of dinucleotides involving a 5´-deoxy-nucleotide is less demanding compared to their ribo-analogues. Moreover, the long flexible alkyne moiety on OdUpG not only makes this initiator molecule less polar compared to EUpG or EdUpG but also facilitates the follow up CuAAC reaction compared to the rigid ethynyl moiety as has been shown in previous studies involving DNA labelling(103,232). Comparative studies on initiator mononucleotides in our lab as well as in others indicated an

increase in incorporation yields with decreasing polarity of the 5´-attached moiety(217), which is in complete accordance to the observation of very high labelling efficiencies by all three polymerases with relatively non-polar initiator OdUpG.

Although T7 RNAP has been widely used for initiator-based RNA 5´-end functionalization, we are not aware of similar approaches using SP6 and T3 RNAP. Given the generality of transcription initiation mechanism by phage polymerases we assumed that SP6 and T3 RNAP can also be used for the same purpose. It is noteworthy here that the transcriptional activity of these polymerases *in vitro* differ from each other depending on the transcript length, NTP and magnesium ion concentrations, salt compositions in the buffer and temperature(237). Therefore it is beneficial to establish a labelling protocol which is compatible with all three different polymerases.

The initiated transcripts have further been functionalized with a diverse array of organic moieties by CuAAC. In this way it is possible to functionalize the target RNA at its 5´-end with a diverse range of biologically relevant functional tags without the need for *de novo* synthesis of new initiator nucleotides or optimization of enzymatic incorporation for each new compound. Therefore these dinucleotides can be regarded as ``universal initiator dinucleotides``.

Finally RNA transcripts primed with our dinucleotide initiator have been subjected to enzymatic phosphorylation followed by splinted ligation to a second RNA strand, thereby converting a terminal modification to an internal one which further broadens the scope of our approach for click-type site-specific, internal modification of RNA.

In a summary, the highly efficient transcriptional labelling with these dinucleotides, their generality for attachment of a wide array of diverse chemical functionalities by CuAAC, their usability for functionalization of long RNAs and finally the compatibility of these modified transcripts with enzymatic ligations surely allowed us to conclude that this is an unprecedented advancement for initiator-based approaches in RNA labelling.

4.2 Chemo-enzymatic labeling of the 3´-end of RNA

4.2.1 Scientific background

After successfully employing our two-step, click chemistry based RNA labeling approach for labeling the RNA 5´-end we sought to expand this strategy towards labeling the 3´-end of RNA as well. It is worth mentioning that for various natural RNAs as well as for *in vitro* transcribed RNAs the 3´-end is more accessible compared to the 5´-end. For example, *in vitro* transcribed RNA (unless otherwise it is primed during transcription initiation with an

initiator nucleotide) contains a triphosphates residue (203-206), similarly, eukaryotic and viral mRNAs are capped at their 5´-end (194) and therefore blocked for further functionalization, ribosomal RNA and various non-coding small RNAs often contain a monophosphate residue at their 5´-end. Moreover our previous strategy to label the RNA 5´-end by transcriptional priming requires access to the corresponding transcription template of the RNA and therefore needs sequence information. Similarly other co-synthetic methods are not applicable for RNA originating from biological sources, e.g., from specific tissues of certain organisms. All co-synthetic chemical or chemo-enzymatic methods strictly require sequence information and in most cases these techniques are applicable only for *de novo* labeling of RNA. Additionally co-synthetic chemical methods (e.g. solid-phase RNA synthesis using phosphoramidite chemistry) are strictly limited towards modifying only short sequences.

Among various currently known post-synthetic strategies for non-*de novo* labeling of the RNA 3´-end the following two are of notable interest.

Ligation-based approach for labeling the RNA 3´-end

This method exploits the relaxed substrate tolerance of T4 RNA ligase 1 (T4 Rnl1) and makes use of suitably modified base analogs (Fig. 4.9)(249,250). This technique has mostly been used for labeling with modified cytidines since cytidine is preferred on the 3´-fragment at the ligation site by T4 Rnl1(250,251). Also in the majority of cases the modification has been attached through either the C-5 or the C-8 of pyrimidines or purines, respectively, to achieve efficient ligation(249,252-256). In this approach, the presence of a 3´-phosphate residue at the modified nucleotide is mandatory to ensure single label incorporation to the target RNA since otherwise the modified end can also be recognized by the ligase as substrate and can therefore be further modified. Phosphorylation of 3´-OH of ribo or deoxyribonucleotides (modified or unmodified) can only be achieved in a chemical manner, since there are no enzymes available which can perform a direct phosphorylation of the 3´-OH of a ribo or deoxyribonucleotide. However, such chemical phosphorylation in presence of other nucleophilic functional groups is highly challenging, which poses an additional barrier for the synthesis of such modified nucleotides. Moreover, RNA modified with this approach needs to be enzymatically dephosphorylated first and then only that modified RNA 3´-end can take part in further enzymatic manipulations.

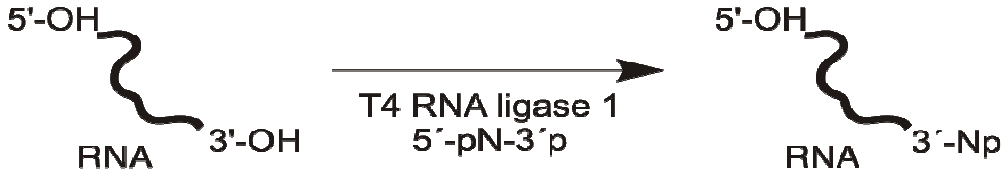


Figure 4.9: Schematic representation of ligation-based approach for labeling the RNA 3´-end. N denotes modified deoxy or ribonucleotide.

Noteworthy to mention here, that this strategy strictly requires the absence of a phosphate group on the 5´-end of the RNA to prevent circularization of the starting material.

Oxidation of 3´-terminal *vicinal* diol for RNA 3´-end labeling

This approach hires the well known reaction for converting *vicinal* diol to *vicinal* dialdehyde (Fig. 4.10) from the field of organic synthesis. Since the only *vicinal* diol present in an RNA molecule is at its 3´-end, this can be selectively oxidized to *vicinal* dialdehyde using controlled amounts of periodic acid (257). Generally, this modified end is conjugated to the required functional group using Schiff´s-base chemistry in a follow-up reaction. However, the harsh reaction conditions involved in this labeling strategy might not be suitable for certain RNA sequences and additionally the 3´-end of the RNA after modification can not further be manipulated by any enzymatic reactions since it does not bear any structural similarities with natural RNA 3´-ends.

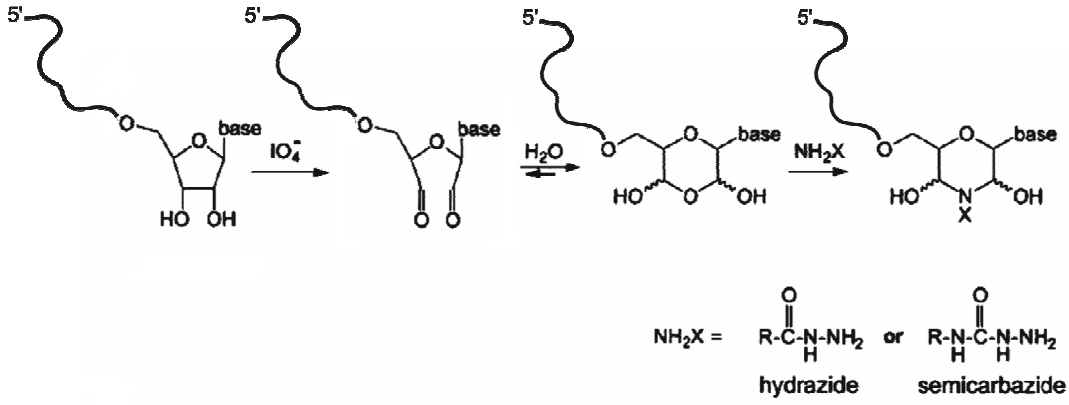


Figure 4.10: Schematic representation of periodate-based approach for labeling the RNA 3´-end. R denotes the required modification.

4.2.2 3´-labeling of RNA by nucleotidyl transferases and copper catalyzed and copper free click chemistry – current work[6]

Scientific background

Therefore we sought to develop a 3´-end labeling technique which will allow us for post-synthetic non-*de novo* labeling of RNA, either of synthetic or biological origin, without any prior knowledge of its sequence. As mentioned earlier in section 1.4 the use of a two-step click chemistry based approach has many advantages over one-step direct labeling and in case of certain bulky modification a two-step approach is rather indispensable – therefore we extended the same principle for labeling the 3´-end of the RNA as well. We decided to introduce the clickable modification to the RNA first and then to attach the required functional group by CuAAC (or with other metal-free click reactions as discussed later), as with our previous labeling approach based on initiator nucleotides. Enzymatic incorporation of a convertible residue to the RNA 3´-terminus might be advantageous in comparison to chemical methods since enzymatic reactions generally involve milder conditions compared to chemical methods. We envisioned that introducing an azide rather than alkyne is beneficial since azides can not only take part in CuAAC but also in various other metal-free click reactions, most notably Strain Promoted Azide Alkyne Cycloaddition (SPAAC)(111) and Bertozzi-Staudinger Ligation(94), therefore broadening the range of various different labels which can be incorporated beyond the scope of only one type of bioorthogonal click reaction. Additionally metal-free click reactions are more suitable for *in vivo* applications.

Here, in our two-step RNA labeling approach we exploit poly(A) polymerases (PAP) to incorporate azide modified nucleotides to the 3´-end of the RNA of interest followed by derivatization of that azide residue with various bio-orthogonal click reactions (Fig. 4.11). We tested four different nucleotidyl transferases [yeast and *E. coli* PAP, Cid 1 poly(U) polymerase (PUP) and terminal deoxynucleotidyl transferase (TdT)] in combination with seven different azido-functionalized NTPs to find out the best enzyme-substrate pair for RNA 3´-end labeling. The advantage of incorporating azide over alkyne has been exemplarily demonstrated by using CuAAC as well as SPAAC to attach sensitive fluorophores to the enzymatically modified RNA. Using our labeling toolbox we could attach a wide range of biologically relevant functional tags to the RNA of interest for all four possible nucleotides (A, C, G, and U), at a wide range of concentrations.

[6] This part of the work has been performed together with M.-L. Winz.

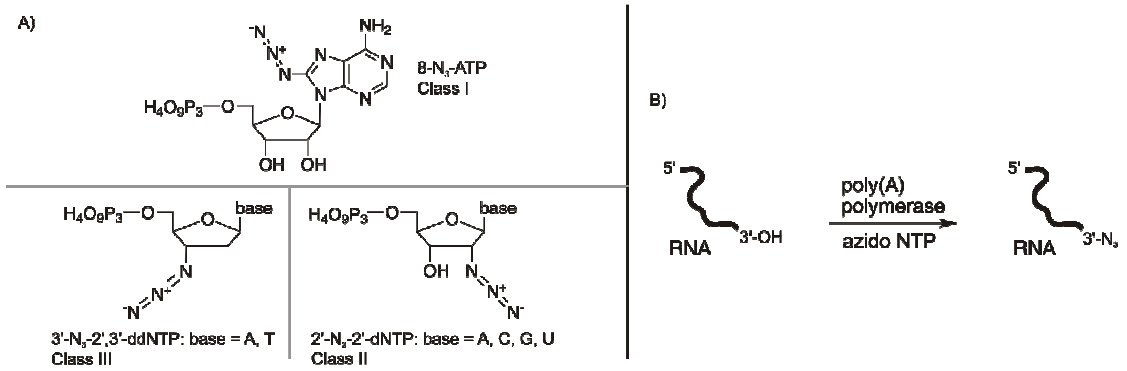


Figure 4.11: A) Different azido-modified NTPs used in this study, **B)** schematic representation of our nucleotidyl transferase-based approach for labeling the RNA 3´-end.

Moreover, after introduction of nucleotides modified at the position C-8 (class-I) or C-2´ (class-II), the resulting 3´-termini contain free hydroxyl groups. Those RNAs could be subjected to enzymatic manipulations involving their modified 3´-termini, namely, addition of poly(A)-tails, 3´-adapter ligation and splinted ligation to other RNA sequences without the need for any prior manipulations of that modified end (contrary to ligation-based approaches where enzymatic dephosphorylation has to be performed first). Doing so, we converted the terminal azide-modifications into internal ones, for which we optimized CuAAC reaction conditions, to achieve site-specific, internal functionalization of RNA. Our approach does not demand the *de novo* synthesis of modified chemical entities and involves only commercially available enzymes and reagents – therefore enzymatic reaction conditions have to be optimized only once and those enzymatically modified RNAs can further be derivatized with the functional group of interest without having to synthesize a new modified nucleotide each time. Due to the commercial availability of an ever growing number of different possible labels that can be introduced by click chemistry, this protocol for site-specific RNA labeling is more accessible for non-chemist users.

4.2.3 Results

Screening of azide modified nucleotides and different nucleotidyl transferases for most efficient enzyme-substrate pair

We have tested all together seven different azide modified nucleotides in combination with four different nucleotidyl transferases to obtain the most efficient enzyme-substrate pair for our labeling purpose. These modified nucleotides can be divided into three classes depending on the modification site (Fig. 4.11). These substrates will allow base as well as backbone functionalization of the target RNA sequences. While the class I and class II substrates will allow further enzymatic manipulations of the 3´-end of the RNA after

modification reaction, the class III substrates, owing to the lack of a 3´-hydroxy function, will not allow any further enzymatic reactions of the modified RNA 3´-end. However, screening of the complete structural spectrum of all different nucleotides against all four different nucleotidyl transferases will facilitate our understanding of the enzyme-substrate interaction, which, in turn is beneficial for designing new modified substrates in future. The four different nucleotidyl transferases tested are yeast and *E. coli* PAP, Cid 1 PUP and TdT. While the choice of yeast and *E. coli* PAP and Cid 1 PUP is obvious, since these enzymes are responsible for non-templated nucleotide addition (A and U respectively) to the RNA 3´-ends in biological systems, the reason for evaluating TdT for the same purpose is not apparently clear. TdT is responsible for non-templated deoxyribonucleotide addition to the 3´-end of DNA. However it also accepts ribonucleotides but generates only very short tails. Since TdT has been used in a previous study to modify the 3´-end of RNA with digoxigenin- and biotin-modified nucleotides (258) and our class III substrates (Fig. 4.11) bear high structural similarities with deoxyribonucleotides, we decided to use this enzyme as well for modifying the RNA 3´-end.

The results of screening are summarized in table 4.1. The efficiency of TdT for modified nucleotide addition on RNA 3´-end was found to be negligible. Among the other three different nucleotidyl transferases yeast PAP exhibited the highest incorporation activity with all different kinds of modified NTPs while Cid 1 PUP accepted 2´-N_3-2´-dNTPs in moderate yield and *E. coli* PAP was able to incorporate only 2´-azido-2´-dATP. The ability of yeast PAP to incorporate all four class II substrates makes our labeling protocol amenable towards modifying RNA 3´-ends with any one of the four different nucleotides of choice, therefore causing no sequence mutation to the RNA after the modification reaction which will help preserving its function.

Table 4.1: Results of the first screening of nucleotidyl transferases for incorporation of N_3-nucleotides.

Modified NTP	# of residues added			
	yeast PAP	*E. coli* PAP	Cid 1 PUP	TdT
8-N_3-ATP	1-2	-	0-2*	-
2'-N_3-2'-dATP	multiple	1(-2)**	0-mult.*	-
2'-N_3-2'-dCTP	1-2	-	0-mult.*	-
2'-N_3-2'-dGTP	1-2	-	0-mult.*	0-1*
2'-N_3-2'-dUTP	1-2	-	0-mult.*	-
3'-N_3-2',3'-ddATP	0-1	-	-	n.d.
3'-N_3-2',3'-ddTTP	0-1	-	-	n.d.
* mostly 0; ** only at long incorporation times or with $MnCl_2$				

With 8-N_3-ATP we mostly observed multiple nucleotide incorporation. We envisioned that, as a labeling strategy, it would be beneficial to introduce a single label per biomolecule, therefore reaction conditions leading to single nucleotide incorporation by the nucleotidyl

transferases need to be established. To this end, we focused only on yeast and *E. coli* PAP together with only sugar modified NTPs for single residue incorporation.

Optimization of single-nucleotide incorporation by yeast and *E. coli* PAP

As mentioned earlier, the RNA 3´-end, after modification with class III substrates, cannot further be manipulated by any other enzymes – thereby limiting our strategy strictly towards only 3´-end modification. For that reason, we focused mainly on single incorporation of 2´-azido nucleotides. Having observed the relatively short tails generated by yeast PAP with 2´-azido-2´-dNTPs we envisioned that it might be possible to control this reaction towards single label incorporation. Under optimized reaction conditions we indeed observed single label incorporation with yeast PAP (except for 2´-N_3-2´-dATP) as summarized in table 4.2.

Table 4.2: Conditions that lead to single nucleotide addition.

Modified NTP	PAP from	Reaction time
2'-N_3-2'-dATP	*E. coli* (0.5 U/µl)	16 h (- $MnCl_2$)
	E. coli (0.25 U/µl)	20 min (+ $MnCl_2$)*
2'-N_3-2'-dCTP	yeast	20 min
2'-N_3-2'-dGTP		2 h
2'-N_3-2'-dUTP		5 min
3'-N_3-2',3'-ddATP		(not optimized)
3'-N_3-2',3'-ddTTP		
8-N_3-ATP	not possible	
* preferred conditions (2.5 mM $MnCl_2$)		

Since the use of yeast PAP with 2´-N_3-2´-dATP led to the addition of multiple residues even at low NTP concentration and short reaction time, we thought *E. coli* PAP might be useful for single label incorporation of this particular modified nucleotide since *E. coli* PAP generally has a lower incorporation activity compared to yeast PAP. Acceptable incorporation yield could be observed only after 16 hours of incubation at 37^0C which also led to severe RNA degradation. Addition of $MnCl_2$ to the reaction mixture to an end concentration of 2.5 mM led to a labeling yield of >50 % after as little as 20 min reaction time. This is in accordance with previous studies with *E. coli* PAP where the enzymatic activity was reported to be increased by 5 fold in presence of manganese ions(259). Single nucleotide incorporation efficiencies for all 2´-N_3-NTPs in triplicates are shown in figure 4.12.

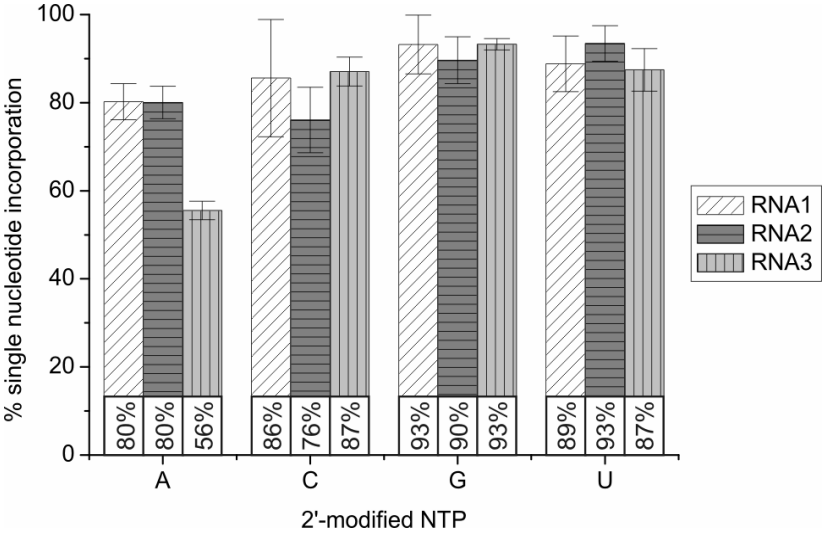


Figure 4.12: Determination of single nucleotide incorporation efficiencies for all 2´-N_3-NTPs by yeast and *E. coli* PAP.

Dynamic range of PAP reaction

Depending on the application, the amount of RNA to be modified may vary considerably. While RNAs isolated from rare biological samples (e.g. specific tissues) can be scarce, naturally abundant RNA or *in vitro* synthesized (chemically or enzymatically) RNA might need to be modified in bigger amounts and therefore in higher concentrations, demanding our labeling protocol to be equally efficient under very high as well as very low RNA concentrations. Thus, to investigate the robustness of our approach, we sought to find out the dynamic range of the yeast PAP reaction where we varied the reaction time employing any one of the four 2´-N_3-modified NTPs and RNA1 at 0.2 and 4 µM concentration, or RNA2 at 4 to 80 µM concentration in a reaction with 2´-N_3-2´-dCTP and –UTP (Tab. 4.3 for sequence information).

Table 4.3: Sequences of oligonucleotides.

Name	Sequence	Features
RNA1	5’-GUGACCGCGGAUCGACUUCACCGC GCAGUG-3’	10 nt loop, 5 nt dangling ends
RNA2	5’-GCAAGCUGACCCUGAAGUUCAU-3’	siRNA pair
RNA3	5’-GAACUUCAGGGUCAGCUUGCCG-3’	
RNA4	5’-pGCCCGACGAUGCUUCAGCAACCAG UGUAAUGGCG-3'	ligation fragment*
DNA1	5'-CGCCATTACACTGGTTGCTGAAGCAT CGTCGGGCCCACTGCGCGGTGAGTCGATC CGCGGTCAC-3'	DNA splint*
DNA2	3’-Ap-5’-5’-pCTGTAGGCACCATCAAT-3’-block	adapter: 3’-block, 5’ adenylated
DNA3	5’-CGCCATTACACTGGTTGCTGAAGCAT CGTCGCGGTGAAGTCGATCCGCGGTCAC-3’	helper DNA, assisting CuAAC
* for splinted ligation of RNA1 (+G) to RNA4		

Although the overall reaction yield was generally lower at 0.2 µM than at 4 µM concentration, we observed similar reaction kinetics. At elevated RNA concentrations, we observed that the relative rate of the reaction was decreased. Nevertheless, efficient addition of a single nucleotide was possible at prolonged reaction times even in case of very high RNA concentration (maximally 2h for 80 µM RNA with 2´-N_3-2´-dCTP). LC-MS analysis of the reaction mixture revealed *ca* 80 % labeling yield. Thus we concluded that our labeling toolbox is indeed applicable for labeling very low as well as very high amounts of RNA - thereby providing a universal RNA 3´-end labeling toolbox capable of modifying widely different amounts of RNA without implementing severe changes to the experimental set up.

CuAAC at the 3´-terminus

As mentioned earlier, the 3´-modified RNAs carrying an azido functional group can be used as versatile modules for derivatization with a diverse array of biologically relevant functional/reporter tags by the use of click chemistry. In this approach the enzymatic reaction for single modified nucleotide incorporation has to be optimized only once and the resulting RNA sequence can be stored and further derivatized at any time with a functional tag of choice without having to synthesize a new modified NTP each time or optimizing the enzymatic incorporation for every single new compound. To test the copper-catalyzed click reaction we used three different alkynes, namely biotin-alkyne, Alexa Fluor 488 alkyne or Alexa Fluor 647 alkyne. RNA modified with any one of the four 2´-N_3-2´-dNTPs was subjected to a ligand accelerated copper-catalyzed azide-alkyne cycloaddition/ligation (CuAAC). We have used a water soluble, C_{3v}-symmetric class-I ligand (for ligand classification see chapter 3), THPTA, as before. The use of a copper-coordinating ligand is mandatory to prevent degradation of the RNA by Fenton´s chemistry initiated by the copper

in aqueous media in presence of oxygen as mentioned earlier in chapter 3. The optimized reaction conditions and the reaction yields are summarized in table 4.4. Some of these data were further confirmed by LC-MS analysis.

Table 4.4: Yields and corresponding reaction conditions of CuAAC at different N_3-RNA concentrations.

[N_3-RNA]	[alkyne]	[Cu(II)]	[THPTA]	[Asc.]	Yield
50 nM	50 µM	100 µM	500 µM	1 mM	> 80%
1 - 10 µM	500 µM	100 µM	500 µM	1 mM	quantitative
50µM	2 mM	500 µM	2.5 mM	5 mM	quantitative
General conditions: 50 mM phosphate buffer, 37°C, 2 h					

Fluorescent labeling of RNA by SPAAC

As previously mentioned, incorporation of azide instead of an alkyne functionality into biomolecules has the added advantage that the incorporated azide can be used not only for copper-catalyzed but also for other, copper-free click reactions, namely strain promoted click reaction and Bertozzi-Staudinger ligation. To prove the generality of our approach we performed strain promoted click reaction for attachment of sensitive fluorescent dyes to the 3´-end of our modified RNA.

Detection limit for fluorescently labeled RNA

After incorporation of either one of the four 2´-N_3-nucleotides and conjugation with an Alexa Fluor 647 alkyne by CuAAC we were able to visualize as low as 100 amol as clear band on 12 % denaturing polyacrylamide sequencing gel. Lower amounts like 50 or 25 amol resulted in blurry but still distinguishable spots on the gel.

Fluorescent labeling of total RNA isolated from *E. coli*

To check the applicability of our labeling approach towards labeling of long RNA of biological origin we have isolated total RNA from *E. coli* and subjected it to tailing with 2´-N_3-2´-dUTP and yeast PAP followed by CuAAC functionalization with Alexa Fluor 647 alkyne. To prevent degradation of RNA sample, the CuAAC protocol had been slightly modified (30 min incubation at 25^0C instead of 2h incubation at 37^0C). In PAGE analysis the appearance of fluorescent bands that perfectly overlay with bands visible after non-specific RNA staining by SYBR Gold indicated successful functionalization of long RNA without any degradation by our RNA labeling approach.

Enzymatic manipulation of the azido-modified 3´-termini of PAP reaction products

In contrary to the class III substrates (Fig. 4.11), enzymatic modification with class I and class II substrates leaves a hydroxyl functional group at the modified RNA 3´-terminus which, in principle, can further take part in various other enzymatic reactions. To test this assumption, we investigated the feasibility of i) further polyadenylation of these modified RNAs with yeast and *E. coli* PAP, ii) carrying out an adapter ligation to attach a short DNA strand to the modified 3´-end and iii) a splint-mediated ligation in order to add a second RNA strand to the modified 3´-end. The results of polyadenylation and adapter ligation are summarized in table 4.5.

Table 4.5: Enzymatic manipulation of 3'-termini of RNA previously modified in PAP reactions.

3'-terminal modified nt	Polyadenylation* by		Adapter ligation
	Yeast PAP	*E. coli* PAP	
none	+ (~ 93%)	+ (> 95%)	+ (> 95%)***
2'-N_3-2'-dA	+ (~ 65%)	+ (> 95%)	+ (~ 75%)***
2'-N_3-2'-dC	- (< 5%)**	+ (> 95%)	+ (~ 90%)***
2'-N_3-2'-dG	- (< 5%)	+ (> 95%)	+ (~ 85%)***
2'-N_3-2'-dU	- (< 5%)	+ (> 95%)	+ (~ 80%)***
3'-N_3-2',3'-ddA	not possible (3'-terminus blocked)		
3'-N_3-2',3'-ddT			
(1x) 8-N_3-A	+ (~84-89%)	n.d.	+ (~ 50-80%)***
(2x) 8-N_3-A	- (< 5%)		(+) (~ 8-40%)***
* in presence of 500 µM ATP, 1 h, 37°C; ** a higher fraction of polyadenylated sequences was attributed to remaining, non-modified RNA; *** consumption of starting material (RNA)			

Polyadenylation was found to be inefficient with yeast PAP in case of RNA strands bearing 2´-N_3-modified nucleotides at their 3´-end whereas *E. coli* PAP can accept these modified ends quite well and catalyzed quantitative polyadenylation. RNAs bearing a single 8-N_3-A at their 3´-end could be polyadenylated in a near quantitative fashion with yeast PAP while RNAs bearing two or multiple 8-N_3-A at their 3´-termini could not be efficiently polyadenylated by the same enzyme. Therefore we concluded that the polyadenylation reaction was inhibited when multiple 8-N_3-A were present on the RNA 3´-end.

We also tested the feasibility of ligating a chemically pre-adenylated DNA adapter to the modified 3´-end of the RNA by T4 RNA ligase 1 and T4 RNA ligase 2 truncated. The RNA strands containing one 8-N_3-A were accepted as efficiently as the unmodified RNA, whereas there was a strong bias against sequences with two 8-N_3-A modifications. For 2´-N_3-nucleotide containing sequences, such a preference was not observed. In addition to the desired ligation product, we observed some concatenated and circularized RNAs all of which resulted from the 5´-radioctive monophosphate (required for visualization in phosphor-imager scan) present at the 5´-terminus of the modified RNA. We did not observe any of these side

products in case of non-labeled RNA lacking a monophosphate at the 5´-end. These observations for side product formation were indeed in accordance with previous reports in the field of adapter ligation (260).

Furthermore the possibility of joining a second RNA strand was tested by splinted ligation. This would convert the 3´-terminal azido-modification into an internal modification within a defined RNA sequence and considerably extend the scope of our approach for RNA labeling. To demonstrate the feasibility of this idea we have used two different phage ligases namely T4 DNA ligase (T4 Dnl) and T4 RNA ligase 2 (T4 Rnl2) for ligation of a 34 nucleotide RNA fragment to a 30 nucleotide long RNA modified with 2´-N_3-2´-dGTP and yeast PAP at its 3´-end. For this purpose we have designed a 64 nucleotide DNA splint (Tab. 4.3 for sequence information) encompassing both ligation fragments completely. Although both enzymes accepted this modified 3´-end for ligation, T4 Rnl2 showed slightly better ligation efficiency compared to T4 Dnl. We observed higher ligation efficiencies for both enzymes at 37^0C for an incubation time of 1h compared to overnight incubation at 16^0C. With a 5 fold excess of the donor strand and an incubation period of 1h at 37^0C we observed *ca* 50 % ligation yield for T4 Dnl and *ca* 80 % for T4 Rnl2. This is the first example of a successful splint-mediated ligation where a non-natural azide functional group is present directly at the ligation site.

As expected, ligation was not possible at all if the fragment carrying the azide modification has first been subjected to CuAAC for attaching a fluorophore and then to ligation. This is due to the small size of the azide functional group compared to a bulky fluorophore and therefore RNA modified with an azide functional group can still be recognized as substrate by ligases whereas RNA modified with a bulky fluorophore cannot. This directly demonstrates the advantage of our two-step labeling approach by click chemistry compared to one-step enzymatic labeling.

CuAAC of internally azide-modified RNA

After ligation our terminal azide modification became an internal modification. It is well known that the reactivity of many functional groups often decreases severely when buried deep inside the tertiary structure of RNA. In fact, a similar observation has also been reported for ethynyl and octadiynyl substitutions in DNA as well as for alkynes in single and double stranded DNAs. Indeed, we observed a decreased CuAAC labeling efficiency (75-85 %) for our internally azide-functionalized RNA even at elevated concentrations of the catalyst and the alkyne reaction partners. The reactive azide due to its positioning in a base paired region of the RNA is not exposed towards the active copper-catalyst and alkyne moiety, which is directly reflected as a reduced yield during CuAAC conjugation. However, the local environment of this azide residue can be conveniently changed by the addition of a

complementary helper DNA strand which forces the azide modification to move into a 9-nucleotide bulge loop position (Fig. 4.13). Under these conditions we again observed quantitative functionalization by CuAAC. Taking into account all of these results, we are able to show that our RNA labeling toolbox is capable of site-specific terminal as well as internal modification.

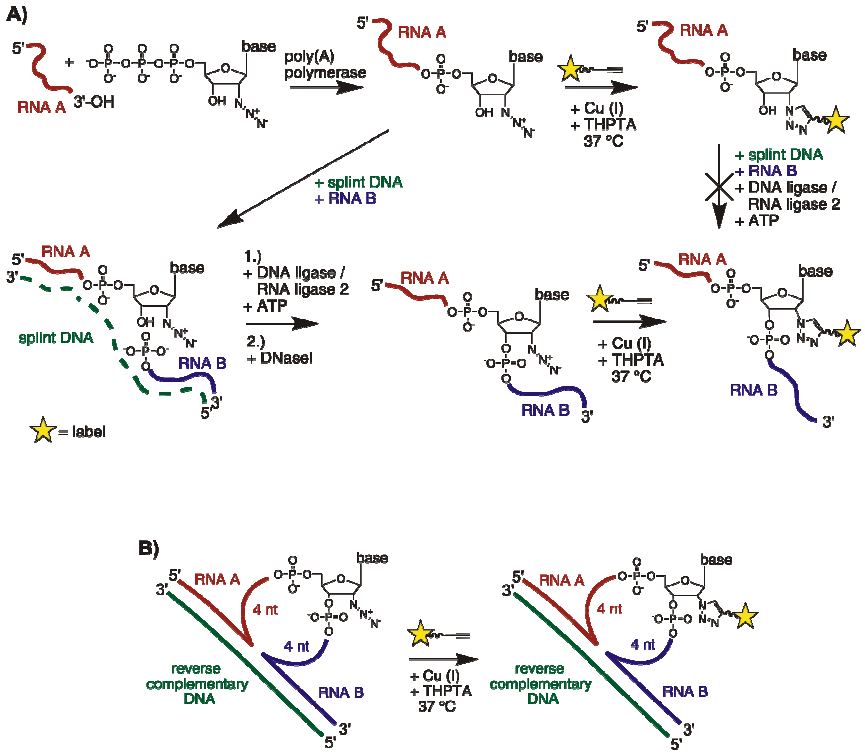


Figure 4.13: A) Schematic representation of splinted-ligation and click modification before or after ligation, **B)** click modification of internal azide in presence of complementary helper DNA strand.

4.2.4 Discussion

In our two-step modular RNA labeling approach we exploit poly(A) polymerases to incorporate azide modified nucleotide to the 3´-end of the RNA of interest followed by derivatization of that azide residue with various bio-orthogonal click reactions. We have demonstrated that in comparison to a one-step enzymatic approach where the modified nucleotide already carries the functional group of interest, our two-step labeling strategy brings a number of advantages. Since the click reaction partners can be changed in a modular way, enzymatic conditions for single incorporation of the modified nucleotide have to be optimized only once and that modified RNA can be conjugated to a number of various different functional entities *via* different click reactions without either having to synthesize a new modified nucleotide each time or to optimize the enzymatic reaction conditions for their incorporation in case of every single new compound. There is certainly a big advantage of incorporating an azide compared to alkyne into the biomolecule because azides, unlike

alkynes, can participate in various other copper-free click reactions as well. We have successfully demonstrated this idea by using CuAAC and SPAAC for the attachment of sensitive fluorophores to RNA after enzymatic azide incorporation at the 3´-end of the RNA of interest. Despite the advantages of incorporating an azide over an alkyne into the biomolecule, examples of azide incorporation into nucleic acids are rare. The chemical instability of azides during standard solid-phase oligonucleotide synthesis (235) has prevented its extensive exploitation for labeling RNA by chemical approaches whereas the alkyne moiety has been widely used for the very same purpose. Therefore enzymatic reactions are an attractive alternative which has the added advantage of not being limited to short sequences.

We have observed that yeast PAP can readily add one (or sometimes more) class I or class II azide-modified NTPs to the 3´-terminus of RNA but can, in most cases, not easily extend these reaction products, even with its preferred substrate ATP. This will ultimately lead to the dominance of RNA sequences carrying a single (or sometimes two) label in the reaction mixture.

Similarly, *E. coli* PAP can well accept 2´-N_3-2´-dATP as substrate, as well as can extend this modified 3´-end with its preferred substrate ATP. However, when both the NTP and the 3´-terminus of the RNA-substrate are 2´-N_3-modified, the *E. coli* PAP reaction is considerably slowed down (by 10- to 20-fold), which again leads to the dominance of RNA species carrying a single 2´-azide modified nucleotide at its 3´-end in the reaction mixture.

Unlike most other RNA 3´-end labeling approaches, we have established a method that is capable of introducing any one of the four (A, C, G and U) nucleotides in a chemically modified form. Contrary to ligase-based approaches, which are often restricted to only one kind of nucleotide incorporation namely cytidine (since cytidine is preferred at ligation sites over the other three nucleotides in the 3´-fragment by phage ligases (250,251)), our approach does not demand the presence of a suitable protecting group on the modified nucleotide to achieve single label incorporation (for example ligase-based labeling approaches involve modified pCp analogs where single label incorporation was achieved by blocking the 3´-hydroxy functional group with a phosphate residue (249)). We rather took advantage of kinetic control of the enzymatic reaction to ensure single label incorporation. This enables the 3´-terminal modification of sequences from any origin (synthetic or natural), causing only minimal mutation, and thereby preserving function.

Both, the PAP reaction and the CuAAC can be adjusted to a wide range of RNA concentrations (at least 200 nM – 80 µM for PAP and at least 50 nM – 50 µM for CuAAC), thereby making our labeling strategy applicable for very low as well as very high amounts of RNA. Since both reactions proceed with high yield and without significant RNA degradation, the overall efficiency of our two-step functionalization approach is very high. As a

consequence, an outstanding sensitivity can be achieved upon fluorescent labeling, enabling the detection of RNA amounts below 100 amol, which exceeds the only comparable report of RNA detection sensitivity after CuAAC labeling (227) by more than three orders of magnitude. Other reports with similar detection limit involve incorporation of multiple convertible residues, thereby causing a severe mutation to the native sequence and most likely hampering its function (187), whereas our labeling protocol achieves the same detection limit with only a single (or at most two) label per biomolecule.

Another major advantage of our labeling approach comes from the kinetic control of the PAP reaction – contrary to previous strategies, this enables us to incorporate a single label to the target RNA sequence, yet leaving an enzymatically manipulatable 3´-end after modification. We have successfully exploited these modified RNA 3´-ends for polyadenylation, adapter ligation and splinted ligation essentially broadening the scope of our approach towards site-specific internal modification. To achieve quantitative CuAAC labeling of azide functional groups present in a rather sterically non-accessible internal position of RNA we have used a very simple yet efficient strategy, which involves the use of a complementary helper DNA which forces the internal azide to move in a nine nucleotide bulge-loop position. With this strategy we again observed quantitative CuAAC labeling.

Noteworthy, due to the possibility of attaching the nucleotide of choice (A, C, G, U) prior to ligation, we are capable of targeting any internal position within a *de novo* synthesized RNA, without mutating its sequence. This should help preserving the integrity of functional RNAs that are to be studied after labeling.

To even further broaden the scope of our labeling approach and to demonstrate the advantages of incorporating an azide over alkyne into the biomolecule another click reaction, namely SPAAC, has been used to attach sensitive fluorophores to the 3´-end of the RNA after modification by PAP. Copper-free reactions are more compatible with living cells than CuAAC and may therefore allow for new *in vivo* applications.

In conclusion we have established a chemo-enzymatic RNA functionalization strategy which does not require any special skills or lab equipments for chemical synthesis. In addition it involves only commercially available enzymes and compounds thereby making this labeling toolbox more available towards molecular biologists. Because of the modular nature of click chemistry this labeling approach is amenable towards incorporation of a diverse array of biologically relevant functional tags to the RNA without having to synthesize a new compound for every single functional group or optimization of enzymatic incorporation for every single new chemical entity. Our labeling protocol is applicable for very low as well as very high amounts of RNA. Moreover, this strategy offers incorporation of any one of the four (A, C, G, U) nucleotides in a chemically modified form, thereby causing no sequence mutation to the target RNA which is beneficial to preserve its function. All in all, this is an

unprecedented advancement in the field of site-specific, chemo-enzymatic RNA functionalization.

4.3 Site-specific, one-pot, simultaneous double-labelling of DNA by orthogonal click reactions[7]

Introducing site-specific chemical modifications into DNA is of high interest in modern life sciences and nanotechnology(261). The number of various different applications which involve the use of modified DNA is almost as large as the number of modifications that has been introduced on to DNA by purely chemical or biochemical means. While a lot of effort has been put into the design of such chemical modifications to improve certain properties of nucleic acids (49,50,52), various strategies by which one can introduce such modifications are rather limited and only little effort has been put into the development of new techniques to introduce such modifications site-specifically to DNA. One crucial class of such modifications is fluorophores which revolutionized our understanding of biological mechanisms by the means of direct visualization. DNA molecules carrying multiple fluorophores at each ends are widely used as hybridization-based probes for *in vivo* RNA imaging(262-264). Additionally DNA molecules modified with two different fluorophores at specific internal positions are indispensable for structural and functional studies involving DNA(265,266). The commonly used strategy to introduce chemical modifications into DNA involves the use of modified phosphoramidites or solid supports during chemical DNA synthesis(31,32). However this approach demands for a *de novo* synthesis of either the phosphoramidite or the solid support for each new compound, so that one can not test a variety of different modifications, e.g. fluorophores, easily whereas in many biophysical experiments it is often necessary to first screen different fluorophores for the best-suited spectral properties. Moreover solid-phase DNA synthesis requires rather high quantities of phosphoramidites which again poses an additional barrier since the synthesis of fluorophore modified phosphoramidites often involve many steps leading to poor overall yield of the final compound. Additionally the harsh reaction conditions involved in phosphoramidite-chemistry does not allow introduction of sensitive chemical entities(235). Coupled with the inherent degenerative nature of chemical oligonucleotide synthesis (Section 1.4), this approach is restricted to short DNA carrying single modification per biomolecule preferably at any one of the ends. For longer DNA carrying multiple internal site-specific modifications, mostly post-synthetic chemo-enzymatic ligation strategies are applied which involves ligation of multiple singly-modified short DNA(40). Another strategy is to use different protecting groups during chemical DNA synthesis followed by sequential deprotection and conjugation with different

[7] This part of the work has been performed together with J. Schoch.

chemical entities in a post-synthetic manner(231). However this approach requires multiple intermediate purifications ultimately leading to a laborious and time-consuming labelling protocol with poor labelling efficiency.

Given the advantages of bioorthogonal click chemistry, an alternative strategy would be to incorporate two different small reactive groups site-specifically into DNA during solid-phase synthesis and to modify them post-synthetically by orthogonal click reactions. Requirements therefore are orthogonal reactivities of the click-reaction partners. As mentioned earlier in section 1.4, the use of click reactions as a two step labelling approach has a range of advantages over one-step direct labelling. One such crucial advantage is that the building blocks can be used as versatile modules to attach a diverse array of different chemical functionalities. Therefore modified phosphoramidite containing the reactive handle has to be synthesized and incorporated only once and that modified DNA sequence can be stored and further functionalized at any time with a broad range of different dyes of choice without having to synthesize a new phosphoramidite for each new different fluorophore molecule.

To this end, we envisioned that copper-catalyzed azide-alkyne cycloaddition (CuAAC) and inverse electron demand Diels Alder cycloaddition (DAinv) might form such a bioorthogonal click reaction pair (Fig. 4.14). This second reaction (DAinv) is relatively new among click reactions and does not require any transition metals, therefore has a high potential towards *in vivo* applications. Additionally in contrast to other metal-free click reactions, most notably strain-promoted azide-alkyne cycloadditions (SPAAC), the reactive alkenes and tetrazines are highly soluble in water; hence this click reaction is not restricted to cell surface labelling. Finally owing to its unprecedented high reaction rate this reaction allows (at least in certain cases) labelling of biomolecule with a one-to-one stoichiometry at sub micromolar biomolecule concentration. We rationalized that the alkynes that are reactive in CuAAC should not be reactive towards DAinv due to their low HOMO energies(126).

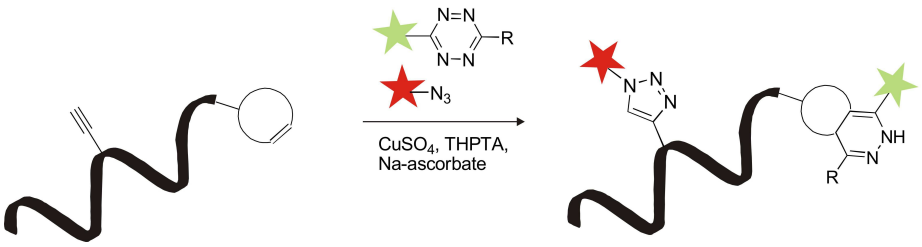


Figure 4.14: Schematic representation of double-modification of a DNA by concurrent DAinv and CuAAC reaction.

Therefore in a first attempt the orthogonality of those two click reactions has been tested. For that purpose DNA molecules bearing different kinds of strained alkenes for DAinv and terminal alkyne for CuAAC were prepared by solid-phase synthesis using the

corresponding phosphoramidites and fully deprotected, purified DNAs were subjected to DAinv or/and CuAAC with various different dyes either carrying tetrazines for DAinv or azides for CuAAC. The crude reaction mixtures, without any further purification, were directly analyzed by LC-MS/MS fragmentation. LC-MS/MS analysis revealed that these two click reactions are indeed absolutely orthogonal to each other. Dyes bearing tetrazine moieties have been exclusively conjugated to only strained-alkenes and dyes carrying azide modification has solely reacted to the terminal alkyne modification. No cross reactions have been found. We further revealed that the reaction sequence can be altered in any way and does not require any intermittent purification and most importantly these two reactions can be performed simultaneously in one pot without compromising the labelling efficiency just by the addition of all components together and incubating the reaction mixture at room temperature. To finally investigate the robustness of our approach for concurrent double-labelling of DNA, we prepared a double stranded 109mer DNA pool carrying 70 random nucleotides flanked by two constant primer binding domains at each end, by PCR with primers each carrying a single modification, namely either a terminal alkyne or a strained alkene. Thereby the resultant double stranded DNA pool was derivatized at each strand with either a terminal alkyne or a strained alkene. This double stranded DNA pool was then subjected to either individual DAinv and CuAAC reactions, or to a simultaneous one-pot DAinv/CuAAC reaction using TAMRA-tetrazine and Cy5-azide. After gel-electrophoretic separation of the reaction mixtures, the TAMRA scan therefore reveals all DAinv products, while all CuAAC products are visualized in Cy5 scan. The bands in both scan overlay perfectly, suggesting attachment of both dyes to the same macromolecular species. The primer-pair used for preparing the pool was also designed for restriction digestion with BspCNI which has been used to demonstrate the specificity and orthogonality of these two click reactions. Therefore the dual labelled PCR product after digestion with BspCNI produced two fragments - one small carrying only the TAMRA dye (i.e., the DAinv product) and one large bearing only Cy5 (i.e., the CuAAC product). No cross-reactivity was observed.

In summary, we found out an orthogonal click reaction pair and successfully applied these two reactions for site-specific, one-pot, simultaneous double-labelling of DNA. Although both of these reactions were known before and has individually been used for biomolecule labelling by our group(123,124,267) as well as by others(81,268), this is the first example of concurrent, one-pot, double-labelling of DNA using two different orthogonal click reactions. We have successfully demonstrated the orthogonality of these two reactions towards each other, therefore the selectivity in site-specific biomolecule labelling. We found no cross-reactivity between these two reactions. Using this strategy we were able to site-specifically modify short as well as long DNA molecules with very high labelling efficiency. These two reactions can be applied in any order or simultaneously in one-pot without

compromising the labelling efficiency, just by adding all components together and incubating the reaction mixture at room temperature. This strategy does not involve any intermittent purification, therefore needs only little hands-on time. Current methodologies for introducing multiple modifications into DNA often involve laborious synthesis of modified phosphoramidites (or solid supports) or complicated enzymatic ligation schemes involving ligation of many smaller fragments – moreover these labelling strategies involve multiple purifications and often lead to very poor overall labelling efficiency. In comparison to those techniques, this is an unprecedented advancement for site-specific, double-labelling of DNA using orthogonal click reactions.

4.4 Conclusion and future work

In a summary, we have developed chemo-enzymatic strategies to functionalize long RNAs site-specifically. We have successfully demonstrated that our two-step functionalization strategy is far more superior to one-step direct enzymatic labelling and this two-step approach is rather indispensable for certain bulky modifications which, due to their severe sterical hindrance, cannot be accepted as a substrate by the enzyme and therefore cannot be inserted into the RNA. Additionally, due to the modular nature of the click reactions, our approaches allow functionalization of the target RNA with a diverse array of different chemical entities without having to synthesize a new molecule every time or optimizing the enzymatic reaction conditions for their incorporations for each single new compound. In contrast to solid-phase RNA synthesis, our approach is not limited to short sequences. We were able to functionalize total RNA from *E. coli*, 209 nucleotide long structured tandem riboswitch aptamers as well as random nucleic acid pool comprising of 233 nucleotides. These experiments clearly demonstrate the sequence independency, usability and robustness of our chemo-enzymatic, site-specific RNA functionalization strategy. Current effort in this regard is directed towards broadening these functionalization strategies. At present, for our initiator-based labelling approach, we are limited to only phage derived RNA polymerases and to only one type of promoter. We have already synthesized initiator nucleotides which posses all advantages of our previously described initiators but can be used in combination with other promoters and RNA polymerases from phages as well from higher organisms. Therefore current effort involves optimization of enzymatic reaction conditions for their efficient incorporation into RNA. Similarly, for our nucleotidyl transferase-based RNA 3´-end labelling approach, presently we are restricted to the fact that the target RNA sequence has to be a substrate of the poly(A) polymerase. We are currently on the verge of using other nucleotidyl transferases so that we can address an even broader range of various different structural RNAs. In a long run, we want to combine our chemo-enzymatic

5´- and 3´-end labelling approaches for preparing circular ribozymes, since circular ribozymes are known to have higher activity, improved nuclease stability and lower divalent metal ion dependence(46) - thereby plausibly allowing an even broader range of applications for RNA-based gene-regulation systems.

In our one-pot, concurrent functionalization approach for DNA with multiple fluorophores, current effort includes the development of superior copper-stabilizing ligands to improve the rate of this reaction. Since the DAinv reaction is extremely fast as well as completely compatible with living systems, we are planning to improve the rate of the CuAAC as well. Development of superior copper-coordinating ligands should not only facilitate the rate of this reaction, but also should decrease the cytotoxic effects induced by copper. Therefore in a long run it might be possible to implement our simultaneous, double-labelling approach for multicolour imaging of biomolecules *in vivo* in collaboration with experts in the respective fields.

5 Affinity based chemical RNomics

5.1 Concept and workflow

As mentioned earlier in section 1.1, riboswitches are solely RNA-based metabolite sensing gene regulatory system, currently accounting for about 3-4 % of the total gene regulation for certain prokaryotes(7,9,11). All riboswitches contain a simple modular architecture – a metabolite sensing aptamer domain and a gene regulatory expression platform. Generally, the aptamer domain binds its cognate metabolite with a very high affinity, often multiple orders of magnitude higher than their *in vitro* selected counterparts. As mentioned earlier in section 1.5, all known riboswitches have been discovered by rational approaches, followed by experimental verification of the candidates, and not by a random experimental screening strategy. These discoveries were made by either investigating mRNAs in pathways where biochemical evidence showed regulation, but no protein regulator had been found, or by genetic screens. Biocomputational approaches were mainly applied to find known aptamer motifs in genomes, to find known riboswitch types in different organisms, or to detect conserved structural elements in the untranslated regions (UTRs) of mRNAs. Given the fact that the *de novo* prediction of RNA three-dimensional structure from sequence is quite inaccurate in the absence of reference structures, it can be assumed that many more riboswitches exist that have not been discovered so far, and that may not be identified at all by the aforesaid approaches.

The goal of this work is to develop a chemically driven screening strategy for the isolation of riboswitches and ``natural aptamers`` that does not rely on rational assumptions, and in a long run the application of this strategy for the isolation of currently unknown riboswitches, which are then to be biochemically characterized.

This approach, which we call ``Affinity-based chemical RNomics``, does not require any prior sequence information of the target RNA and solely rely on the fact that all currently known riboswitches bind to their cognate metabolite with a very high affinity. Ideally this strategy should allow isolation of riboswitches from any organism, given that a total RNA isolate is available for that organism. One of the central concepts in this strategy is the conversion of metabolites to the corresponding trifunctional photoprobes.

The general concept of our approach has been outlined in figure 5.1. The three crucial features of our trifunctional photoaffinity probes are:

The metabolite - this part of the molecule is responsible for efficient binding of our photoprobe to the target RNA sequence.

A photoactivatable functionality - in our case an aryl azide which generates a reactive nitrene intermediate upon irradiation with UV-light and forms covalent bond with any nucleophiles nearby.

A convertible affinity handle – this is a terminal alkyne in our case, which can be converted to an affinity handle by CuAAC with biotin azide.

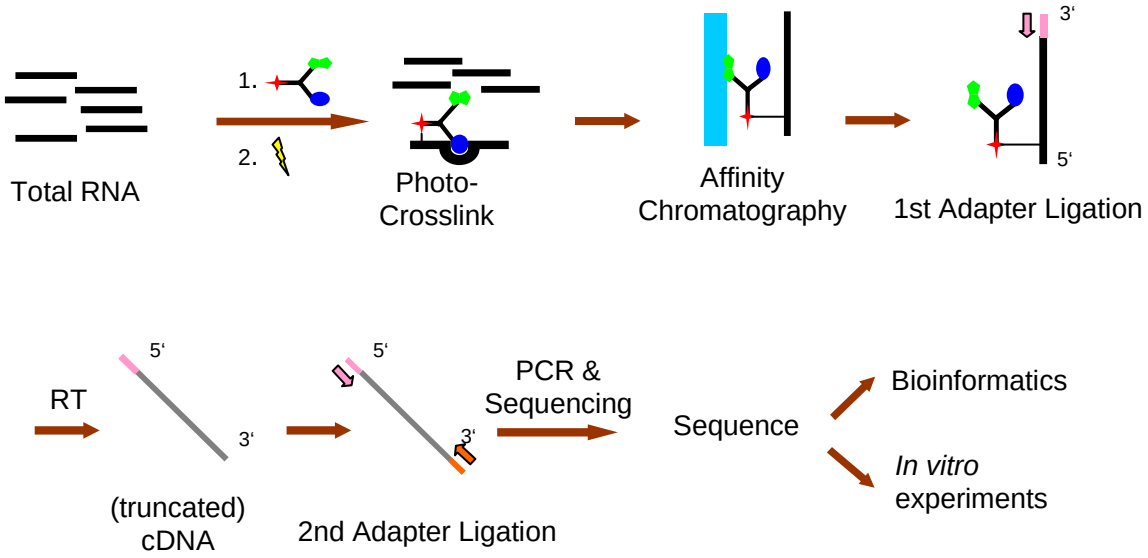


Figure 5.1: Workflow for the isolation of new riboswitches by photoaffinity crosslinking. Symbols: blue oval: metabolite; red asterisk: photoreactive group; green heptagon: convertible affinity handle.

After binding their cognate RNAs in a total RNA preparation, these affinity probes are covalently crosslinked to the target RNA sequences by irradiation with UV light, thereby resulting in a covalent transfer of the convertible affinity handle to the target RNA sequence. Thus now the target sequences can be selectively fished out from the rest of the unrelated sequences by performing a CuAAC conjugation with biotin azide followed by affinity purification on a Streptavidin-agarose column. To our conclusion, this click chemistry-based convertible affinity tag approach is superior compared to the direct attachment of an affinity handle because of i) small size of the CuAAC reaction partners, thereby causing only minimal interference during the event of riboswitch binding to the small molecule (Section 1.5.1) and ii) poor water solubility of biotinylated compounds. This convertible affinity handle is clearly an advantage over genomic SELEX approach as mentioned earlier in section 1.5.1. Furthermore, due to the unprecedented high reaction rate of aryl azide crosslinking (often in microsecond time scale), this approach might even enable us to isolate transiently interacting RNA species. The affinity purified RNA sequences can then be subjected to a ligation scheme to attach constant sequences at each end required for amplification as depicted in figure 5.1. Previous studies from our laboratory as well as from others(147,149,150) indicated stalling during reverse transcription, and therefore incomplete cDNA synthesis over UV-crosslinked RNA templates. Hence, we decided to perform the second adapter ligation at

the cDNA level. We have used chemically pre-adenylated adapters in combination with T4 RNA ligase 2 truncated and T4 RNA ligase 1 to achieve quantitative 3´-RNA adapter ligation. The second adepter ligation at the cDNA level has been achieved following a procedure similar to TD-PCR(269-271). The amplified sequences are to be subjected for illumine sequencing. Due to the enormous sampling depth achieved in modern deep sequencing techniques, we envisioned that the use of deep sequencing in this project might be extremely beneficial.

Since this type of experimental screening approach has not been previously used to isolate riboswitches, we decided to validate our approach first for this purpose. Requirements would therefore be the synthesis of a trifunctional molecule bearing a metabolite for which there exists already a riboswitch. However, in addition to fish out known riboswitches, different attachment sites of the same metabolite, can lead to the isolation of riboswitches responsive to the very same metabolite but with different recognition topology.

5.2 Design of the photoaffinity probes – compromise between best design and synthetic effort

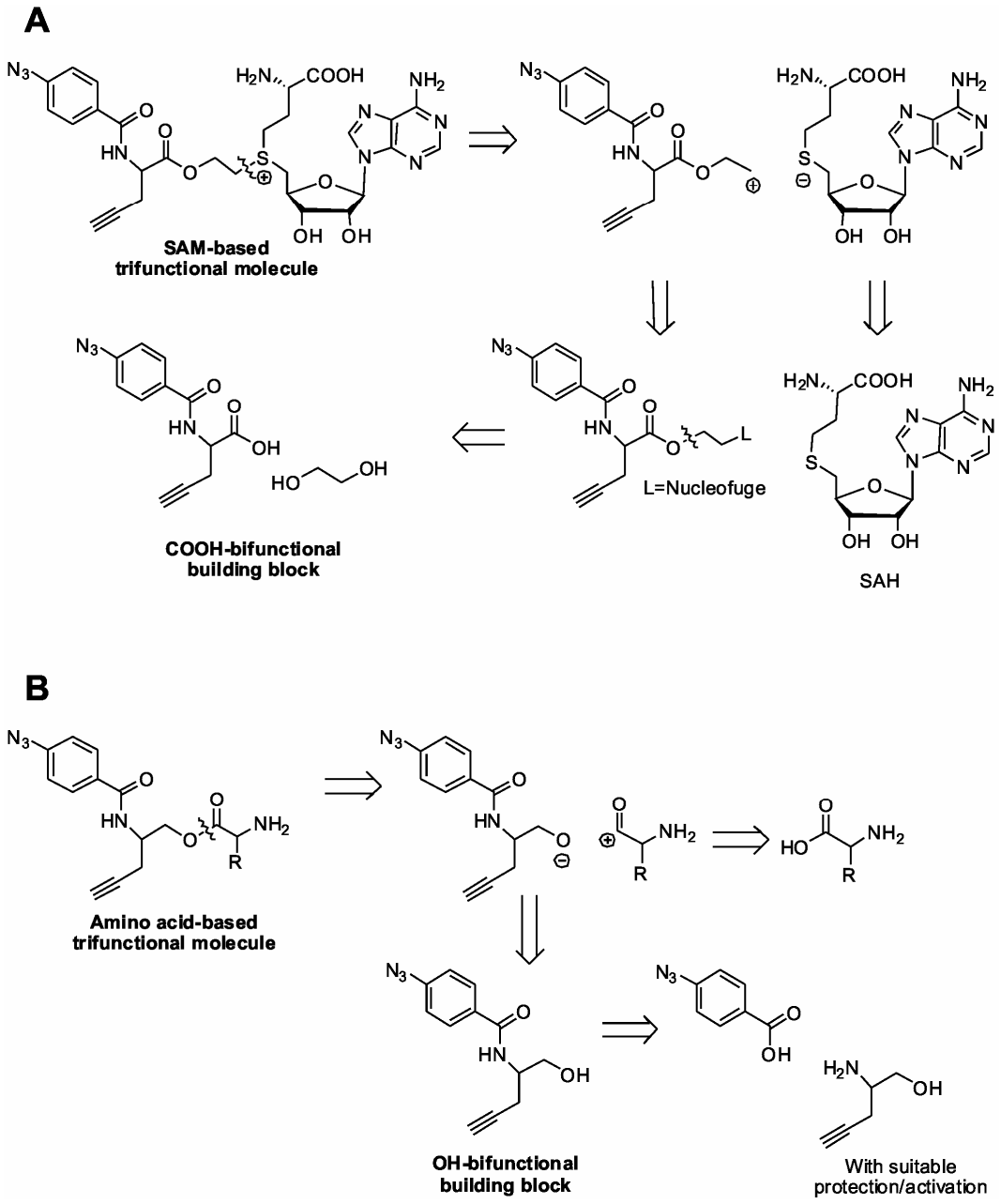
Detailed analysis of biochemical data (and biophysical data whenever available) of a given riboswitch would allow us to find out the best derivatization position for its cognate metabolite. However, in most cases this leads to a target oriented synthesis and demands for a *de novo* synthesis of each different molecule. Although this is ideal from the viewpoint of biology, it is extremely laborious and to our conclusion absolutely non-judicious choice from the standpoint of organic synthesis. The reason for this can be divided into the two following stages.

Firstly, in the absence of biophysical data (mostly NMR data), theoretical prediction for the energy minimized conformer of even a small molecule is quite inaccurate. Therefore rationally designed trifunctional photoprobe might not show predictable behaviour during binding with the corresponding riboswitch. Secondly, this task becomes even more challenging when the riboswitch recognizes its cognate metabolite at a conformation which lies in a local energy minimum rather than in a global one.

Thus, it is a prudent choice to establish a synthetic methodology which will allow quick derivatization of various different metabolites as well as same metabolite but through different attachment sites rather than a *de novo* target oriented synthesis approach, although this strategy does not meet the criteria of the best rational design in every cases. Therefore a convergent synthetic route has been established leading to the syntheses of a diverse array of different bifunctional molecules each carrying a separate functional entity for further attachment of the target metabolite molecule.

5.3 Convergent synthetic approach

Retrosynthetic analyses of SAM- and amino acid-based based trifunctional photoprobes are shown in scheme 5.1. Among various different theoretically possible routes we have taken the one involving only natural donor and acceptor synthons.



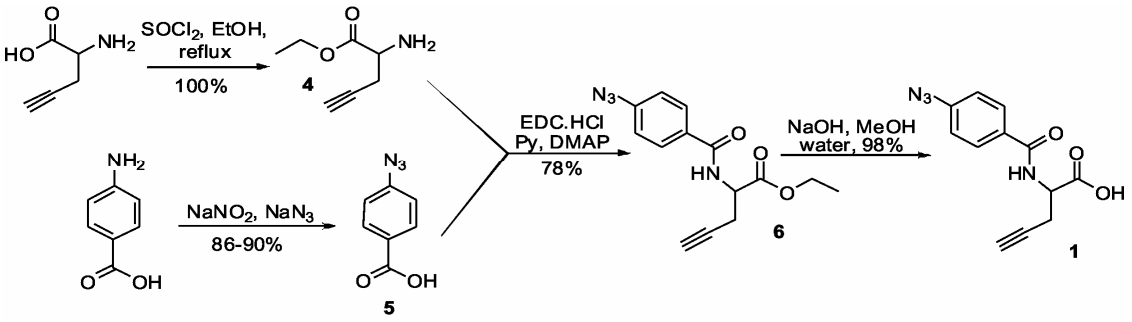
Scheme 5.1: Retrosynthetic analysis of SAM and amino acid-based trifunctional molecule.

5.4 Synthesis of first generation bifunctional building units

5.4.1 Synthesis of carboxylic acid bearing bifunctional molecule

The synthesis of the bifunctional molecule bearing a carboxylic acid for further derivatization with a metabolite is shown in scheme 5.2. This building module can be synthesized *via* a convergent synthetic route in only four steps with high yields as depicted in the scheme 5.2. The first step of the synthesis involves the selective esterification of the carboxylic acid moiety of the racemic propargylglycine with ethanol in presence of an unprotected primary amino group. It is obvious that standard carboxylic acid activation reagents for coupling to the corresponding alcohol will lead to the polymerization of the starting compound due to the presence of the free amino group. Therefore, to achieve this conversion we have used a modified version of Fischer´s esterification technique. Traditional Fischer´s esterification involves the use of a very strong and hygroscopic acid (most commonly concentrated sulfuric acid) in combination with high excess of one reaction component (generally the alcohol counterpart due to their lower boiling point compared to the acid in the majority of cases and thereby causing easier removal of the excess reagent) to drive the reaction equilibrium towards product formation. In our reaction, ethanol is used as a solvent as well as reactant to facilitate the reaction equilibrium towards product formation. Due to the known hydration problem of terminal triple bonds in strongly acidic media (which is further catalyzed by soft metal ions like Hg^{2+}), we envisioned that the commonly used concentrated sulfuric acid might not be suitable for our purpose. Therefore we have adapted to an *in situ* acid generation system, where we have used only 3 equivalents of thionyl chloride. The reaction does not proceed in this case through the formation of acid chloride. Rather thionyl chloride reacts with the solvent, in this case ethanol, thereby generating hydrochloric acid which not only catalyzes the esterification reaction but also provides a transient protection of the primary amino functionality by protonation. It is worth mentioning here that this esterification strategy works best for volatile alcohols. Higher boiling alcohols (e.g. all aromatic alcohols or higher aliphatic alcohols) lead to the generation of side products which are often difficult to remove from the desired product plausibly due to the thermal stability of higher alkyl sulfates formed during this reaction.

The photocrosslinker **5** can be conveniently prepared from *p*-aminobenzoic acid following previously described procedure(272). Compounds **4** and **5** can be coupled together using water soluble carbodiimide in combination with a nucleophilic catalyst, DMAP, as indicated in scheme 5.2. The final carboxylic acid bifunctional building unit **1** can be obtained in high yield by a standard saponification of compound **6** (scheme 5.2).

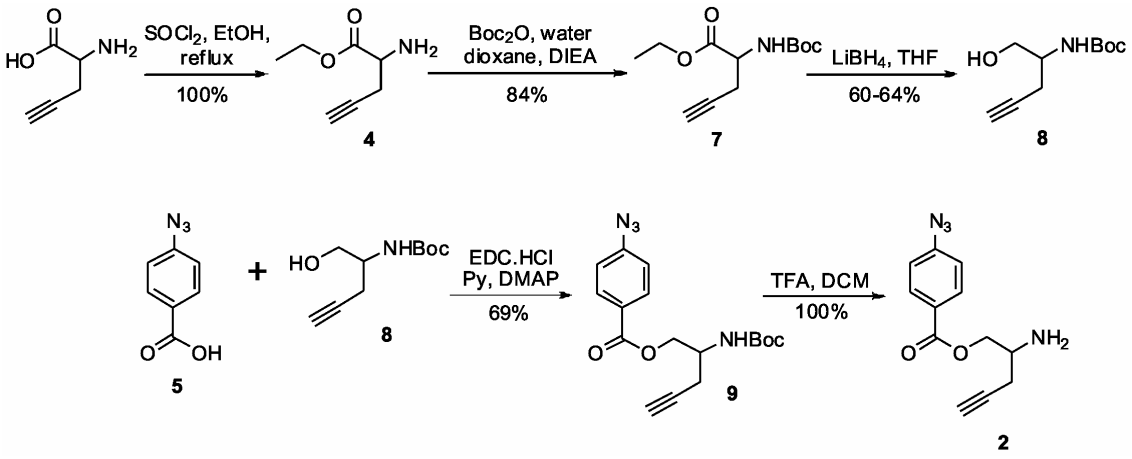


Scheme 5.2: Synthesis of carboxylic acid bearing bifunctional molecule

It is important to mention here that aryl azides are not stable towards heat, therefore reactions conditions requiring high temperatures should strictly be avoided through out this complete synthetic project. Furthermore, due to the presence of alkyne and azide in the same molecule it is recommended to avoid transition metal catalyzed reactions during the course of these syntheses.

5.4.2 Synthesis of primary amine bearing bifunctional molecule

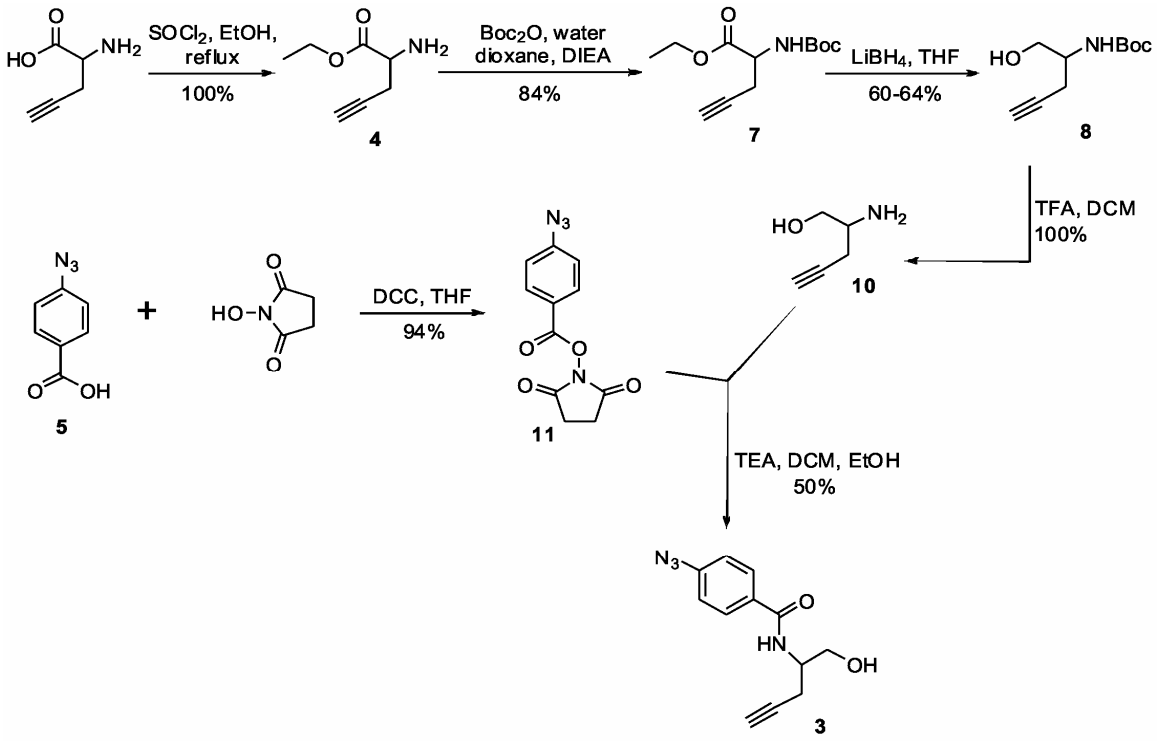
Similarly, compound **2** can be synthesized following a similar convergent synthetic route as depicted in scheme 5.3. To our conclusion a higher yield can be achieved for the Boc-protection reaction when performed with compound **4** rather than directly on propargylglycine. The reason can either be the poor solubility of propargylglycine compared to compound **4** in the reaction mixture or the Boc_2O mediated activation of free carboxylic acid followed by reaction with the free amino group thereby leading to polymerization. The most crucial step in this synthesis is the selective reduction of the ester linkage in presence of a urethane (in Boc functional group) functionality and the terminal triple bond. This selective reduction can be achieved by lithium borohydride in THF. We have found this reducing agent particularly useful for our purpose as also discussed later. Traditional carbodiimide coupling between compounds **8** and **5** led to the formation of compound **9** in 69 % yield followed by a conventional Boc deprotection with TFA in DCM to finally yield our bifunctional amine building module (**2**) in high yield (scheme 5.3).



Scheme 5.3: Synthesis of primary amine bearing bifunctional molecule

5.4.3 Synthesis of primary alcohol bearing bifunctional molecule

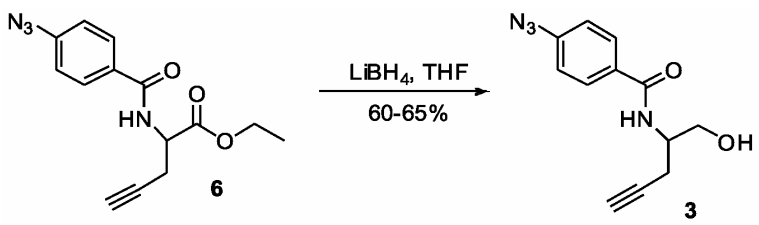
The synthesis of compound **3** involves the highest number of steps compared to the other two building modules described previously, as depicted in scheme 5.4. Compound **10** was synthesized through a number of sequential protection and deprotection steps whereas this, in principle, could be achieved by a direct reduction of the propargylglycine. Repeated efforts to achieve compound **10** *via* direct reduction of propargylglycine by various hydride transfer reagents led to little or no product formation. This is plausibly due to the fact that compound **10** bears a β-aminol architecture which are known for their excellent metal chelation property, water solubility, sensitivity towards aqueous acids and volatility. Since all known hydride transfer based reduction methodology relies on aqueous work up, it was impossible to synthesize compound **10** by a direct reduction of propargylglycine. We did not observe any such problems during the reduction of compound **7** to compound **8** with lithium borohydride in THF, since the amino group is protected in this case, therefore cannot participate in metal chelation. Since compound **10** possess two different nucleophilic functional groups, it will not be possible to achieve a selective coupling of the amino group to the carboxylic acid moiety from **5** by means of direct coupling as has been performed in all previous cases. Rather we have adapted to NHS-coupling strategy which is well known for coupling of amines to carboxylic acids in aqueous media. However, even under optimized conditions we were able to achieve only 50 % yield (using stoichiometric amounts of both coupling partners) which is quite low in comparison to all other steps as depicted in scheme 5.4. Therefore we have developed an alternative synthetic route using lithium borohydride based on the selective reduction of an ester functional group in presence of aromatic amide, azide and terminal alkyne as discussed in section 5.4.4.



Scheme 5.4: Synthesis of primary alcohol bearing bifunctional molecule

5.4.4 Alternative route for the synthesis of primary alcohol bearing bifunctional molecule

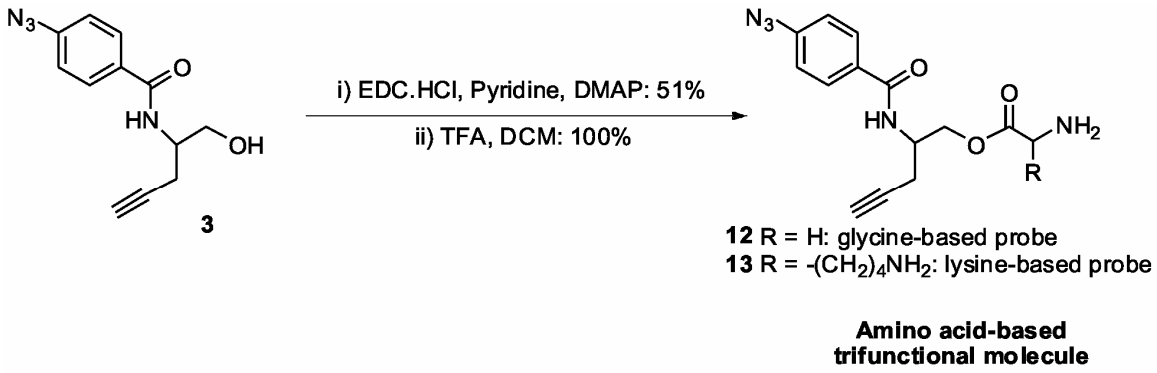
This alternative synthetic route is shown in scheme 5.5. This involves the selective reduction of one of our central building blocks, namely compound **6**. Lithium borohydride seemed to be particularly useful in this case. This pathway allows the synthesis of the –OH bifunctional molecule in a much shorter route as compared to our previous synthetic scheme (scheme 5.5). Furthermore, in this case one single compound, namely compound **6**, can be used as a precursor for the synthesis of two different, -COOH and –OH bifunctional molecule.



Scheme 5.5: Alternative synthetic route for primary alcohol bearing bifunctional molecule

5.5 Synthesis of amino acid (glycine and lysine) -based trifunctional photoprobes[8]

Previous studies(155,273) with glycine and lysine responsive riboswitches reported moderate affinity of these regulatory RNA elements towards the corresponding ester derivatives of the respective amino acids. Therefore, we ideated that derivatization of these amino acids with our first generation bifunctional building module carrying a primary alcohol might be a prudent starting point for our trifunctional photoprobes because of i) the relatively simple synthesis involved and ii) the chemical stability of these molecules compared to some other trifunctional photoprobes (e.g. SAM-based trifunctional molecules (scheme 5.1) are extremely labile). Therefore we have established the synthesis of these two amino acid based trifunctional molecule as indicated in scheme 5.6. The synthesis involves activation of the carboxylic acids of suitably protected glycine or lysine using water soluble carbodiimide followed by coupling with compound **3** to obtain the respective fully protected trifunctional molecule in 51 % yield (for compound **12**). The successive Boc deprotection with TFA in DCM led to the formation of the corresponding fully unprotected trifunctional photoprobe in quantitative yield (scheme 5.6).



Scheme 5.6: Synthesis of amino acid-based trifunctional photoprobes

5.6 Photoaffinity tagging with lysine-based trifunctional molecule

After establishing the synthesis of our trifunctional photoprobe, we optimized the photoaffinity tagging (UV-crosslinking) conditions *in vitro* for lysine-based trifunctional molecule, compound **13,** with *in vitro* transcribed lysine riboswitch from *Bacillus subtilis*. One example is shown in figure 5.2. As expected, we observed higher yields for photoaffinity tagging with increasing probe equivalents. Furthermore, we have also tested the unspecific photo-tagging behavior of one of our bifunctional molecules, compound **3**, and observed little or no tagging.

[8] Yields depicted in scheme 5.6 are for glycine-based trifunctional molecule. These two steps corresponding to the synthesis of lysine trifunctional photoprobe have been performed by Benjamin Strauss.

It is worth mentioning here, that our trifunctional photoprobe **13** showed certain extent of unspecific crosslinking with unrelated sequences when used in buffers with neutral pH. To our rationale, this is due to the presence of free amino groups in compound **13**, which is indispensable for recognition by the lysine riboswitch. This primary amine becomes protonated in buffers with neutral pH and therefore can non-specifically interact with negatively charged RNA-phosphate backbone. As mentioned earlier, due to the unprecedented high reaction rate of aryl azide crosslinking, it is even possible to isolate transient interaction partners, which is to our belief, also a contributing parameter here for this unspecific photo-tagging. Furthermore, since the reactive nitrene formed upon photoactivation of aryl azides are susceptible to nucleophiles, especially primary amines, we recommend avoiding tris-based buffers for photoaffinity tagging experiments involving aryl azides. These photoaffinity tagging experiments although sound similar to well established UV-crosslinking experiments in traditional RNA biochemistry, are fundamentally different since the photoreactive moiety is present on a freely diffusible small molecule, whereas traditional UV-crosslinking experiments involve a photoreactive hybridization strand where high local concentration of the photocrosslinker is achieved *via* a stable duplex formation(272).

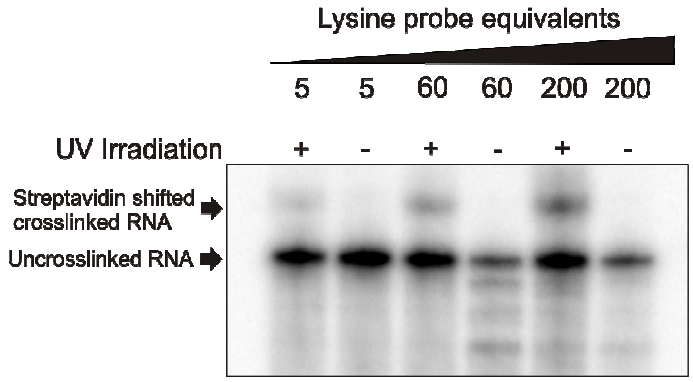


Figure 5.2: *In vitro* photoaffinity tagging by compound 13 for lysine responsive riboswitch from *Bacillus subtilis*. RNA end concentration was maintained at 1 µM in these experiments.

Finally, encouraged by our *in vitro* test photoaffinity tagging experiments, we isolated total RNA from *Bacillus subtilis* and subjected that to *in vitro* photoaffinity tagging experiment following a protocol optimized in the aforesaid experiments. The isolated crosslinked sequences were then subjected to the previously described ligation protocol to attach constant sequences required for amplification. The amplified samples are to be subjected to illumina sequencing for experimental validation/characterization for their biological activities.

5.7 Conclusion and future work

In a summary, for our affinity-based chemical RNomics approach for the discovery of new riboswitches, currently we are at a position to convert many other metabolites to their corresponding trifunctional analogues. This is due to the availability of our convergent synthetic approach as well as a diverse range of different bifunctional building blocks available in the laboratory. Furthermore, we also have developed the synthesis of a set of second generation, chain-extended bifunctional building units. However, the future work of this project truly depends on the deep sequencing data. One major problem, as described previously, with our trifunctional lysine-based photoaffinity probe is its unspecific interaction to nucleic acid backbone. Current effort includes optimization in buffer compositions and pH to suppress this unspecific crosslinking. Presently we are at a stage of validating our approach. Therefore we can also use other metabolite-based trifunctional molecules for this purpose. Requirements would therefore be the existence of a known riboswitch for that particular metabolite. One example could be a trifunctional molecule based on glycine for which we have already established the synthesis. This molecule should not have the aforesaid unspecific crosslinking problem since it lacks a primary amino group, therefore would not be positively charged in a neutral pH environment. Depending on the sequencing data, it might be beneficial to prepare bacterial RNA sample which has depleted ribosomal and transfer RNAs. After finding the known riboswitch sequence in our data-set (thereby establishing the proof of principle for our affinity based chemical RNomics approach), efforts will be directed towards validating the other sequences (if any) for their biological function by various *in vitro* and *in vivo* experiments.

6 Conclusion and future work

In our ligation related work, we have successfully established a splinted-ligation strategy to prepare dual fluorophore-labelled full-length riboswitch constructs for bulk- and SM-FRET measurements for studying ligand induced folding dynamics of SAM-I riboswitch. To our conclusion, T4 RNA ligase 2 is much more efficient in joining nicks of RNA in a double stranded environment compared to T4 DNA ligase. We have optimized the ligation scheme to improve the overall yield of the full-length ligation product from 1 % to 10 %. However, we believe that further optimizations of the pH of the ligation buffer might improve the ligation yield. It is noteworthy here that we have observed incorrect annealing of some of the ligation fragments when only four fragments are used together with the DNA-splint, previously used for the 5-way ligation. This indicates that the correct formation of ligation competent complex (LCC) can be prevented, ultimately leading to decreased overall yield of the product. This seems to be a general problem for all transcriptionally acting riboswitches. This is due to the fact that for all transcriptionally acting riboswitches, the 3´-end of the P1 stem has certain sequence complementarity to the 5´-end of the terminator hairpin. Therefore even though longer ligation fragments are used (in our case each fragment consists of 34 nucleotides which is almost as large as gene-specific primers used in our laboratory to amplify specific riboswitch transcription templates from isolated genomic DNA) there can always be some extent of incorrect annealing. This can be prevented by using multiple smaller DNA splints, each capable of hybridizing to a particular RNA fragment and afterwards leaving a sticky end for overlap with its neighbouring fragment. Thus, all RNA fragments can be separately annealed to its specific DNA splint, then brought together into one reaction vessel and allowed for the formation of the complete LCC by the help of the sticky ends present in each heteroduplex. This ligation strategy is currently under investigation in our laboratory.

In case of our aim to address ligand induced folding dynamics of riboswitches, we have observed similar folding phenomena for the complete riboswitch with its expression platform as reported earlier for only the aptamer region. However we made a couple of observations in studies involving a slightly different, yet non-cognate metabolite, which can not be explained by any known facts about this riboswitch to date. We believe that these data might further shed light on the gene regulation of riboswitches by premature transcription termination and hopefully settle down the long standing discussion about the ``switching`` behaviour of these regulatory RNA elements *in vivo*.

During the course of the aforementioned work, we realized the limitations of existing nucleic acid functionalization strategies. Therefore we decided to use bioorthogonal click reactions as part of our functionalization strategy. After choosing the best click reaction for

our purpose and having its conditions optimized we have developed chemo-enzymatic strategies to functionalize long RNAs site-specifically. We have successfully demonstrated that our two-step functionalization strategy is far more superior to one-step direct enzymatic labelling and this two-step approach is rather indispensable for certain bulky modifications which, due to their severe sterical hindrance, cannot be accepted as a substrate by the enzyme and therefore cannot be inserted into the RNA. Additionally, due to the modular nature of the click reactions, our approaches allow functionalization of the target RNA with a diverse array of different chemical entities without having to synthesize a new molecule every time or optimizing the enzymatic reaction conditions for their incorporations for each single new compound. In contrast to solid-phase RNA synthesis, our approach is not limited to short sequences. We were able to functionalize total RNA from *E. coli*, 209 nucleotide long structured tandem riboswitch aptamers as well as random nucleic acid pool comprising of 233 nucleotides. These experiments clearly demonstrate the sequence independency, usability and robustness of our chemo-enzymatic, site-specific RNA functionalization strategy. Current effort in this regard is directed towards broadening these functionalization strategies. At present, for our initiator-based labelling approach, we are limited to only phage derived RNA polymerases and to only one type of promoter. We have already synthesized initiator nucleotides which posses all advantages of our previously described initiators but can be used in combination with other promoters and RNA polymerases from phages as well from higher organisms. Therefore current effort involves optimization of enzymatic reaction conditions for their efficient incorporation into RNA. Similarly, for our nucleotidyl transferase-based RNA 3´-end labelling approach, presently we are restricted to the fact that the target RNA sequence has to be a substrate of the poly(A) polymerase. We are currently on the verge of using other nucleotidyl transferases so that we can address an even broader range of various different structural RNAs. In a long run, we want to combine our chemo-enzymatic 5´- and 3´-end labelling approaches for preparing circular ribozymes, since circular ribozymes are known to have higher activity, improved nuclease stability and lower divalent metal ion dependence(46) - thereby plausibly allowing an even broader range of applications for RNA-based gene-regulation systems.

In our one-pot, concurrent functionalization approach for DNA with multiple fluorophores, current effort includes development of other copper-stabilizing ligand to improve the rate of this reaction. Since the DAinv reaction is extremely fast as well as completely compatible with living systems, we are planning to improve the rate of the CuAAC as well. Development of superior copper-coordinating ligands should not only facilitate the rate of this reaction, but also should decrease the cytotoxic effects induced by copper. Therefore in a long run it might be possible to implement our simultaneous, double-labelling

approach for multicolour imaging of biomolecules *in vivo* in collaboration with experts in the respective fields.

For our experimental RNomics approach, currently we are at a position to convert many other metabolites to their corresponding trifunctional analogues. This is due to the availability of our convergent synthetic approach as well as a diverse range of different bifunctional building blocks available in the laboratory. However, the future work of this project truly depends on the deep sequencing data. One major problem, as described previously, with our trifunctional lysine-based photoaffinity probe is its unspecific interaction to nucleic acid backbone. Current effort includes optimization in buffer compositions and pH to suppress this unspecific crosslinking. Presently we are at a stage of validating our approach. Therefore we can also use other metabolite-based trifunctional molecules for this purpose. Requirements would therefore be the existence of a known riboswitch for that particular metabolite. One example could be a trifunctional molecule based on glycine for which we have already established the synthesis. This molecule should not have the aforesaid unspecific crosslinking problem since it lacks a primary amino group, therefore would not be positively charged in a neutral pH environment. Depending on the sequencing data, it might be beneficial to prepare bacterial RNA sample which has depleted ribosomal and transfer RNAs. After finding the known riboswitch sequence in our data-set (thereby establishing the proof of principle for our affinity based chemical RNomics approach), efforts will be directed towards validating the other sequences (if any) for their biological function by various *in vitro* and *in vivo* experiments.

7 Experimental section

7.1 General molecular biology methods

Enzymes were purchased from commercial providers and unless otherwise stated used following a protocol recommend by the manufacturer. Gels were stained (wherever applicable) by SYBR Gold following conditions recommend by the manufacturer (Invitrogen). All polyacrylamide gels were visualized by a Typhoon 9400 variable mode scanner. Oligonucleotides bearing fluorophores are scanned in the fluorescence channel by excitation with a suitable laser on Typhoon. For visualizing radioactive bands, storage phosphor screens (GE Healthcare) were exposed to the gels and scanned by excitation with a 633 nm laser. If not otherwise stated, during all quantifications radioactive bands were background-corrected (to the object average of a region in the gel not containing any radioactive bands), and quantified using ImageQuant software (Molecular Dynamics; version 5.2). Oligonucleotides were purchased from either IBA or Dharmocon and used directly without any further purification. The overlay pictures for multiple fluorophore labeled oligonucleotides were created by using ImageQuant and splinted-ligation yields were calculated by normalizing the fluorescence of a Cy5 scan over one lane. Agarose gels were pre-stained with Ethidium bromide and visualized in a alpha-imager. For Streptavidin-shift gels, glycerol gel-loading buffer is used instead of denaturing formamide-gel loading buffer. For gel analysis of fluorophore labeled oligonucleotides, xylene cyanol and bromophenol blue were omitted in the gel-loading buffer to prevent interference with fluorescence scan.

Table 7.1: List of commercial providers for enzymes used in this study

Enzyme	Supplier
Shrimp alkaline phosphatase (SAP)	Fermentas
Calf intestine alkaline phosphatase (CIAP)	Fermentas
Polynucleotide kinase (PNK)	Fermentas
T4 DNA ligase (T4 Dnl)	Fermentas
T4 RNA ligase 1 (T4 Rnl1)	NEB
T4 RNA ligase 2 (T4 Rnl2)	NEB
T4 RNA ligase 2 truncated (T4 Rnl2 tr)	NEB
DNase I	Fermentas
T7 RNA polymerase (T7 RNAP)	Fermentas
T3 RNA polymerase (T3 RNAP)	Promega
SP6 RNA polymerase (SP6 RNAP)	Fermentas
Taq DNA polymerase	Biozyme
Polymerase-X Hybrid DNA polymerase	Roboklon

7.2 Bulk FRET measurements

All fluorescence measurements were performed using Jasco FP-6500 Spectrofluorometer (Labor- und Datentechnik GMBH Deutschland, Gross-Umstadt). The emission spectrum was recorded with the following device settings: excitation wavelength 532 nm, data pitch 0.5, scanning speed 1000 nm/min, emission band-pass 550 – 750 nm. The bulk FRET efficiencies (E_{FRET}) of different constructs were calculated by comparing the fluorescence emission intensities of the donor (Cy3) and the acceptor (Cy5) dyes using the following equation:

$$E = \frac{I_A}{I_A + I_D}$$

I_A: Maximum emission intensity of the acceptor dye (Cy5)
I_D: Maximum emission intensity of the donor dye (Cy3)

For refolding of the riboswitch 2 pmol of RNA was incubated for 1 min at 90 °C in 15 µl of Tris-buffer (50 mM Tris-HCl pH 8.5 at 22 ^{0}C, 20 mM $MgCl_2$, 100 mM KCl) followed by cooling at room temperature for 15 min. After 2 min of equilibration the emission spectrum (550 – 750 nm) was recorded. The data were corrected using a device specific correction function. Of three consecutive measurements the maximum intensities of the donor as well as the acceptor were extracted and the means and standard deviations were calculated. For all six constructs measurements were performed with RNA-only (native state), with 2 nmol of S-adenosyl-L-homocysteine (SAH) and 2 nmol of S-adenosyl-L-methionine (SAM) dissolved in SAM-diluent (10 % (v/v) ethanol in 5 mM H_2SO_4) at 10 °C and 25 °C.

7.3 Quantitative enzymatic phosphorylation of the RNA 5´-end for ligation

To a solution of dye labeled RNA (end concentration 40 µM) in Kinase Ligation (KL) buffer (50 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$) was added ATP and DTT to an end concentration of 2 and 5 mM respectively followed by the addition of T4 PNK (10 u/µl, Fermentas, St. Leon-Rot, Germany) to an end concentration of 0.75 u/µl. The resulting mixture was incubated at 37 °C for 60 min while shaking at 600 rpm and afterwards subjected to phenol-ether extraction. For extraction of the RNA one volume of aq-phenol (lower phase) was added to the reaction mixture and vortexed thoroughly. Layers were separated by centrifugation at 13000 rpm and the organic phase was discarded. This procedure was repeated twice. To the aqueous phase was added three volume of aq-ether (upper phase) followed by vortexing and separating the layers by centrifugation at 13000 rpm. This procedure was repeated twice. Subsequently the RNA was precipitated by adding 5 M ammonium acetate (pH 5.6) to a final concentration of 0.5 M and 2.5x volumes of -80 °C ethanol. The pellet was dried by using the Concentrator 5301 (Eppendorf, Wesseling-Berzdorf, Germany) and the RNA was re-dissolved in water.

Table 7.2: List of sequences used to prepare dye-labeled riboswitch constructs by splinted-ligation.

Name	Sequence	Features
AJ2- AS-yitJ-1-Cy5	5´-AUAUCCGUCy5UCUUAUCAAGAGAAGCAGAGGGACUG-3´	Ligation fragment
AJ3- AS-yitJ-2-Cy3	5´-GCCCGACGAUCy3GCUUCAGCAACCAGUGUAAUGGCG-3´	Ligation fragment
AJ4- AS-yitJ-3-Cy3	5´-AUCAGCCAUGACUCy3AAGGUGCUAAAUCCAGCAAGC-3´	Ligation fragment
AJ5- AS-yitJ-4-Cy5	5´-UCGAACAGCUCy5UGGAAGAUAAGAAGAGACAAAAUC-3´	Ligation fragment
AJ6- AS-yitJ-5-Cy3	5´-ACUGACAAAGUCUUCy3CUUCUUAAGAGGACUUUUU Biotin-3´	Ligation fragment
AJ7- AS-yitJ-1	5´-pAUAUCCGUUCUUAUCAAGAGAAGCAGAGGGACUG-3´	Ligation fragment
AJ8- AS-yitJ-2	5´-pGCCCGACGAUGCUUCAGCAACCAGUGUAAUGGCG-3´	Ligation fragment
AJ9- AS-yitJ-3	5´-pAUCAGCCAUGACUAAGGUGCUAAAUCCAGCAAGC-3´	Ligation fragment
AJ10- AS-yitJ-4	5´-pUCGAACAGCUUGGAAGAUAAGAAGAGACAAAAUC-3´	Ligation fragment
AJ11- AS-yitJ-5	5´-pACUGACAAAGUCUUCUUCUUAAGAGGACUUUUU Biotin-3´	Ligation fragment
AJ1- AS-yitJ DNA splint	5´-Fluorescein-AAAAAGTCCTCTTAAGAAGAAGACTTTGTCAGTGATTTTGTCTCTTCTTATCTTCCAAGCTGTTCGAGCTTGCTGGATTTAGCACCTTAGTCATGGCTGATCGCCATTACACTGGTTGCTGAAGCATCGTCGGGCCAGTCCCTCTGCTTCTCTTGATAAGAACGGATAT-3´	DNA splint

7.4 Splinted ligation

The dye-labeled SAM-riboswitch constructs were prepared via splinted ligation. For the formation of ligation competent complex (LCC) 5 ligation fragments at an end concentration of 10 µM each were annealed to the DNA splint at an end concentration of 9.5 µM by heating at 90 °C for 1 min followed by cooling down for 15 min at room temperature in Ligation buffer (1x KL-buffer, 50 µM ATP, 5 mM DTT). Subsequently T4 RNA ligase 2 was added and the mixture was incubated at 37 °C while shaking at 600 rpm for various durations as indicated in every experiment. For digesting the DNA splint, DNase I (50 u/µl, Fermentas, St. Leon-Rot, Germany) was added to an end concentration of 0.5 u/µl and the resulting mixture was incubated at 37 °C while shaking at 600 rpm for various durations as indicated in every experiment. Analytical ligations were performed in a total volume of 20 µl and preparative ligations in 300 µl. For inactivation of DNase I one volume of FA-buffer (90 ml Formamide, 10 ml 10x Tris/Borate/EDTA-buffer) was added followed by a heat-shock treatment at 90 °C for 15 sec then cooling in ice.

7.5 Analysis and purification of the ligated constructs

Analysis and purification of ligated constructs has been performed using 8 % denaturing polyacrylamide gels. Prior to sample loading, gels were pre-run at 15 W and 150 mA for 15 min. For analysis, gels were scanned using Typhoon 9400 Variable Mode Imager (Amersham Bioscience, Freiburg, Germany) using the following settings:
Fluorescein channel: Excitation with green laser (532 nm) and 526 nm SP emission filter. Cy3 channel: excitation with the same laser and 580 nm BP 30 emission filter. Cy5 channel: excitation with red laser (633 nm) and 670 nm BP 30 emission filter.

In preparative gels samples were loaded twice (preparative and analytical lanes) and preparative lanes were covered with an aluminium foil to avoid dye bleaching during scanning procedure; only analytical lanes were exposed towards Typhoon scan.

From preparative gels bands containing desired constructs were excised and the RNA was eluted using 0.5 M ammonium acetate (pH 5.6) solution by overnight shaking at 650 rpm at 20 ^{0}C followed by ethanol precipitation as described before. The RNA pellet after drying was re-dissolved in water and concentration was checked using the nano-drop (Peqlab, Nanodrop ND-1000 Spectrophotometer).

7.6 General protocol for CuAAC

We always prefer *in situ* generation of Cu(I) by the reduction of Cu(II) salts. For this reduction purpose we recommend sodium ascorbate or ascorbic acids rather than any other phosphine-based reducing agents since phosphine-based reducing agents also reduce the reactive azide when present in high concentration. In every experiment, 5 equivalents of the copper-coordinating ligand and 10 equivalents of the Cu(II) reducing agent with respect to Cu(II) was used. We had best experience by using a combination of sodium ascorbate and $CuSO_4$. Although $Cu(OAc)_2$ is also reported for the same purpose we observed slightly reduced efficiency in the CuAAC labeling when this salt is used as a source of Cu(II). The Cu(II) salt was always added last to prevent biomolecule degradation. Unless otherwise stated, all CuAAC reactions were performed in 50 mM sodium phosphate buffer (pH 7.0) at 37 ^{0}C. Reaction durations are depicted in individual experiments. During labeling of very large RNAs by CuAAC we recommend rather lower temperature. For labeling DNA, the reaction temperature can be increased up to 40^{o}C to facilitate the reaction rate. We strongly recommend avoiding buffers containing primary amines. The reaction can be alternatively performed in acetate-based buffers. For a nucleic acid end concentration of 1-10 µM in the CuAAC reaction mixture, 500 µM of Cu(II) and 60-100 µM of the other reaction partner generally leads to quantitative labeling. Obviously the reaction rate can be improved using

one component in high excess. Addition of DMSO or any other donor solvents leads to a decrease in CuAAC labeling efficiency when THPTA is used as a ligand. Generally the oligonucleotide after CuAAC labeling is either directly analyzed on a gel or immediately purified. In case of direct gel analysis of the reaction mixture, equal volume of formamide gel-loading buffer was added.

7.7 General protocol for *in vitro* transcription

Table 7.3: Protocol for transcriptional labeling.

Components	End concentration
DNA template	0.2 µM
Transcription buffer	1 X
DTT	10 mM
BSA	0.04 mg/ml
UTP/ATP/CTP	Individual experiments
GTP	Individual experiments
Initiator nucleotide	4 mM
α-^{32}P-CTP	0.8 µCi/µl
T3/SP6/T7 RNAP	1-2 u/µl

All components except the enzyme were mixed together and the mixture was heated to 70 ^{0}C for two minutes followed by cooling down to room temperature for 15 min to allow the template to anneal. Afterwards the polymerase was added and the reaction mixture was incubated for 2 hrs followed by either removal of enzyme by phenol-ether extraction and ethanol-precipitation of the transcript from 0.5 M NH_4OAc buffer with the help of glycogen as carrier to an end concentration of 0.2 µg/µl or direct purification of the transcript by preparative gel electrophoresis. In addition to the aforesaid annealing protocol, synthetic DNA templates were annealed prior to the transcription in hybridization buffer (10 mM Tris-HCl pH 7.5 with 80 mM NaCl). This additional annealing step was omitted when PCR product was used as transcription template. Generally the DNA template was added at the end to the transcription reaction mixture and the whole mixture containing everything was never placed in ice to avoid spermidine facilitated precipitation of template DNA.

Table 7.4: Composition of 1X transcription buffer.

Components	End concentrations
Tris-HCl buffer pH 8.1 at 25 °C	40 mM
Spermidine	1 mM
$MgCl_2$	22 mM
Triton-X-100	0.01 %

Table 7.5: Sequences used as transcription templates

Sequence	Features
5´-<u>TCTAATACGACTCACTATA</u>GGAGCTCAGCCTACGAGCCTGAGCC-3´	T7 template sense
5´-GGCTCAGGCTCGTAGGCTGAGCTCCTATAGTGAGTCGTATTAGA-3´	T7 template antisense
5´-<u>TCAATTAACCCTCACTAAA</u>GGGAGACAGCCTACGAGCCTGAGCC-3´	T3 template sense
5´-GGCTCAGGCTCGTAGGCTGTCTCCCTTTAGTGAGGGTTAATTGA-3´	T3 template antisense
5´-<u>TCATTTAGGTGACACTATA</u>GAAGAGCAGCCTACGAGCCTGAGCC-3´	SP6 template sense
5´-GGCTCAGGCTCGTAGGCTGCTCTTCTATAGTGTCACCTAAATGA-3´	SP6 template antisense
5´-TCTAATACGACTCACTATA-3´	T7 template sense
5´-AGATTATGCTGAGTGATATCCAGTAGTCGGAAGTGA**GG**-3´	T7 template antisense methoxy*

*The two bold G´s are carrying the 2´-methoxy modification

7.8 General protocol for quantitative biotinylation of alkyne bearing transcripts

7.8.1 For transcripts carrying ethynyl moiety

Table 7.6: Components of CuAAC labeling for ethynyl modified transcript.

Components	End concentration
Azide component	0.75 mM
$CuSO_4$	0.5 mM
Sodium ascorbate	5 mM
THPTA	2.5 mM
Sodium phosphate buffer pH 7.0	50 mM

All components were mixed together and the reaction mixture was incubated for 2 hr at 37 ^{0}C. Afterwards the labeled transcript was purified by ethanol precipitation with glycogen as a carrier. This protocol allows quantitative labeling at as low as 5 nM transcript concentrations. The use of such high excess of the azide counterpart ensures quantitative labeling even in the case of some extent of co-precipitation of the initiator dinucleotide with the transcript.

7.8.2 For transcripts carrying octadiynyl moiety

Table 7.7: Components of CuAAC labeling for octadiynyl modified transcript.

Components	End concentration
Azide component	0.25 mM
$CuSO_4$	0.5 mM
Sodium ascorbate	5 mM
THPTA	2.5 mM
Sodium phosphate buffer pH 7.0	50 mM

The long flexible alkyne moiety is sterically much more accessible compared to the previously described ethynyl moiety thereby requiring rather low concentrations of the azide component for quantitative functionalization.

7.9 General organic synthesis methods

Unless otherwise stated, all chemicals and dry solvents were purchased from commercial providers and used directly without any further manipulation. TLC was performed on Macherey-Nagel pre-coated POLYGRAM SILG/UV-254 plates and visualized using UV light. Column chromatography was performed using silica gel (0.04 mm, 60 Å, Baker). Organic extracts were dried over anhydrous sodium sulphate. Solvents were evaporated with a rotary evaporator under reduced pressure at ≤ 35 ^{0}C (unless otherwise stated). All organic compounds are vacuum dried at room temperature. All compounds bearing aryl azide are strictly protected from light and have been subjected to temperatures strictly not higher than 40 ^{0}C during any kinds of synthetic handling. Aromatic azides are potentially explosive. NMR spectra were recorded on a Varian-300 MHz spectrometer or Varian-500 MHz spectrometer. EI and FAB mass spectra were recorded using Finnigan MAT 8200 and MALDI were recorded using Bruker BiFlex III. HR-ESI was recorded using - Bruker micrOTOF-Q II. FT-IR was recorded using JASCO FT/IR 4100 spectrometer.

7.10 General procedure for the selective esterification of unprotected amino acids

A solution of the amino acid in the corresponding alcohol (3 ml of alcohol per mmol of the amino acid) was cooled to 0^{0}C followed by the dropwise addition of thionyl chloride (3 equivalents with respect to the amino acid). The resulting mixture was stirred at 0^{0}C for another 5 min followed by refluxing until the completion of the reaction (TLC). The crude reaction mixture was directly evaporated in a rotary evaporator and dried under vacuum

overnight. The resulting solid (hydrochloride salt of the esterified amino acid) does not require any further purification and can be directly used in the next step.

7.11 General procedure for ^{t}butyloxycarbonyl protection

To a solution of the amine in 1:1 water-dioxane (4 ml of each solvent per mmol of the amine) was added 3 equivalents of DIEA dropwise at 0 ^{0}C. The ice-bath was removed after 5 min followed by the dropwise addition of Boc_2O at room temperature. The reaction mixture was left stirring at room temperature until finished (TLC, generally overnight) followed by the addition of 0.05 M HCl until just neutral. Dioxane was removed using rotary evaporator and the aqueous part was partitioned with DCM or chloroform followed by washing with 0.05 M HCl to remove traces of DIEA. The organic layer was evaporated and dried in vacuum overnight. In most of the cases, the crude product does not require any further purification and contains only *tert*-butanol, therefore can be used directly in the next step.

7.12 General procedure for water soluble carbodiimide (EDC.HCl) mediated coupling

1.5-2.0 equivalents of EDC.HCl and 0.1-0.2 equivalents of DMAP were added to a pyridine (5 ml per mmol of the carboxylic acid) solution of the carboxylic acid reaction partner and the resulting mixture was stirred for *ca* 30 min. A visible change in the appearance of the solution generally indicates the successful activation of the carboxylic acid. A pyridine (3 ml per mmol of the alcohol or amine) solution of the amine or the alcohol reactant was added drop-wise to the aforesaid activated carboxylic acid solution at room temperature and the resulting reaction mixture was left stirring at room temperature until the completion of the reaction (TLC). The crude reaction mixture was evaporated to remove the pyridine and redissolved in DCM or $CHCl_3$ followed by washing with saturated NH_4HCO_3 solution. The organic layer was dried, evaporated and purified by flash column chromatography.

Alternatively this coupling reaction can be performed in DCM or $CHCl_3$. EDC.HCl is preferred over EDC because of its higher stability for long term storage. Instead of using NH_4HCO_3 solution, $NaHCO_3$ solution can also be used, however this did not lead to the best results at our hands.

7.13 General procedure for lithium borohydride reduction of esters to the corresponding alcohols

To a solution of the ester in anhydrous THF (0.5 ml of THF per mmol of the ester) was added 1.6 equivalent of a 2.0 M solution of lithium borohydride in THF and the resulting reaction mixture was left stirring at room temperature until completion (TLC). Afterwards the reaction mixture was cooled in ice and slowly quenched by 50 % aqueous acetic acid. The resulting mixture was diluted with water, followed by the partitioning with DCM. The resultant organic layer was occasionally washed with very dilute aqueous acetic acid, evaporated and dried under vacuum. The crude product, if required, can further be purified by flash column chromatography.

7.14 General procedure for saponification

To a solution of the ester in methanol (5 ml of the solvent per mmol of the ester) was added 3 equivalents of 1.0 M NaOH followed by stirring at room temperature until the TLC reveals complete consumption of staring material. The reaction mixture was neutralized by the addition of equal amount of 1.0 M HCl followed by cooling in ice. This generally leads to the precipitation of the saponified acid, which can be filtered and washed with ice cold water. The resulting solid has to be thoroughly dried under vacuum overnight and does not require any further purification. Alternatively when no precipitation occurs, methanol should be evaporated from the reaction mixture and the product can be recovered by partitioning with suitable organic solvents. Care should be taken during neutralization. Addition of concentrated as well as excess acid should be avoided to prevent precipitation of high amounts of NaCl. Little amounts of NaCl can be removed during washing the precipitated acid with ice cold water. However many carboxylic acids are partly soluble in water.

7.15 General procedure for Boc deprotection

TFA (0.6 ml per mmol of the substrate) was added to a solution of the Boc-protected substrate in DCM or $CHCl_3$ (5 ml of the solvent per mmol of the substrate) at 0^0C. The resulting mixture was left stirring at 0^0C for another 5 min and then continued stirring at room temperature until TLC revealed complete consumption of starting material. The reaction mixture was evaporated in a rotary evaporator and thoroughly vacuum dried overnight. This dried product can be directly used in any further application. If required, the resulting deprotected amine can be partitioned using a 4 to 1 $CHCl_3$-isopropanol or 8 to 2 DCM-methanol mixture as organic solvent.

7.16 Analytical data of synthesized molecules

Ethyl 2-aminopent-4-ynoate (4)

^{1}H NMR (300 MHz, CD_3OD)

δ 1.34 (t, 3H), 2.67 (t, 1H), 2.84 - 3.00 (m, 2H), 4.24 – 4.29 (m, 1H), 4.30 – 4.39 (m, 2H)

^{13}C NMR (300 MHz, CD_3OD)

δ 14.41, 21.30, 52.65, 64.05, 75.26, 76.86

FAB-MS

$[M+3H]^+ = C_7H_1NO_2$ m/z 143.2

4-azidobenzoic acid (5)

Synthesized according to previously published procedure(272)

Ethyl 2-(4-azidobenzamido)pent-4-ynoate (6)

^{1}H NMR (300 MHz, CD_3OD)

δ 1.28 (t, 3H), 2.39 (t, 1H), 2.75 – 2.92 (m, 2H), 4.19 – 4.2 (dq, 2H), 4.69 – 4.74 (m, 1H), 7.15 - 7.19 (m, 2H), 7.87 - 7.92 (m, 2H)

^{13}C NMR (300 MHz, CD_3OD)

δ 13.26, 20.86, 20.89, 52.36, 61.55, 70.99, 79.03, 118.83, 129.26, 130.34, 144.05, 167.98, 170.72

FAB-MS

$[M+H]^+ = C_{14}H_{14}N_4O_3$ m/z 287.1

2-(4-azidobenzamido)pent-4-ynoic acid (1)

^{1}H NMR (300 MHz, CD_3OD)

δ 2.36 (t, 1H), 2.76 – 2.94 (m, 2H), 4.70 – 4.75 (m, 1H), 7.14 - 7.19 (m, 2H), 7.88 - 7.93 (m, 2H)

^{13}C NMR (300 MHz, CD_3OD)

δ 21.10, 52.19, 70.77, 79.31, 118.82, 129.22, 130.52, 143.96, 167.86, 172.49

ESI-MS

$[M+Na]^+$ = m/z 281.06425

Ethyl 2-(*tert*-butoxycarbonylamino)pent-4-ynoate (7)

^{1}H NMR (300 MHz, CD_3OD)

δ 1.28(t, 3H), 1.45 (s, 9H), 1.52 (s, 3H), 2.36 (t, 1H), 2.64 – 2.68 (m, 2H), 4.12 – 4.30 (m, 3H)

^{13}C NMR (300 MHz, CD_3OD)

δ 14.53, 22.70, 27.62, 28.71, 54.06, 62.61, 68.17, 72.27, 80.05, 80.86, 86.43, 148.41, 157.69, 172.39

FAB-MS

$[M+1H]^+ = C_{12}H_{19}NO_4$ m/z 242.2

***tert*-butyl 1-hydroxypent-4-yn-2-ylcarbamate (8)**

^{1}H NMR (500 MHz, CD_3OD)

δ 1.44(s, 9H), 2.28 (s, 1H), 2.34 – 2.47 (m, 2H), 3.55 – 3.61 (m, 2H), 3.65 – 3.69 (m, 1H)

^{13}C NMR (500 MHz, CD_3OD)

δ 21.80, 28.77, 52.80, 54.84, 63.69, 71.29, 80.29, 81.55, 157.97

FAB-MS

$[M+3H]^+ = C_{10}H_{17}NO_3$ m/z 202.1

2-(*tert*-butoxycarbonylamino)pent-4-ynyl 4-azidobenzoate (9)

^{1}H NMR (300 MHz, CD_3OD)

δ 1.42 (s, 9H), 1.52 (s, 3H), 2.37 (t, 1H), 2.49 – 2.52 (m, 2H), 4.00 – 4.10 (m, 1H), 4.28 – 4.44 (m, 2H), 7.15 - 7.18 (m, 2H), 8.05 – 8.08 (m, 2H)

2-aminopent-4-ynyl 4-azidobenzoate (2)

^{1}H NMR (300 MHz, CD_3OD)

δ 1.56 (s, 1H), 2.65 (t, 1H), 2.77 – 2.80 (m, 2H), 3.78 - 3.85 (m, 1H), 4.50 – 4.65 (dq, 2H), 7.16 - 7.20 (m, 2H), 8.10 – 8.15 (m, 2H)

^{13}C NMR (300 MHz, CD_3OD)

δ 20.59, 27.77, 50.37, 64.65, 74.56, 77.86, 120.15, 126.92, 132.91, 147.07, 166.62

FAB-MS

$[M+3H]^+ = C_{12}H_{12}N_4O_2$ m/z 247.1

2-aminopent-4-yn-1-ol (10)

^{1}H NMR (300 MHz, CD_3OD)

δ 2.54 (t, 1H), 2.59 – 2.63 (m, 2H), 3.33 – 3.40 (m, 1H), 3.67 – 3.84 (m, 2H)

FAB-MS

$[M+3H]^+ = C_5H_9NO$ m/z 102.1

4-azido-N-(1-hydroxypent-4-yn-2-yl)benzamide (3)

^{1}H NMR (300 MHz, CD_3OD)

δ 2.31 (t, 1H), 2.47 – 2.65 (m, 2H), 3.72 - 3.74 (m, 2H), 4.14 – 4.26 (m, 1H), 7.12 - 7.15 (m, 2H), 7.86 – 7.89 (m, 2H)

^{13}C NMR (300 MHz, CD_3OD)

δ 21.42, 52.54, 63.56, 71.42, 81.56, 119.95, 130.38, 132.26, 132.87, 144.92, 169.30, **FAB-MS**

$[M+1H]^+ = C_{12}H_{12}N_4O_2$ m/z 244.1

2-(4-azidobenzamido)pent-4-ynyl 2-aminoacetate (12)

^{1}H NMR (300 MHz, CD_3OD)

δ 1.57 (s, 2H), 2.41 (t, 1H), 2.59 – 2.63 (m, 2H), 3.86 (s, 2H), 4.33 – 4.38 (m, 1H), 4.47 – 4.58 (m, 2H), 7.14 - 7.18 (m, 2H), 7.86 – 7.88 (m, 2H)

^{13}C NMR (300 MHz, CD_3OD)

δ 21.63, 26.11, 26.79, 27.77, 34.80, 40.97, 67.46, 72.20, 80.52, 120.04, 130.44, 130.48, 131.85, 145.26, 168.58, 169.42,

FAB-MS

$[M+3H]^+ = C_{14}H_{15}N_5O_3$ m/z 304.1

7.17 General protocol for *in vitro* photoaffinity tagging

Table 7.8: Protocol for photoaffinity tagging experiments.

Components	End concentration
Lysine riboswitch or total RNA	*ca* 1 µM or 68 ng/µl
Lysine trifunctional ester probe (**13**)	5-200 eqv. as depicted in each expt´s
$MgCl_2$	5 mM
Sodium phosphate buffer pH 7.0	50 mM

All components were mixed together in a proportion as depicted in table 7.8. RNA folding was performed by heating the reaction mixture to 70 ^{0}C for two minutes followed by cooling down to room temperature for 15 min. Each of the samples was divided into equal halves and only one half of the sample was subjected to photoaffinity tagging by the irradiation with 254 nm UV-light for 15 min at 2-4 ^{0}C without any polystyrene filter. Afterwards, all samples were diluted with water followed by a rigorous extraction of the aqueous layer with aq-phenol, BuOH-EDTA and diethylether. The crosslinked RNA was ethanol-precipitated from a 0.5 M NH_4OAc buffer with the help of glycogen as a carrier at and end concentration of 0.8 µg/µl. This was directly followed by a CuAAC conjugation with biotin azide to quantitatively introduce a biotinyl residue for calculating the photo-tagging yield. The same CuAAC conditions, as described before for quantitative biotinylation of transcripts carrying octadiynyl residues, can be used for this purpose.

8 References

1. Ferre-D'Amare, A.R. and Winkler, W.C. (2011) The roles of metal ions in regulation by riboswitches. *Met Ions Life Sci*, **9**, 141-173.
2. Li, P.T., Vieregg, J. and Tinoco, I., Jr. (2008) How RNA unfolds and refolds. *Annu Rev Biochem*, **77**, 77-100.
3. Serganov, A. and Patel, D.J. (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat Rev Genet*, **8**, 776-790.
4. Breaker, R.R. Prospects for riboswitch discovery and analysis. *Mol Cell*, **43**, 867-879.
5. Henkin, T.M. (2008) Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes Dev*, **22**, 3383-3390.
6. Montange, R.K. and Batey, R.T. (2008) Riboswitches: emerging themes in RNA structure and function. *Annu Rev Biophys*, **37**, 117-133.
7. Roth, A. and Breaker, R.R. (2009) The structural and functional diversity of metabolite-binding riboswitches. *Annu Rev Biochem*, **78**, 305-334.
8. Serganov, A. (2009) The long and the short of riboswitches. *Curr Opin Struct Biol*, **19**, 251-259.
9. Winkler, W.C. and Breaker, R.R. (2005) Regulation of bacterial gene expression by riboswitches. *Annu Rev Microbiol*, **59**, 487-517.
10. Bastet, L., Dube, A., Masse, E. and Lafontaine, D.A. (2011) New insights into riboswitch regulation mechanisms. *Mol Microbiol*, **80**, 1148-1154.
11. Mandal, M. and Breaker, R.R. (2004) Gene regulation by riboswitches. *Nat Rev Mol Cell Biol*, **5**, 451-463.
12. Smith, A.M., Fuchs, R.T., Grundy, F.J. and Henkin, T.M. (2010) Riboswitch RNAs: regulation of gene expression by direct monitoring of a physiological signal. *RNA Biol*, **7**, 104-110.
13. Cheah, M.T., Wachter, A., Sudarsan, N. and Breaker, R.R. (2007) Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature*, **447**, 497-500.
14. Thore, S., Leibundgut, M. and Ban, N. (2006) Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. *Science*, **312**, 1208-1211.
15. Nudler, E. (2006) Flipping riboswitches. *Cell*, **126**, 19-22.
16. Winkler, W.C. (2005) Riboswitches and the role of noncoding RNAs in bacterial metabolic control. *Curr Opin Chem Biol*, **9**, 594-602.
17. Mandal, M., Boese, B., Barrick, J.E., Winkler, W.C. and Breaker, R.R. (2003) Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. *Cell*, **113**, 577-586.
18. Houlberg, U. and Jensen, K.F. (1983) Role of hypoxanthine and guanine in regulation of Salmonella typhimurium pur gene expression. *J Bacteriol*, **153**, 837-845.
19. Wang, J.X. and Breaker, R.R. (2008) Riboswitches that sense S-adenosylmethionine and S-adenosylhomocysteine. *Biochem Cell Biol*, **86**, 157-168.
20. Saint-Girons, I., Belfaiza, J., Guillou, Y., Perrin, D., Guiso, N., Barzu, O. and Cohen, G.N. (1986) Interactions of the Escherichia coli methionine repressor with the metF operator and with its corepressor, S-adenosylmethionine. *J Biol Chem*, **261**, 10936-10940.
21. Poiata, E., Meyer, M.M., Ames, T.D. and Breaker, R.R. (2009) A variant riboswitch aptamer class for S-adenosylmethionine common in marine bacteria. *RNA*, **15**, 2046-2056.
22. Roy, R., Hohng, S. and Ha, T. (2008) A practical guide to single-molecule FRET. *Nat Methods*, **5**, 507-516.
23. Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C. and Ha, T. (2008) Advances in single-molecule fluorescence methods for molecular biology. *Annu Rev Biochem*, **77**, 51-76.
24. Dammertz, K., Hengesbach, M., Helm, M., Nienhaus, G.U. and Kobitski, A.Y. (2011) Single-molecule FRET studies of counterion effects on the free energy landscape of human mitochondrial lysine tRNA. *Biochemistry*, **50**, 3107-3115.
25. Hengesbach, M., Kobitski, A., Voigts-Hoffmann, F., Frauer, C., Nienhaus, G.U. and Helm, M. (2008) RNA intramolecular dynamics by single-molecule FRET. *Curr Protoc Nucleic Acid Chem*, **Chapter 11**, Unit 11 12.
26. Kobitski, A.Y., Hengesbach, M., Helm, M. and Nienhaus, G.U. (2008) Sculpting an RNA conformational energy landscape by a methyl group modification--a single-molecule FRET study. *Angewandte Chemie (International ed*, **47**, 4326-4330.

27. Kobitski, A.Y., Hengesbach, M., Seidu-Larry, S., Dammertz, K., Chow, C.S., van Aerschot, A., Nienhaus, G.U. and Helm, M. Single-molecule FRET reveals a cooperative effect of two methyl group modifications in the folding of human mitochondrial tRNA(Lys). *Chem Biol*, **18**, 928-936.
28. Kobitski, A.Y., Nierth, A., Helm, M., Jäschke, A. and Nienhaus, G.U. (2007) Mg2+-dependent folding of a Diels-Alderase ribozyme probed by single-molecule FRET analysis. *Nucleic Acids Res*, **35**, 2047-2059.
29. Voigts-Hoffmann, F., Hengesbach, M., Kobitski, A.Y., van Aerschot, A., Herdewijn, P., Nienhaus, G.U. and Helm, M. (2007) A methyl group controls conformational equilibrium in human mitochondrial tRNA(Lys). *J Am Chem Soc*, **129**, 13382-13383.
30. Beaucage, S.L. and Iyer, R.P. (1992) Advances in the Synthesis of Oligonucleotides by the Phosphoramidite Approach. *Tetrahedron*, **48**, 2223-2311.
31. Beaucage, S.L. and Iyer, R.P. (1993) The Synthesis of Modified Oligonucleotides by the Phosphoramidite Approach and Their Applications. *Tetrahedron*, **49**, 6123-6194.
32. Beaucage, S.L. and Iyer, R.P. (1993) The Functionalization of Oligonucleotides Via Phosphoramidite Derivatives. *Tetrahedron*, **49**, 1925-1963.
33. Beaucage, S.L. and Iyer, R.P. (1993) The Synthesis of Specific Ribonucleotides and Unrelated Phosphorylated Biomolecules by the Phosphoramidite Method. *Tetrahedron*, **49**, 10441-10488.
34. Lonnberg, H. (2009) Solid-phase synthesis of oligonucleotide conjugates useful for delivery and targeting of potential nucleic acid therapeutics. *Bioconjug Chem*, **20**, 1065-1094.
35. Dolinnaya, N.G., Sokolova, N.I., Ashirbekova, D.T. and Shabarova, Z.A. (1991) The use of BrCN for assembling modified DNA duplexes and DNA-RNA hybrids; comparison with water-soluble carbodiimide. *Nucleic Acids Res*, **19**, 3067-3072.
36. Mitra, D. and Damha, M.J. (2007) A novel approach to the synthesis of DNA and RNA lariats. *J Org Chem*, **72**, 9491-9500.
37. Carriero, S. and Damha, M.J. (2003) Template-mediated synthesis of lariat RNA and DNA. *J Org Chem*, **68**, 8328-8338.
38. Dolinnaya, N.G., Tsytovich, A.V., Sergeev, V.N., Oretskaya, T.S. and Shabarova, Z.A. (1991) Structural and kinetic aspects of chemical reactions in DNA duplexes. Information on DNA local structure obtained from chemical ligation data. *Nucleic Acids Res*, **19**, 3073-3080.
39. Dolinnaya, N.G., Sokolova, N.I., Gryaznova, O.I. and Shabarova, Z.A. (1988) Site-directed modification of DNA duplexes by chemical ligation. *Nucleic Acids Res*, **16**, 3721-3738.
40. Paredes, E., Evans, M. and Das, S.R. (2011) RNA labeling, conjugation and ligation. *Methods*, **54**, 251-259.
41. Moore, M.J. and Query, C.C. (2000) Joining of RNAs by splinted ligation. *Methods Enzymol*, **317**, 109-123.
42. Kurschat, W.C., Muller, J., Wombacher, R. and Helm, M. (2005) Optimizing splinted ligation of highly structured small RNAs. *RNA*, **11**, 1909-1914.
43. Bullard, D.R. and Bowater, R.P. (2006) Direct comparison of nick-joining activity of the nucleic acid ligases from bacteriophage T4. *Biochem J*, **398**, 135-144.
44. Stark, M.R., Pleiss, J.A., Deras, M., Scaringe, S.A. and Rader, S.D. (2006) An RNA ligase-mediated method for the efficient creation of large, synthetic RNAs. *RNA*, **12**, 2014-2019.
45. Nishigaki, K., Taguchi, K., Kinoshita, Y., Aita, T. and Husimi, Y. (1998) Y-ligation: An efficient method for ligating single-stranded DNAs and RNAs with T4 RNA ligase. *Molecular Diversity*, **4**, 187-190.
46. Wang, L. and Ruffner, D.E. (1998) Oligoribonucleotide circularization by 'template-mediated' ligation with T4 RNA ligase: synthesis of circular hammerhead ribozymes. *Nucleic Acids Res*, **26**, 2502-2504.
47. Imai, M., Richardson, M.A., Ikegami, N., Shatkin, A.J. and Furuichi, Y. (1983) Molecular cloning of double-stranded RNA virus genomes. *Proc Natl Acad Sci U S A*, **80**, 373-377.
48. Lambden, P.R., Cooke, S.J., Caul, E.O. and Clarke, I.N. (1992) Cloning of noncultivatable human rotavirus by single primer amplification. *J Virol*, **66**, 1817-1822.
49. Wachowius, F. and Hobartner, C. Chemical RNA modifications for studies of RNA structure and dynamics. *Chembiochem*, **11**, 469-480.
50. Mayer, G. and Heckel, A. (2006) Biologically active molecules with a "light switch". *Angewandte Chemie (International ed*, **45**, 4900-4921.
51. Weil, T.T., Parton, R.M. and Davis, I. Making the message clear: visualizing mRNA localization. *Trends Cell Biol*, **20**, 380-390.
52. Shim, M.S. and Kwon, Y.J. Efficient and targeted delivery of siRNA in vivo. *Febs J*, **277**, 4814-4827.

53. Schelhaas, M. and Waldmann, H. (1996) Protecting group strategies in organic synthesis. *Angewandte Chemie-International Edition in English*, **35**, 2056-2083.
54. Tietze, L.F. (1996) Domino reactions in organic synthesis. *Chemical Reviews*, **96**, 115-136.
55. Grosjean, H. (2009) DNA and RNA Modification Enzymes.
56. Padilla, R. and Sousa, R. (2002) A Y639F/H784A T7 RNA polymerase double mutant displays superior properties for synthesizing RNAs with non-canonical NTPs. *Nucleic Acids Res*, **30**, e138.
57. Sousa, R. and Padilla, R. (1995) A mutant T7 RNA polymerase as a DNA polymerase. *Embo J*, **14**, 4609-4621.
58. Ricchetti, M. and Buc, H. (1993) E. coli DNA polymerase I as a reverse transcriptase. *Embo J*, **12**, 387-396.
59. Van de Sande, J.H., Loewen, P.C. and Khorana, H.G. (1972) Studies on polynucleotides. 118. A further study of ribonucleotide incorporation into deoxyribonucleic acid chains by deoxyribonucleic acid polymerase I of Escherichia coli. *J Biol Chem*, **247**, 6140-6148.
60. Tabor, S. and Richardson, C.C. (1989) Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage T7 DNA polymerase and Escherichia coli DNA polymerase I. *Proc Natl Acad Sci U S A*, **86**, 4076-4080.
61. Roychoudhury, R., Jay, E. and Wu, R. (1976) Terminal labeling and addition of homopolymer tracts to duplex DNA fragments by terminal deoxynucleotidyl transferase. *Nucleic Acids Res*, **3**, 863-877.
62. Sano, H. and Feix, G. (1974) Ribonucleic acid ligase activity of deoxyribonucleic acid ligase from phage T4 infected Escherichia coli. *Biochemistry*, **13**, 5110-5115.
63. Hermanson, G.T. (2008) Bioconjugate Techniques.
64. Kolb, H.C., Finn, M.G. and Sharpless, K.B. (2001) Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angewandte Chemie (International ed*, **40**, 2004-2021.
65. Orski, S.V., Poloukhtine, A.A., Arumugam, S., Mao, L., Popik, V.V. and Locklin, J. (2010) High density orthogonal surface immobilization via photoactivated copper-free click chemistry. *J Am Chem Soc*, **132**, 11024-11026.
66. Decreau, R.A., Collman, J.P. and Hosseini, A. (2010) Electrochemical applications. How click chemistry brought biomimetic models to the next level: electrocatalysis under controlled rate of electron transfer. *Chem Soc Rev*, **39**, 1291-1301.
67. Huisgen, R. (1989) Kinetics and Reaction-Mechanisms - Selected Examples from the Experience of 40 Years. *Pure and Applied Chemistry*, **61**, 613-628.
68. Ess, D.H. and Houk, K.N. (2008) Theory of 1,3-dipolar cycloadditions: distortion/interaction and frontier molecular orbital models. *J Am Chem Soc*, **130**, 10187-10198.
69. Tornoe, C.W., Christensen, C. and Meldal, M. (2002) Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J Org Chem*, **67**, 3057-3064.
70. Rostovtsev, V.V., Green, L.G., Fokin, V.V. and Sharpless, K.B. (2002) A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angewandte Chemie (International ed*, **41**, 2596-2599.
71. Kolb, H.C. and Sharpless, K.B. (2003) The growing impact of click chemistry on drug discovery. *Drug Discov Today*, **8**, 1128-1137.
72. Wang, Q., Chan, T.R., Hilgraf, R., Fokin, V.V., Sharpless, K.B. and Finn, M.G. (2003) Bioconjugation by copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J Am Chem Soc*, **125**, 3192-3193.
73. Wu, P., Malkoch, M., Hunt, J.N., Vestberg, R., Kaltgrad, E., Finn, M.G., Fokin, V.V., Sharpless, K.B. and Hawker, C.J. (2005) Multivalent, bifunctional dendrimers prepared by click chemistry. *Chem Commun (Camb)*, 5775-5777.
74. Chan, T.R., Hilgraf, R., Sharpless, K.B. and Fokin, V.V. (2004) Polytriazoles as copper(I)-stabilizing ligands in catalysis. *Org Lett*, **6**, 2853-2855.
75. Kalisiak, J., Sharpless, K.B. and Fokin, V.V. (2008) Efficient synthesis of 2-substituted-1,2,3-triazoles. *Org Lett*, **10**, 3171-3174.
76. Wu, P., Feldman, A.K., Nugent, A.K., Hawker, C.J., Scheel, A., Voit, B., Pyun, J., Frechet, J.M., Sharpless, K.B. and Fokin, V.V. (2004) Efficiency and fidelity in a click-chemistry route to triazole dendrimers by the copper(i)-catalyzed ligation of azides and alkynes. *Angewandte Chemie (International ed*, **43**, 3928-3932.
77. Yoo, E.J., Ahlquist, M., Kim, S.H., Bae, I., Fokin, V.V., Sharpless, K.B. and Chang, S. (2007) Copper-catalyzed synthesis of N-sulfonyl-1,2,3-triazoles: controlling selectivity. *Angewandte Chemie (International ed*, **46**, 1730-1733.

78. Yoo, E.J., Ahlquist, M., Bae, I., Sharpless, K.B., Fokin, V.V. and Chang, S. (2008) Mechanistic studies on the Cu-catalyzed three-component reactions of sulfonyl azides, 1-alkynes and amines, alcohols, or water: dichotomy via a common pathway. *J Org Chem*, **73**, 5520-5528.
79. Mamidyala, S.K. and Finn, M.G. (2010) In situ click chemistry: probing the binding landscapes of biological molecules. *Chem Soc Rev*, **39**, 1252-1261.
80. Amblard, F., Cho, J.H. and Schinazi, R.F. (2009) Cu(I)-catalyzed Huisgen azide-alkyne 1,3-dipolar cycloaddition reaction in nucleoside, nucleotide, and oligonucleotide chemistry. *Chem Rev*, **109**, 4207-4220.
81. El-Sagheer, A.H. and Brown, T. (2010) Click chemistry with DNA. *Chem Soc Rev*, **39**, 1388-1405.
82. Best, M.D. (2009) Click chemistry and bioorthogonal reactions: unprecedented selectivity in the labeling of biological molecules. *Biochemistry*, **48**, 6571-6584.
83. Gramlich, P.M.E., Wirges, C.T., Manetto, A. and Carell, T. (2008) Postsynthetic DNA Modification through the Copper-Catalyzed Azide-Alkyne Cycloaddition Reaction. *Angewandte Chemie-International Edition*, **47**, 8350-8358.
84. Le Droumaguet, C., Wang, C. and Wang, Q. (2010) Fluorogenic click reaction. *Chem Soc Rev*, **39**, 1233-1239.
85. Angell, Y.L. and Burgess, K. (2007) Peptidomimetics via copper-catalyzed azide-alkyne cycloadditions. *Chem Soc Rev*, **36**, 1674-1689.
86. Papavassiliou, A.G. (2009) Footprinting DNA-protein interactions in native polyacrylamide gels by chemical nucleolytic activity of 1,10-phenanthroline-copper. *Methods Mol Biol*, **543**, 163-199.
87. Chen, C.B., Milne, L., Landgraf, R., Perrin, D.M. and Sigman, D.S. (2001) Artificial nucleases. *Chembiochem*, **2**, 735-740.
88. Nagaraj, R.H., Sell, D.R., Prabhakaram, M., Ortwerth, B.J. and Monnier, V.M. (1991) High correlation between pentosidine protein crosslinks and pigmentation implicates ascorbate oxidation in human lens senescence and cataractogenesis. *Proc Natl Acad Sci U S A*, **88**, 10257-10261.
89. Liu, P.Y., Jiang, N., Zhang, J., Wei, X., Lin, H.H. and Yu, X.Q. (2006) The oxidative damage of plasmid DNA by ascorbic acid derivatives in vitro: The first research on the relationship between the structure of ascorbic acid and the oxidative damage of plasmid DNA. *Chemistry & Biodiversity*, **3**, 958-966.
90. Fry, S.C. (1998) Oxidative scission of plant cell wall polysaccharides by ascorbate-induced hydroxyl radicals. *Biochem J*, **332 (Pt 2)**, 507-515.
91. Agarwal, R. and Gupta, M.N. (1994) Copper Affinity Precipitation as an Initial Step in Protein Purification. *Biotechnology Techniques*, **8**, 655-658.
92. Uchida, K. and Kawakishi, S. (1986) Selective oxidation of imidazole ring in histidine residues by the ascorbic acid-copper ion system. *Biochem Biophys Res Commun*, **138**, 659-665.
93. Liu, Y., Sun, G., David, A. and Sayre, L.M. (2004) Model studies on the metal-catalyzed protein oxidation: structure of a possible His-Lys cross-link. *Chem Res Toxicol*, **17**, 110-118.
94. van Berkel, S.S., van Eldijk, M.B. and van Hest, J.C.M. (2011) Staudinger Ligation as a Method for Bioconjugation. *Angewandte Chemie-International Edition*, **50**, 8806-8827.
95. Kohn, M. and Breinbauer, R. (2004) The Staudinger ligation-a gift to chemical biology. *Angewandte Chemie (International ed*, **43**, 3106-3116.
96. Sen Gupta, S., Kuzelka, J., Singh, P., Lewis, W.G., Manchester, M. and Finn, M.G. (2005) Accelerated bioorthogonal conjugation: a practical method for the ligation of diverse functional molecules to a polyvalent virus scaffold. *Bioconjug Chem*, **16**, 1572-1579.
97. Rodionov, V.O., Presolski, S.I., Gardinier, S., Lim, Y.H. and Finn, M.G. (2007) Benzimidazole and related ligands for Cu-catalyzed azide-alkyne cycloaddition. *J Am Chem Soc*, **129**, 12696-12704.
98. Campbell-Verduyn, L.S., Mirfeizi, L., Dierckx, R.A., Elsinga, P.H. and Feringa, B.L. (2009) Phosphoramidite accelerated copper(i)-catalyzed [3 + 2] cycloadditions of azides and alkynes. *Chem Commun (Camb)*, 2139-2141.
99. Teyssot, M.L., Chevry, A., Traikia, M., El-Ghozzi, M., Avignant, D. and Gautier, A. (2009) Improved copper(I)-NHC catalytic efficiency on huisgen reaction by addition of aromatic nitrogen donors. *Chemistry*, **15**, 6322-6326.
100. Rodionov, V.O., Fokin, V.V. and Finn, M.G. (2005) Mechanism of the ligand-free CuI-catalyzed azide-alkyne cycloaddition reaction. *Angewandte Chemie (International ed*, **44**, 2210-2215.
101. Rodionov, V.O., Presolski, S.I., Diaz, D.D., Fokin, V.V. and Finn, M.G. (2007) Ligand-accelerated Cu-catalyzed azide-alkyne cycloaddition: a mechanistic report. *J Am Chem Soc*, **129**, 12705-12712.

102. Hong, V., Presolski, S.I., Ma, C. and Finn, M.G. (2009) Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angewandte Chemie (International ed*, **48**, 9879-9883.
103. Gierlich, J., Burley, G.A., Gramlich, P.M., Hammond, D.M. and Carell, T. (2006) Click chemistry as a reliable method for the high-density postsynthetic functionalization of alkyne-modified DNA. *Org Lett*, **8**, 3639-3642.
104. Gierlich, J., Gutsmiedl, K., Gramlich, P.M.E., Schmidt, A., Burley, G.A. and Carell, T. (2007) Synthesis of highly modified DNA by a combination of PCR with alkyne-bearing triphosphates and click chemistry. *Chemistry-a European Journal*, **13**, 9486-9494.
105. Gramlich, P.M.E., Wirges, C.T., Gierlich, J. and Carell, T. (2008) Synthesis of modified DNA by PCR with alkyne-bearing purines followed by a click reaction. *Organic Letters*, **10**, 249-251.
106. Burley, G.A., Gierlich, J., Mofid, M.R., Nir, H., Tal, S., Eichen, Y. and Carell, T. (2006) Directed DNA metallization. *Journal of the American Chemical Society*, **128**, 1398-1399.
107. Fischler, M., Sologubenko, A., Mayer, J., Clever, G., Burley, G., Gierlich, J., Carell, T. and Simon, U. (2008) Chain-like assembly of gold nanoparticles on artificial DNA templates via 'click chemistry'. *Chemical Communications*, 169-171.
108. Rozkiewicz, D.I., Gierlich, J., Burley, G.A., Gutsmiedl, K., Carell, T., Ravoo, B.J. and Reinhoudt, D.N. (2007) Transfer printing of DNA by "Click" chemistry. *Chembiochem*, **8**, 1997-2002.
109. Gaetke, L.M. and Chow, C.K. (2003) Copper toxicity, oxidative stress, and antioxidant nutrients. *Toxicology*, **189**, 147-163.
110. Prescher, J.A., Dube, D.H. and Bertozzi, C.R. (2004) Chemical remodelling of cell surfaces in living animals. *Nature*, **430**, 873-877.
111. Agard, N.J., Prescher, J.A. and Bertozzi, C.R. (2004) A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *Journal of the American Chemical Society*, **126**, 15046-15047.
112. Agard, N.J., Baskin, J.M., Prescher, J.A., Lo, A. and Bertozzi, C.R. (2006) A comparative study of bioorthogonal reactions with azides. *ACS Chem Biol*, **1**, 644-648.
113. Baskin, J.M., Prescher, J.A., Laughlin, S.T., Agard, N.J., Chang, P.V., Miller, I.A., Lo, A., Codelli, J.A. and Bertozzi, C.R. (2007) Copper-free click chemistry for dynamic in vivo imaging. *Proc Natl Acad Sci U S A*, **104**, 16793-16797.
114. Sletten, E.M. and Bertozzi, C.R. (2008) A hydrophilic azacyclooctyne for Cu-free click chemistry. *Org Lett*, **10**, 3097-3099.
115. Guo, J., Chen, G., Ning, X., Wolfert, M.A., Li, X., Xu, B. and Boons, G.J. (2010) Surface modification of polymeric micelles by strain-promoted alkyne-azide cycloadditions. *Chemistry*, **16**, 13360-13366.
116. Ning, X., Temming, R.P., Dommerholt, J., Guo, J., Ania, D.B., Debets, M.F., Wolfert, M.A., Boons, G.J. and van Delft, F.L. (2010) Protein modification by strain-promoted alkyne-nitrone cycloaddition. *Angewandte Chemie (International ed*, **49**, 3065-3068.
117. Ning, X., Guo, J., Wolfert, M.A. and Boons, G.J. (2008) Visualizing metabolically labeled glycoconjugates of living cells by copper-free and fast huisgen cycloadditions. *Angewandte Chemie (International ed*, **47**, 2253-2255.
118. Blackman, M.L., Royzen, M. and Fox, J.M. (2008) Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity. *J Am Chem Soc*, **130**, 13518-13519.
119. Clavier, G. and Audebert, P. (2010) s-Tetrazines as building blocks for new functional molecules and molecular materials. *Chem Rev*, **110**, 3299-3314.
120. Devaraj, N.K., Hilderbrand, S., Upadhyay, R., Mazitschek, R. and Weissleder, R. (2010) Bioorthogonal turn-on probes for imaging small molecules inside living cells. *Angewandte Chemie (International ed*, **49**, 2869-2872.
121. Devaraj, N.K., Upadhyay, R., Haun, J.B., Hilderbrand, S.A. and Weissleder, R. (2009) Fast and sensitive pretargeted labeling of cancer cells through a tetrazine/trans-cyclooctene cycloaddition. *Angewandte Chemie (International ed*, **48**, 7013-7016.
122. Devaraj, N.K., Weissleder, R. and Hilderbrand, S.A. (2008) Tetrazine-based cycloadditions: application to pretargeted live cell imaging. *Bioconjug Chem*, **19**, 2297-2299.
123. Schoch, J., Wiessler, M. and Jäschke, A. (2010) Post-synthetic modification of DNA by inverse-electron-demand Diels-Alder reaction. *J Am Chem Soc*, **132**, 8846-8847.
124. Schoch, J., Ameta, S. and Jäschke, A. (2011) Inverse electron-demand Diels-Alder reactions for the selective and efficient labeling of RNA. *Chem Commun (Camb)*.
125. van Der Wel, G.K., Wijnen, J.W. and Engberts, J.B. (1996) Solvent Effects on a Diels-Alder Reaction Involving a Cationic Diene: Consequences of the Absence of Hydrogen-Bond Interactions for Accelerations in Aqueous Media. *J Org Chem*, **61**, 9001-9005.

126. Thalhammer, F., Wallfahrer, U. and Sauer, J. (1990) Reactivity of Simple Open-Chain and Cyclic Dienophiles in Inverse-Type Diels-Alder Reactions. *Tetrahedron Letters*, **31**, 6851-6854.
127. Han, H.S., Devaraj, N.K., Lee, J., Hilderbrand, S.A., Weissleder, R. and Bawendi, M.G. Development of a bioorthogonal and highly efficient conjugation method for quantum dots using tetrazine-norbornene cycloaddition. *J Am Chem Soc*, **132**, 7838-7839.
128. Haun, J.B., Devaraj, N.K., Hilderbrand, S.A., Lee, H. and Weissleder, R. (2010) Bioorthogonal chemistry amplifies nanoparticle binding and enhances the sensitivity of cell detection. *Nat Nanotechnol*, **5**, 660-665.
129. Haun, J.B., Devaraj, N.K., Marinelli, B.S., Lee, H. and Weissleder, R. (2011) Probing intracellular biomarkers and mediators of cell activation using nanosensors and bioorthogonal chemistry. *ACS Nano*, **5**, 3204-3213.
130. Haun, J.B., Castro, C.M., Wang, R., Peterson, V.M., Marinelli, B.S., Lee, H. and Weissleder, R. (2011) Micro-NMR for rapid molecular analysis of human tumor samples. *Sci Transl Med*, **3**, 71ra16.
131. Reiner, T., Earley, S., Turetsky, A. and Weissleder, R. (2010) Bioorthogonal small-molecule ligands for PARP1 imaging in living cells. *Chembiochem*, **11**, 2374-2377.
132. Wieland, M. and Hartig, J.S. (2008) Artificial riboswitches: synthetic mRNA-based regulators of gene expression. *Chembiochem*, **9**, 1873-1878.
133. Wieland, M. and Hartig, J.S. (2008) Improved aptazyme design and in vivo screening enable riboswitching in bacteria. *Angewandte Chemie (International ed*, **47**, 2604-2607.
134. Barrick, J.E., Corbino, K.A., Winkler, W.C., Nahvi, A., Mandal, M., Collins, J., Lee, M., Roth, A., Sudarsan, N., Jona, I. *et al.* (2004) New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc Natl Acad Sci U S A*, **101**, 6421-6426.
135. Weinberg, Z., Barrick, J.E., Yao, Z., Roth, A., Kim, J.N., Gore, J., Wang, J.X., Lee, E.R., Block, K.F., Sudarsan, N. *et al.* (2007) Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res*, **35**, 4809-4819.
136. Barrick, J.E. and Breaker, R.R. (2007) The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol*, **8**, R239.
137. Winkler, W., Nahvi, A. and Breaker, R.R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, **419**, 952-956.
138. Serganov, A., Polonskaia, A., Phan, A.T., Breaker, R.R. and Patel, D.J. (2006) Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature*, **441**, 1167-1171.
139. Lorenz, C., von Pelchrzim, F. and Schroeder, R. (2006) Genomic systematic evolution of ligands by exponential enrichment (Genomic SELEX) for the identification of protein-binding RNAs independent of their expression levels. *Nat Protoc*, **1**, 2204-2212.
140. Lorenz, C., Gesell, T., Zimmermann, B., Schoeberl, U., Bilusic, I., Rajkowitsch, L., Waldsich, C., von Haeseler, A. and Schroeder, R. (2010) Genomic SELEX for Hfq-binding RNAs identifies genomic aptamers predominantly in antisense transcripts. *Nucleic Acids Res*, **38**, 3794-3808.
141. Zimmermann, B., Bilusic, I., Lorenz, C. and Schroeder, R. (2010) Genomic SELEX: a discovery tool for genomic aptamers. *Methods*, **52**, 125-132.
142. Huttenhofer, A. and Vogel, J. (2006) Experimental approaches to identify non-coding RNAs. *Nucleic Acids Res*, **34**, 635-646.
143. Terpe, K. (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol*, **60**, 523-533.
144. Srisawat, C. and Engelke, D.R. (2001) Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. *Rna*, **7**, 632-641.
145. Vodovozova, E.L. (2007) Photoaffinity labeling and its application in structural biology. *Biochemistry (Mosc)*, **72**, 1-20.
146. Brunner, J. (1993) New photolabeling and crosslinking methods. *Annu Rev Biochem*, **62**, 483-514.
147. Kishore, S., Jaskiewicz, L., Burger, L., Hausser, J., Khorshid, M. and Zavolan, M. (2011) A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat Methods*, **8**, 559-564.
148. Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M. and Ule, J. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol*, **17**, 909-915.

149. Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M. and Ule, J. (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol*, **17**, 909-915.
150. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M. *et al.* (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, **141**, 129-141.
151. Burgin, A.B. and Pace, N.R. (1990) Mapping the active site of ribonuclease P RNA using a substrate containing a photoaffinity agent. *Embo J*, **9**, 4111-4118.
152. Mironov, A.S., Gusarov, I., Rafikov, R., Lopez, L.E., Shatalin, K., Kreneva, R.A., Perumov, D.A. and Nudler, E. (2002) Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell*, **111**, 747-756.
153. Serganov, A. and Patel, D.J. (2009) Amino acid recognition and gene regulation by riboswitches. *Biochim Biophys Acta*.
154. Pikovskaya, O., Serganov, A.A., Polonskaia, A., Serganov, A. and Patel, D.J. (2009) Preparation and crystallization of riboswitch-ligand complexes. *Methods Mol Biol*, **540**, 115-128.
155. Serganov, A., Huang, L. and Patel, D.J. (2008) Structural insights into amino acid binding and gene control by a lysine riboswitch. *Nature*, **455**, 1263-1267.
156. Serganov, A., Yuan, Y.R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A.T., Hobartner, C., Micura, R., Breaker, R.R. and Patel, D.J. (2004) Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem Biol*, **11**, 1729-1741.
157. Huang, L., Serganov, A. and Patel, D.J. (2010) Structural insights into ligand recognition by a sensing domain of the cooperative glycine riboswitch. *Mol Cell*, **40**, 774-786.
158. Klein, D.J., Edwards, T.E. and Ferre-D'Amare, A.R. (2009) Cocrystal structure of a class I preQ1 riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. *Nat Struct Mol Biol*, **16**, 343-344.
159. Klein, D.J. and Ferre-D'Amare, A.R. (2006) Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science*, **313**, 1752-1756.
160. Edwards, T.E. and Ferre-D'Amare, A.R. (2006) Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition. *Structure*, **14**, 1459-1468.
161. Kulshina, N., Baird, N.J. and Ferre-D'Amare, A.R. (2009) Recognition of the bacterial second messenger cyclic diguanylate by its cognate riboswitch. *Nat Struct Mol Biol*, **16**, 1212-1217.
162. Blouin, S., Chinnappan, R. and Lafontaine, D.A. (2011) Folding of the lysine riboswitch: importance of peripheral elements for transcriptional regulation. *Nucleic Acids Res*, **39**, 3373-3387.
163. Blouin, S. and Lafontaine, D.A. (2007) A loop loop interaction and a K-turn motif located in the lysine aptamer domain are important for the riboswitch gene regulation control. *RNA*, **13**, 1256-1267.
164. Eschbach, S., St-Pierre, P., Penedo, J.C. and Lafontaine, D.A. (2012) Folding of the SAM-I riboswitch: A tale with a twist. *RNA Biol*, **9**.
165. Heppell, B., Blouin, S., Dussault, A.M., Mulhbacher, J., Ennifar, E., Penedo, J.C. and Lafontaine, D.A. (2011) Molecular insights into the ligand-controlled organization of the SAM-I riboswitch. *Nat Chem Biol*, **7**, 384-392.
166. Heppell, B. and Lafontaine, D.A. (2008) Folding of the SAM aptamer is determined by the formation of a K-turn-dependent pseudoknot. *Biochemistry*, **47**, 1490-1499.
167. Heppell, B., Mulhbacher, J., Penedo, J.C. and Lafontaine, D.A. (2009) Application of fluorescent measurements for characterization of riboswitch-ligand interactions. *Methods Mol Biol*, **540**, 25-37.
168. Lemay, J.F. and Lafontaine, D.A. (2007) Core requirements of the adenine riboswitch aptamer for ligand binding. *RNA*, **13**, 339-350.
169. Lemay, J.F., Penedo, J.C., Mulhbacher, J. and Lafontaine, D.A. (2009) Molecular basis of RNA-mediated gene regulation on the adenine riboswitch by single-molecule approaches. *Methods Mol Biol*, **540**, 65-76.
170. Lemay, J.F., Penedo, J.C., Tremblay, R., Lilley, D.M. and Lafontaine, D.A. (2006) Folding of the adenine riboswitch. *Chem Biol*, **13**, 857-868.
171. Mulhbacher, J. and Lafontaine, D.A. (2007) Ligand recognition determinants of guanine riboswitches. *Nucleic Acids Res*, **35**, 5568-5580.
172. Gusarov, I. and Nudler, E. (1999) The mechanism of intrinsic transcription termination. *Mol Cell*, **3**, 495-504.

173. Nudler, E. and Gusarov, I. (2003) Analysis of the intrinsic transcription termination mechanism and its control. *Methods Enzymol*, **371**, 369-382.
174. Yarnell, W.S. and Roberts, J.W. (1999) Mechanism of intrinsic transcription termination and antitermination. *Science*, **284**, 611-615.
175. Lim, J., Winkler, W.C., Nakamura, S., Scott, V. and Breaker, R.R. (2006) Molecular-recognition characteristics of SAM-binding riboswitches. *Angewandte Chemie (International ed*, **45**, 964-968.
176. Winkler, W.C., Nahvi, A., Sudarsan, N., Barrick, J.E. and Breaker, R.R. (2003) An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat Struct Biol*, **10**, 701-707.
177. McDaniel, B.A., Grundy, F.J. and Henkin, T.M. (2005) A tertiary structural element in S box leader RNAs is required for S-adenosylmethionine-directed transcription termination. *Mol Microbiol*, **57**, 1008-1021.
178. Winkler, W.C., Grundy, F.J., Murphy, B.A. and Henkin, T.M. (2001) The GA motif: an RNA element common to bacterial antitermination systems, rRNA, and eukaryotic RNAs. *RNA*, **7**, 1165-1172.
179. Montange, R.K. and Batey, R.T. (2006) Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature*, **441**, 1172-1175.
180. Sabanayagam, C.R., Eid, J.S. and Meller, A. (2005) Using fluorescence resonance energy transfer to measure distances along individual DNA molecules: corrections due to nonideal transfer. *J Chem Phys*, **122**, 061103.
181. Ho, C.K. and Shuman, S. (2002) Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains. *Proc Natl Acad Sci U S A*, **99**, 12709-12714.
182. Nandakumar, J., Ho, C.K., Lima, C.D. and Shuman, S. (2004) RNA substrate specificity and structure-guided mutational analysis of bacteriophage T4 RNA ligase 2. *J Biol Chem*, **279**, 31337-31347.
183. Nandakumar, J. and Shuman, S. (2005) Dual mechanisms whereby a broken RNA end assists the catalysis of its repair by T4 RNA ligase 2. *J Biol Chem*, **280**, 23484-23489.
184. Yin, S., Ho, C.K. and Shuman, S. (2003) Structure-function analysis of T4 RNA ligase 2. *J Biol Chem*, **278**, 17601-17608.
185. Lu, C., Ding, F., Chowdhury, A., Pradhan, V., Tomsic, J., Holmes, W.M., Henkin, T.M. and Ke, A. (2010) SAM recognition and conformational switching mechanism in the Bacillus subtilis yitJ S box/SAM-I riboswitch. *J Mol Biol*, **404**, 803-818.
186. Jao, C.Y. and Salic, A. (2008) Exploring RNA transcription and turnover in vivo by using click chemistry. *Proc Natl Acad Sci U S A*, **105**, 15779-15784.
187. Hammond, D.M., Manetto, A., Gierlich, J., Azov, V.A., Gramlich, P.M.E., Burley, G.A., Maul, M. and Carell, T. (2007) DNA photography: An ultrasensitive DNA-detection method based on photographic techniques. *Angewandte Chemie-International Edition*, **46**, 4184-4187.
188. Guan, L., van der Heijden, G.W., Bortvin, A. and Greenberg, M.M. (2011) Intracellular detection of cytosine incorporation in genomic DNA by using 5-ethynyl-2'-deoxycytidine. *Chembiochem*, **12**, 2184-2190.
189. Qu, D.Z., Wang, G.X., Wang, Z., Zhou, L., Chi, W.L., Cong, S.J., Ren, X.S., Liang, P.Z. and Zhang, B.L. (2011) 5-Ethynyl-2 '-deoxycytidine as a new agent for DNA labeling: Detection of proliferating cells. *Analytical Biochemistry*, **417**, 112-121.
190. Rao, H., Sawant, A.A., Tanpure, A.A. and Srivatsan, S.G. (2012) Posttranscriptional chemical functionalization of azide-modified oligoribonucleotides by bioorthogonal click and Staudinger reactions. *Chem Commun (Camb)*.
191. Weisbrod, S.H. and Marx, A. (2008) Novel strategies for the site-specific covalent labelling of nucleic acids. *Chem Commun (Camb)*, 5675-5685.
192. Wilson, D.S. and Szostak, J.W. (1999) In vitro selection of functional nucleic acids. *Annual Review of Biochemistry*, **68**, 611-647.
193. Kuwahara, M. and Sugimoto, N. (2010) Molecular evolution of functional nucleic acids with chemical modifications. *Molecules*, **15**, 5423-5444.
194. Su, W., Slepenkov, S., Grudzien-Nogalska, E., Kowalska, J., Kulis, M., Zuberek, J., Lukaszewicz, M., Darzynkiewicz, E., Jemielity, J. and Rhoads, R.E. (2011) Translation, stability, and resistance to decapping of mRNAs containing caps substituted in the triphosphate chain with BH3, Se, and NH. *RNA*, **17**, 978-988.
195. Huang, F., He, J., Zhang, Y. and Guo, Y. (2008) Synthesis of biotin-AMP conjugate for 5' biotin labeling of RNA through one-step in vitro transcription. *Nat Protoc*, **3**, 1848-1861.
196. Harrison, B. and Zimmerman, S.B. (1986) T4 polynucleotide kinase: macromolecular crowding increases the efficiency of reaction at DNA termini. *Anal Biochem*, **158**, 307-315.

197. Harrison, B. and Zimmerman, S.B. (1986) Stabilization of T4 polynucleotide kinase by macromolecular crowding. *Nucleic Acids Res*, **14**, 1863-1870.
198. Jorgensen, E.D., Durbin, R.K., Risman, S.S. and McAllister, W.T. (1991) Specific contacts between the bacteriophage T3, T7, and SP6 RNA polymerases and their promoters. *J Biol Chem*, **266**, 645-651.
199. Lee, S.S. and Kang, C. (1993) Two base pairs at -9 and -8 distinguish between the bacteriophage T7 and SP6 promoters. *J Biol Chem*, **268**, 19299-19304.
200. Lee, S.S., Park, S.K., Cho, I.H. and Kang, C. (1993) All 4 bases of both strands at -9 and -8 in T7 promoter are needed to be substituted by SP6-specific bases to switch promoter specificity. *Biochem Mol Biol Int*, **31**, 1017-1021.
201. Joho, K.E., Gross, L.B., McGraw, N.J., Raskin, C. and McAllister, W.T. (1990) Identification of a region of the bacteriophage T3 and T7 RNA polymerases that determines promoter specificity. *J Mol Biol*, **215**, 31-39.
202. Kuzmine, I., Gottlieb, P.A. and Martin, C.T. (2003) Binding of the priming nucleotide in the initiation of transcription by T7 RNA polymerase. *J Biol Chem*, **278**, 2819-2823.
203. Krupp, G. (1988) RNA synthesis: strategies for the use of bacteriophage RNA polymerases. *Gene*, **72**, 75-89.
204. Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res*, **15**, 8783-8798.
205. Krieg, P.A. and Melton, D.A. (1984) Functional messenger RNAs are produced by SP6 in vitro transcription of cloned cDNAs. *Nucleic Acids Res*, **12**, 7057-7070.
206. Martin, C.T., Muller, D.K. and Coleman, J.E. (1988) Processivity in early stages of transcription by T7 RNA polymerase. *Biochemistry*, **27**, 3966-3974.
207. Kohli, V. and Temsamani, J. (1993) Comparison of in vitro transcriptions using various types of DNA templates. *Anal Biochem*, **208**, 223-227.
208. Butler, E.T. and Chamberlin, M.J. (1982) Bacteriophage SP6-specific RNA polymerase. I. Isolation and characterization of the enzyme. *J Biol Chem*, **257**, 5772-5778.
209. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res*, **12**, 7035-7056.
210. Schenborn, E.T. and Mierendorf, R.C., Jr. (1985) A novel transcription property of SP6 and T7 RNA polymerases: dependence on template structure. *Nucleic Acids Res*, **13**, 6223-6236.
211. Wolf, J., Dombos, V., Appel, B. and Muller, S. (2008) Synthesis of guanosine 5'-conjugates and their use as initiator molecules for transcription priming. *Org Biomol Chem*, **6**, 899-907.
212. Fusz, S., Eisenfuhr, A., Srivatsan, S.G., Heckel, A. and Famulok, M. (2005) A ribozyme for the aldol reaction. *Chem Biol*, **12**, 941-950.
213. Fusz, S., Srivatsan, S.G., Ackermann, D. and Famulok, M. (2008) Photocleavable initiator nucleotide substrates for an aldolase ribozyme. *J Org Chem*, **73**, 5069-5077.
214. Eisenfuhr, A., Arora, P.S., Sengle, G., Takaoka, L.R., Nowick, J.S. and Famulok, M. (2003) A ribozyme with michaelase activity: synthesis of the substrate precursors. *Bioorg Med Chem*, **11**, 235-249.
215. Sengle, G., Eisenfuhr, A., Arora, P.S., Nowick, J.S. and Famulok, M. (2001) Novel RNA catalysts for the Michael reaction. *Chem Biol*, **8**, 459-473.
216. Seelig, B. and Jäschke, A. (1999) Ternary conjugates of guanosine monophosphate as initiator nucleotides for the enzymatic synthesis of 5'-modified RNAs. *Bioconjug Chem*, **10**, 371-378.
217. Fiammengo, R., Musilek, K. and Jäschke, A. (2005) Efficient preparation of organic substrate-RNA conjugates via in vitro transcription. *J Am Chem Soc*, **127**, 9271-9276.
218. Tarasow, T.M., Tarasow, S.L. and Eaton, B.E. (1997) RNA-catalysed carbon-carbon bond formation. *Nature*, **389**, 54-57.
219. Huang, F. (2003) Efficient incorporation of CoA, NAD and FAD into RNA by in vitro transcription. *Nucleic Acids Res*, **31**, e8.
220. Huang, F., Wang, G., Coleman, T. and Li, N. (2003) Synthesis of adenosine derivatives as transcription initiators and preparation of 5' fluorescein- and biotin-labeled RNA through one-step in vitro transcription. *RNA*, **9**, 1562-1570.
221. Li, N., Yu, C. and Huang, F. (2005) Novel cyanine-AMP conjugates for efficient 5' RNA fluorescent labeling by one-step transcription and replacement of [gamma-32P]ATP in RNA structural investigation. *Nucleic Acids Res*, **33**, e37.
222. Schlatterer, J.C. and Jäschke, A. (2006) Universal initiator nucleotides for the enzymatic synthesis of 5'-amino- and 5'-thiol-modified RNA. *Biochem Biophys Res Commun*, **344**, 887-892.

223. Zhang, L., Sun, L.L., Cui, Z.Y., Gottlieb, R.L. and Zhang, B.L. (2001) 5 '-sulfhydryl-modified RNA: Initiator synthesis, in vitro transcription, and enzymatic incorporation. *Bioconjugate Chem.*, **12**, 939-948.
224. Huang, F. and Shi, Y. (2010) Synthesis of symmetrical thiol-adenosine conjugate and 5' thiol-RNA preparation by efficient one-step transcription. *Bioorg Med Chem Lett*, **20**, 6254-6257.
225. Ryu, Y., Kim, K.J., Roessner, C.A. and Scott, A.I. (2006) Decarboxylative Claisen condensation catalyzed by in vitro selected ribozymes. *Chem Commun (Camb)*, 1439-1441.
226. Tsukiji, S., Pattnaik, S.B. and Suga, H. (2003) An alcohol dehydrogenase ribozyme. *Nat Struct Biol*, **10**, 713-717.
227. Paredes, E. and Das, S.R. (2011) Click chemistry for rapid labeling and ligation of RNA. *Chembiochem*, **12**, 125-131.
228. Pfander, S., Fiammengo, R., Kirin, S.I., Metzler-Nolte, N. and Jäschke, A. (2007) Reversible site-specific tagging of enzymatically synthesized RNAs using aldehyde-hydrazine chemistry and protease-cleavable linkers. *Nucleic Acids Res*, **35**, e25.
229. Sengle, G., Jenne, A., Arora, P.S., Seelig, B., Nowick, J.S., Jäschke, A. and Famulok, M. (2000) Synthesis, incorporation efficiency, and stability of disulfide bridged functional groups at RNA 5'-ends. *Bioorg Med Chem*, **8**, 1317-1329.
230. Salic, A. and Mitchison, T.J. (2008) A chemical method for fast and sensitive detection of DNA synthesis in vivo. *Proc Natl Acad Sci U S A*, **105**, 2415-2420.
231. Gramlich, P.M.E., Warncke, S., Gierlich, J. and Carell, T. (2008) Click-click-click: Single to triple modification of DNA. *Angewandte Chemie-International Edition*, **47**, 3442-3444.
232. Wirges, C.T., Gramlich, P.M.E., Gutsmiedl, K., Gierlich, J., Burley, G.A. and Carell, T. (2007) Pronounced effect of DNA hybridization on click reaction efficiency. *Qsar & Combinatorial Science*, **26**, 1159-1164.
233. Kellner, S., Seidu-Larry, S., Burhenne, J., Motorin, Y. and Helm, M. (2011) A multifunctional bioconjugate module for versatile photoaffinity labeling and click chemistry of RNA. *Nucleic Acids Res*, **39**, 7348-7360.
234. Motorin, Y., Burhenne, J., Teimer, R., Koynov, K., Willnow, S., Weinhold, E. and Helm, M. (2011) Expanding the chemical scope of RNA:methyltransferases to site-specific alkynylation of RNA for click labeling. *Nucleic Acids Res*, **39**, 1943-1952.
235. Aigner, M., Hartl, M., Fauster, K., Steger, J., Bister, K. and Micura, R. (2011) Chemical synthesis of site-specifically 2'-azido-modified RNA and potential applications for bioconjugation and RNA interference. *Chembiochem*, **12**, 47-51.
236. Pitulle, C., Kleineidam, R.G., Sproat, B. and Krupp, G. (1992) Initiator oligonucleotides for the combination of chemical and enzymatic RNA synthesis. *Gene*, **112**, 101-105.
237. Pokrovskaya, I.D. and Gurevich, V.V. (1994) In vitro transcription: preparative RNA yields in analytical scale reactions. *Anal Biochem*, **220**, 420-423.
238. Schenborn, E.T. (1998) Transcription in vitro using bacteriophage RNA polymerases. *Methods Mol Biol*, **86**, 209-219.
239. Stengel, G., Urban, M., Purse, B.W. and Kuchta, R.D. (2010) Incorporation of the fluorescent ribonucleotide analogue tCTP by T7 RNA polymerase. *Anal Chem*, **82**, 1082-1089.
240. Srivatsan, S.G. and Tor, Y. (2007) Fluorescent pyrimidine ribonucleotide: synthesis, enzymatic incorporation, and utilization. *J Am Chem Soc*, **129**, 2044-2053.
241. Srivatsan, S.G. and Tor, Y. (2007) Synthesis and enzymatic incorporation of a fluorescent pyrimidine ribonucleotide. *Nat Protoc*, **2**, 1547-1555.
242. Pawar, M.G. and Srivatsan, S.G. (2011) Synthesis, photophysical characterization, and enzymatic incorporation of a microenvironment-sensitive fluorescent uridine analog. *Org Lett*, **13**, 1114-1117.
243. Kao, C., Rudisser, S. and Zheng, M. (2001) A simple and efficient method to transcribe RNAs with reduced 3' heterogeneity. *Methods*, **23**, 201-205.
244. Kao, C., Zheng, M. and Rudisser, S. (1999) A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA*, **5**, 1268-1272.
245. Fontanel, M.L., Bazin, H. and Teoule, R. (1994) Sterical recognition by T4 polynucleotide kinase of non-nucleosidic moieties 5'-attached to oligonucleotides. *Nucleic Acids Res*, **22**, 2022-2027.
246. Fontanel, M.L., Bazin, H. and Teoule, R. (1993) 32P labeling of nonnucleosidic moieties 5'-attached to oligonucleotides. *Anal Biochem*, **214**, 338-340.
247. van Houten, V., Denkers, F., van Dijk, M., van den Brekel, M. and Brakenhoff, R. (1998) Labeling efficiency of oligonucleotides by T4 polynucleotide kinase depends on 5'-nucleotide. *Anal Biochem*, **265**, 386-389.

248. Masento, M.S., Hewer, A., Grover, P.L. and Phillips, D.H. (1989) Enzyme-mediated phosphorylation of polycyclic hydrocarbon metabolites: detection of non-adduct compounds in the 32P-postlabelling assay. *Carcinogenesis*, **10**, 1557-1559.
249. Brennan, C.A. and Gumport, R.I. (1985) T4 RNA ligase catalyzed synthesis of base analogue-containing oligodeoxyribonucleotides and a characterization of their thermal stabilities. *Nucleic Acids Res*, **13**, 8665-8684.
250. Bain, J.D. and Switzer, C. (1992) Regioselective ligation of oligoribonucleotides using DNA splints. *Nucleic Acids Res*, **20**, 4372.
251. Barrio, J.R., Barrio, M.C., Leonard, N.J., England, T.E. and Uhlenbeck, O.C. (1978) Synthesis of modified nucleoside 3',5'-bisphosphates and their incorporation into oligoribonucleotides with T4 RNA ligase. *Biochemistry*, **17**, 2077-2081.
252. Wower, J., Rosen, K.V., Hixson, S.S. and Zimmermann, R.A. (1994) Recombinant photoreactive tRNA molecules as probes for cross-linking studies. *Biochimie*, **76**, 1235-1246.
253. Wower, J., Hixson, S.S. and Zimmermann, R.A. (1988) Photochemical cross-linking of yeast tRNA(Phe) containing 8-azidoadenosine at positions 73 and 76 to the Escherichia coli ribosome. *Biochemistry*, **27**, 8114-8121.
254. Whitfeld, P.R. and Markham, R. (1953) Natural configuration of the purine nucleotides in ribonucleic acids; chemical hydrolysis of the dinucleoside phosphates. *Nature*, **171**, 1151-1152.
255. Paulsen, H. and Wintermeyer, W. (1984) Incorporation of 1,N6-ethenoadenosine into the 3' terminus of tRNA using T4 RNA ligase. 1. Preparation of yeast tRNAPhe derivatives. *Eur J Biochem*, **138**, 117-123.
256. Paulsen, H. and Wintermeyer, W. (1984) Incorporation of 1,N6-ethenoadenosine into the 3' terminus of tRNA using T4 RNA ligase. 2. Preparation and ribosome interaction of fluorescent Escherichia coli tRNAMetf. *Eur J Biochem*, **138**, 125-130.
257. Hansske, F. and Cramer, F. (1979) Modification of the 3' terminus of tRNA by periodate oxidation and subsequent reaction with hydrazides. *Methods Enzymol*, **59**, 172-181.
258. Rosemeyer, V., Laubrock, A. and Seibl, R. (1995) Nonradioactive 3'-end-labeling of RNA molecules of different lengths by terminal deoxynucleotidyltransferase. *Anal Biochem*, **224**, 446-449.
259. Sippel, A.E. (1973) Purification and characterization of adenosine triphosphate: ribonucleic acid adenyltransferase from Escherichia coli. *Eur J Biochem*, **37**, 31-40.
260. Hafner, M., Renwick, N., Brown, M., Mihailovic, A., Holoch, D., Lin, C., Pena, J.T., Nusbaum, J.D., Morozov, P., Ludwig, J. *et al.* (2011) RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA*, **17**, 1697-1712.
261. Khakshoor, O. and Kool, E.T. (2011) Chemistry of nucleic acids: impacts in multiple fields. *Chem Commun (Camb)*, **47**, 7018-7024.
262. Juskowiak, B. (2011) Nucleic acid-based fluorescent probes and their analytical potential. *Anal Bioanal Chem*, **399**, 3157-3176.
263. Itzkovitz, S. and van Oudenaarden, A. (2011) Validating transcripts with probes and imaging technology. *Nat Methods*, **8**, S12-19.
264. Bao, G., Rhee, W.J. and Tsourkas, A. (2009) Fluorescent probes for live-cell RNA detection. *Annu Rev Biomed Eng*, **11**, 25-47.
265. Myong, S., Bruno, M.M., Pyle, A.M. and Ha, T. (2007) Spring-loaded mechanism of DNA unwinding by hepatitis C virus NS3 helicase. *Science*, **317**, 513-516.
266. Greulich, K.O. (2005) Single-molecule studies on DNA and RNA. *Chemphyschem*, **6**, 2458-2471.
267. Winz, M.L., Samanta, A., Benzinger, D. and Jäschke, A. (2012) Site-specific terminal and internal labeling of RNA by poly(A) polymerase tailing and copper-catalyzed or copper-free strain-promoted click chemistry. *Nucleic Acids Res*.
268. Debets, M.F., van Berkel, S.S., Dommerholt, J., Dirks, A.T., Rutjes, F.P. and van Delft, F.L. (2011) Bioconjugation with strained alkenes and alkynes. *Acc Chem Res*, **44**, 805-815.
269. Komura, J. and Riggs, A.D. (1998) Terminal transferase-dependent PCR: a versatile and sensitive method for in vivo footprinting and detection of DNA adducts. *Nucleic Acids Res*, **26**, 1807-1811.
270. Chen, H.H., Castanotto, D., LeBon, J.M., Rossi, J.J. and Riggs, A.D. (2000) In vivo, high-resolution analysis of yeast and mammalian RNA-protein interactions, RNA structure, RNA splicing and ribozyme cleavage by use of terminal transferase-dependent PCR. *Nucleic Acids Res*, **28**, 1656-1664.
271. Schmidt, W.M. and Mueller, M.W. (1996) Controlled ribonucleotide tailing of cDNA ends (CRTC) by terminal deoxynucleotidyl transferase: a new approach in PCR-mediated analysis of mRNA sequences. *Nucleic Acids Res*, **24**, 1789-1791.

272. Wombacher, R. and Jäschke, A. (2008) Probing the active site of a diels-alderase ribozyme by photoaffinity cross-linking. *J Am Chem Soc*, **130**, 8594-8595.
273. Kwon, M. and Strobel, S.A. (2008) Chemical basis of glycine riboswitch cooperativity. *RNA*, **14**, 25-34.

Abbreviations

AP-quenching	Aminopurine fluorescence quenching
BP	Band-pass
Boc_2O	Di-*tert*-butyl dicarbonate
CuAAC	Copper-catalyzed Azide-Alkyne Cycloaddition
DAinv	Inverse electron demand Diels-Alder cycloaddition
DCM	Dichloromethane
DIEA	Diisopropylethylamine
DMAP	4-(Dimethylamino)pyridine
EDC.HCl	*N*-(3-Dimethylaminopropyl)-*N*′-ethylcarbodiimide hydrochloride
KL buffer	Kinase Ligation buffer
FA-buffer	Formamide gel-loading buffer
FRET	Fluorescence Resonance Energy Transfer
LCC	Ligation competent complex
LP	Long-pass
NHS	N-Hydroxysuccinimide
PAGE	Polyacrylamide gel electrophoresis
PAP	Poly(A) polymerase
PNK	Polynucleotide kinase
PUP	Poly(U) polymerase
PTPP	Pyrithiamine pyrophosphate
RNAP	RNA polymerase
SAH	S-adenosyl-L-homocysteine
SAM	S-adenosyl-L-methionine
SELEX	Systematic evolution of ligands by exponential enrichment
SM-FRET	Single Molecule Fluorescence Resonance Energy Transfer
SP	Short-pass
SPAAC	Strain Promoted Azide Alkyne Cycloaddition
Strep-EMSA	Streptavidin Electrophoretic Mobility Shift Assay
T4 Dnl	T4 DNA ligase
T4 Rnl1	T4 RNA ligase 1
T4 Rnl2	T4 RNA ligase 2

TBTA	Tris-(benzyltriazolylmethyl) amine
TdT	Terminal deoxynucleotidyl transferase
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
THPTA	Tris-(3-hydroxypropyltriazolylmethyl) amine
TLC	Thin Layer Chromatography
TPP	Thiamine pyrophosphate