Optimierung und Analyse von synthetischen Tetrazyklin-Tandem-Riboswitchen durch *machine learning*

vom Fachbereich Biologie der Technischen Universität Darmstadt

zur Erlangung des Grades

Doctor rerum naturalium

(Dr. rer. nat.)

Dissertation

von Ann-Christin Groher

Erstgutachterin: Prof. Dr. Beatrix Süß

Zweitgutachter Prof. Dr. Johannes Kabisch

Darmstadt 2021

Groher, Ann-Christin: Optimierung und Analyse von synthetischen Tetrazyklin-Tandem-Riboswitchen durch *machine learning*

Darmstadt, Technische Universität Darmstadt,
Jahr der Veröffentlichung der Dissertation auf TUprints: 2021
URN: urn:nbn:de:tuda-tuprints-178913
Tag der mündlichen Prüfung: 31.03.2021

Für Flo, Noah, Charlotte
und meine geliebte
Oma

Danksagung

Ganz besonders möchte ich **Prof. Dr. Beatrix Süß** für die Möglichkeit danken, meine Doktorarbeit in ihrem Labor durchführen zu können und mich zudem sehr herzlich für hervorragende Betreuung, Unterstützung und den stetigen wissenschaftlichen Austausch bedanken.

Sehr herzlich möchte ich mich auch bei **Prof. Dr. Johannes Kabisch** für die Übernahme des Zweitgutachtens bedanken.

Ich bedanke mich auch bei meinen Prüfern **Prof. Dr. Heinz Köppl** und **Prof. Dr. Torsten Waldminghaus** für die Bereitschaft meiner Dissertationsprüfung beizuwohnen.

Dr. Sven Jager danke ich ebenfalls sehr herzlich für die tolle und produktive Zusammenarbeit während unseres gemeinsamen Projektes.

Ich danke allen Beteiligten des interdisziplinären **Projekt CompuGene** für den interessanten Austausch und die gute Zusammenarbeit.

Auch möchte ich mich sehr bei allen aktuellen und ehemaligen Mitarbeitern des Süßlabs bedanken. **Dr. Alexander Wittmann** danke ich, für die Betreuung meines ersten Praktikums in der Arbeitsgruppe. **Dr. Marc Vogel** und **Dr. Michael Vockenhuber** und **Dunja Sehn** danke ich, dass ihr für meine Fragen stets ein offenes Ohr hattet. Ich danke insbesondere **Leon, Franzi, Anne, Annette, Stephen, Stella, Jeannine und Julia** dafür, dass ihr einfach super nette Kollegen wart und für die zum Teil urkomischen Mittagspausen.

Ich danke auch meinen **Eltern** und meinen **Schwestern**. Danke für eure Unterstützung und dafür, dass ihr immer für mich da wart.

Mein größter Dank gilt dir **Flo**, meinem **Ehemann, Freund** und **Postdoc**, für die Unterstützung und Liebe in den letzten 12 Jahren. Danke, dass du immer für mich da warst, mich immer unterstützt hast und immer an mich geglaubt hast.

Teile dieser Arbeit wurden bereits publiziert:

Groher, A.-C., Jager, S., Schneider, C., Groher, F., Hamacher, K., & Suess, B. (2018a). Tuning the Performance of Synthetic Riboswitches using Machine Learning. *ACS Synthetic Biology*, *8*(1), 34–44. http://doi.org/10.1021/acssynbio.8b00207

1 Zusammenfassung

Die RNA erfüllt in der Zelle eine Vielzahl an verschiedenen Aufgaben, die zum Teil eng mit der Genregulation verknüpft sind. Bekannte Beispiele sind unter anderem Riboswitche, die sowohl natürlich vorkommen als auch synthetisch hergestellt werden können. Riboswitche können im 5´untranslatierten Bereich (UTR) von Genen die Translation beeinflussen und so als effektive Werkzeuge zur Kontrolle der Genexpression wirken. Jedoch ist die Effizienz synthetisch hergestellter Riboswitche zur Genregulation meist begrenzt und bedarf eines langen Optimierungsprozesses. Eine Perspektive bieten automatisierte Verfahren, welche auch in der Synthetischen Biologie einen immer höheren Stellenwert bekommen.

Diese Arbeit beschäftigt sich mit der Entwicklung eines *machine learning*-Programms zur Optimierung synthetischer Riboswitche, welche in den Zellen der Bäckerhefe ihre Anwendung finden, sowie der Analyse der generierten Daten im Hinblick auf ihre biophysikalischen Parameter und Sequenzmotive. Das Tetrazyklin (TC)-Dimer LG3, welches aus zwei TC-Aptameren besteht, die sich nur in ihrem Endstamm P1 unterscheiden, diente hier als Vorlage des Optimierungsprozesses. Durch die Veränderung der Sequenz des P1-Stammes, lässt sich die Basalexpression und der Schaltfaktor dieses Riboswitches beeinflussen. Das *machine learning*-Programm wurde mit Daten trainiert, die sich aus der Sequenz berechnen lassen: Länge des P1-Stamms, dessen GC-Gehalt, die minimale freie Energie (MFE oder ΔG), die Entropie (Shannon) von P1 sowie der Wasserstoffbrückenbindung der beteiligte Basenpaare (H-Bindung), die Schmelztemperatur von P1 (Tm P1) sowie die Schmelztemperatur des kompletten Aptamers (Tm). Es wurden insgesamt drei *machine learning*-Runden durchgeführt, wobei in der 3. Runde erstmals eine signifikante Verbesserung des mittleren Schaltfaktors der Riboswitche beobachtet werden konnte. Nach der 3. Runde wurde das Programm um ein *deep learning*-Programm erweitert und so zusätzlich ein Trainieren auf Sequenzdaten des Stammes ermöglicht. Mit der Kombination der beiden Programme wurde ein außergewöhnlich guter Riboswitch gefunden (R4-G8), der einen Schaltfaktor von 40-fach und die Stammsequenz 5´AGGTGACC3´ aufweist. Nachfolgende Analysen der Daten ergaben, dass ein bestimmter Bereich biophysikalischer Parameter und bestimmte Sequenzmotive innerhalb des P1-Stamms das Vorkommen gut schaltender Riboswitche begünstigt und sich R4-G8 mit seinen biophysikalischen Parametern und seiner Sequenz sehr wahrscheinlich an seinem individuellen Optimum befindet, da jede weitere Veränderung der Sequenz zu einer Verschlechterung des Schaltfaktors führte.

Im letzten Abschnitt dieser Arbeit wurden drei verschiedene TC-Aptamere mit P1-Stämmen aus den vorrangegangenen *machine learning*-Runden mit einem Tobramycin-Aptamer zu einem funktionalen NOR-Gate fusioniert. Dabei wurde ein TC-Tobramycin-Hybrid erzeugt, bei welchem das Tobramycin-Aptamer auf den P2-Stamm des TC-Aptamers gesetzt wurde. Beide Aptamere können sowohl unabhängig voneinander als auch gemeinsam ihren Liganden binden und die Translation inhibieren.

1.1 English summary

RNA performs a variety of different tasks in the cell, many of them closely linked to gene regulation. Well-known examples include riboswitches, which are both naturally occurring and synthetically produced. Riboswitches can affect translation in the 5'untranslated region (UTR) of genes and thus act as effective tools to control gene expression. However, the efficiency of synthetically produced riboswitches for gene regulation is usually limited and requires a long optimization process. Automated methods, which are also of increasing importance in synthetic biology, offer one perspective.

This work focuses on the development of a machine learning program for the optimization of synthetic riboswitches, which find their use in the cells of baker's yeast, and the analysis of the generated data with respect to their biophysical parameters and sequence motifs. The tetracycline (TC) dimer LG3, which consists of two TC aptamers differing only in their terminal stem P1, was used here as a template of the optimization process. By changing the sequence of the P1 stem, the basal expression and switching factor of this riboswitch can be influenced. The machine learning program was trained with data that can be calculated from the sequence: Length of the P1 stem, its GC content, the minimum free energy (MFE or ΔG), the entropy (Shannon) of P1 as well as the hydrogen bonding of the involved base pairs (H-bonding), the melting temperature of P1 (Tm P1) as well as the melting temperature of the complete aptamer (Tm). Three machine learning rounds were conducted in total, with a significant improvement in the mean switching factor of the riboswitches being observed for the first time in the 3rd round. After the 3rd round, a deep learning program was added to the program, allowing additional training on sequence data of the stem. With the combination of the two programs, an extremely good riboswitch was found (R4-G8), with a switching factor of 40-fold and the stem sequence 5'AGGTGACC3'. Subsequent analyses of the data revealed that a specific range of biophysical parameters and certain sequence motifs within the P1 stem favored the occurrence of good switching riboswitches, and that R4-G8's biophysical parameters and sequence were most likely at its individual optimum, since any further change in sequence resulted in a decrease of switching factor.

In the last section of this work, three different TC aptamers with P1 stems from the previous machine learning rounds were fused with a tobramycin aptamer to form a functional NOR gate. This generated a TC-tobramycin hybrid in which the tobramycin aptamer was placed on the P2 stem of the TC aptamer. Both aptamers can bind their ligand and inhibit translation both independently and together.

2 Einleitung

2.1 Synthetische Biologie

Die Biologie beschäftigt sich auf wissenschaftlicher Ebene mit allen Bereichen des Lebens. So werden zum Beispiel Entwicklungen, Verhalten, Physiologie und Biochemie von Lebewesen untersucht und erforscht. In der Synthetische Biologie, als dem neuesten Teilgebiet der Biologie, kommen vor allem ingenieurswissenschaftliche Ansätze auf biologischen Grundlagen zum Tragen. Ein Ziel der Synthetischen Biologie ist es, biologische Systeme zu designen, welche nicht in der Natur existieren (EC et al. 2005). Auch werden ingenieurswissenschaftliche Prinzipien und ein *re-design* existierender Prinzipen angewandt, um biologische Prozesse besser zu verstehen. Dabei sind vor allem Prinzipien wie Standardisierung, Modularität, digitale Logik und mathematisch vorhersehbares Verhalten von Bedeutung (Bartley et al. 2017).

In den letzten Jahren wurde der Synthetischen Biologie ein immer größerer Stellenwert zugemessen. Im Unterschied zur „klassischen Biologie“ ermöglicht die Synthetische Biologie die aktive Veränderung von Vorgängen, Prozessen und ganzen Organismen hin zu etwas gänzlich Neuem. Einzelne Moleküle können verändert und designt werden, die wiederum Stoffwechselwege und Organismen verändern und beeinflussen. So hat die Synthetische Biologie in vielen Teilbereichen der Biologie Einzug gehalten und ist aus der Grundlagenforschung wie auch aus der Diagnostik nicht mehr weg zu denken (Slomovic et al. 2015). Obgleich der Begriff „Synthetische Biologie“ erstmals 1911 vom französischen Biophysiker Stéphan Leduc benutzt wurde, liegen dessen Ursprünge in des zweiten Hälfte des letzten Jahrhunderts. Zwar kann die Synthetischen Biologie auf jeder biologischen Ebene, von der DNA, RNA über Proteine bis hin zu ganzen Organismen operieren, jedoch wurde die Untersuchung und Entwicklung von molekularen Regelkreisen, deren Grundlage auf Francois Jacobs und Jaques Monods wegweisende Erkenntnissen über das Lac-Operon in *Escherichia coli* basiert, zu einem ihrer bedeutendsten Meilensteine. Es wurden Methoden entwickelt, mit deren Hilfe es möglich war, neue regulatorische Systeme aus molekularen Komponenten zusammenzubauen (Cameron et al. 2014). Der Durchbruch wurde schließlich Mitte der 1990er Jahre mit der Entwicklung der automatisierten DNA-Sequenzierung, der Verbesserung von *high-throughput*-Techniken zur Messung von RNA, Proteinen, Lipiden und Metaboliten und optimierten Klonierungsmethoden erzielt (Ideker et al. 2001; Jeong et al. 2000; Westerhoff & Palsson 2004). Hier wurde die Grundlage zur Berechnung zellulärer Netzwerke sowie deren Nachbau gelegt.

E. coli entwickelte sich schnell zu dem Organismus, in welchem die ersten, künstlich hergestellten genetischen Schaltkreise untersucht werden konnten, da er sehr gut erforscht ist, seine biochemischen Stoffwechselwege sehr gut entschlüsselt sind und er mit einfachen Mitteln relativ

schnell genetisch manipuliert werden kann. Weiter angelehnt an die Ingenieurswissenschaften und mit dem Ziel eines mathematischen Verständnisses biologischer Vorgänge, wurde Anfang 2000 die Entwicklung erster genetischen Schaltkreise publiziert, welche analog zu elektrischen Schaltkreisen funktionierten (McAdams & Shapiro 1995; McAdams & Arkin 2000; Cameron et al. 2014). Die RNA, welche sich mit einfachen Methoden gut manipulieren lässt, entwickelte sich in den darauffolgenden Jahren zu einem wichtigen Werkzeug im Schaltkreisdesign. So konnte der Entwurf von synthetischen Schaltkreisen von der hauptsächlich transkriptioneller Kontrolle auf posttranskriptionelle und translatorische Kontrollmechanismen ausgeweitet werden (Cameron et al. 2014). Eine Möglichkeit, die Genexpression mittels RNA zu kontrollieren und damit genetische Schaltkreise aufzubauen, bieten sogenannte Riboswitche (F. Groher & Suess 2014). Riboswitche sind RNA-Strukturen, mit welchen man die Genexpression auf mRNA-Ebene kontrollieren kann. Der Einsatz von RNA-basierten Regulatoren entwickelt sich derzeit zu einem viel erforschten Bereich innerhalb der Synthetischen Biologie und sowohl regulatorische, nichtkodierende kleine RNAs (*Toehold-Switche* und *Stars*) als auch Aptamer-basierte Riboswitche bieten eine Vielzahl an Möglichkeiten für die Generierung von logischen Gattern und Schaltkreisen in der Zelle.

2.2 Grundlagen der Molekularbiologie und das zentrale Dogma der Biologie

Die Regulation der Genexpression auf mRNA-Ebene ist nicht nur ein Forschungsgebiet der Synthetischen Biologie, sondern auch der Molekularbiologie. Wie sich aus dem Namen ableiten lässt, beschäftigt sich die Molekularbiologie mit Stoffen und Prozessen, die auf molekularer Ebene innerhalb biologischer Systeme passieren. Kohlenstoff, Wasserstoff und Sauerstoff spielen als Grundgerüst biologischer Moleküle eine zentrale Rolle. Kohlenstoffe, Proteine, Lipide und Nukleinsäuren, die zu den wichtigsten biologischen Stoffen gehören, verfügen in ihren Grundbausteinen, den Monosacchariden, Aminosäuren und Carbonsäuren, über mindestens zwei Kohlenstoffatome sowie mehrere Wasserstoff- und Sauerstoffatomen. Ebenso unverzichtbar für biologische Moleküle sind die Atome Stickstoff, Schwefel und Phosphor.

Das zentrale Dogma der Molekularbiologie wurde erstmals von Francis Crick beschrieben (Crick 1970) und beschreibt die Übertragung von Informationen, ausgehend von der DNA über die RNA hin zu Proteinen. Die DNA, welche aus den Nukleobasen Alanin, Thymin, Cytosin und Guanin sowie einer Desoxyribose und einem Phosphatrest besteht, enthält als Abfolge dieser Basen den genetischen Code einer jeden Zelle. Dieser wird mit Hilfe einer Polymerase in die RNA übertragen (transkribiert), welche schließlich mit Hilfe des Ribosoms in Proteine übersetzt (translatiert) wird. Sie sind nicht nur wichtig als Bausteine selbst, sondern katalysieren als Enzyme nahezu alle Prozesse innerhalb einer

Zelle. Die RNA nimmt in diesem Dogma die Rolle als übersetzender Teil zwischen der DNA und den Proteinen ein (Abbildung 2.2). Mittlerweile weiß man jedoch, dass die RNA über ihre Rolle als Übersetzter hinaus weitaus mehr Aufgaben in der Zelle erfüllt.

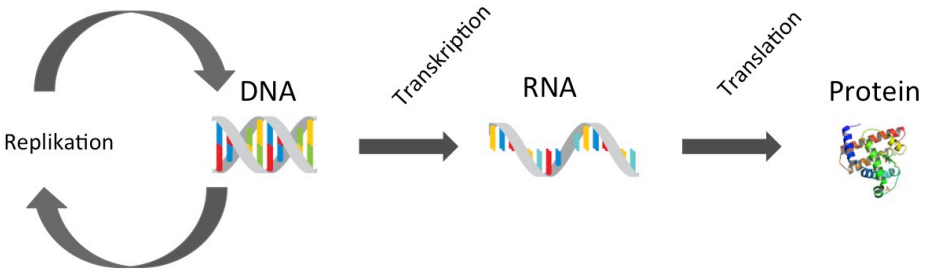


Abbildung 2.2 Das Dogma der Molekularbiologie. Die DNA kann über Replikation dupliziert werden. Von der DNA kann über Transkription die RNA synthetisiert werden und von dieser über Translation die Proteine. Die Genexpression kann unter anderem auf den Ebenen der Transkription und Translation gesteuert werden.

2.3 Die vielseitige Rolle der RNA in der Zelle

Während die DNA stets als Doppelhelix vorliegt, über zwei Desoxyribose-Phosphat-Rückgrate verfügt und sehr stabil ist, verfügt die RNA über ein Ribose-Phosphat-Rückgrat und liegt daher in der Zelle instabiler und flexibler vor. Die einzigen Unterschiede zur DNA sind chemisch gesehen somit nur das Vorhandensein der Base Uracil (anstatt Thymin), was die Bildung von Wobble-Basenpaaren ermöglicht sowie das Vorhandensein einer OH-Gruppe am C2-Atom der Ribose (Abbildung 2.3). In der DNA wird diese OH-Gruppe durch ein einfaches Wasserstoffatom ausgetauscht, was hier zusätzlich zur Stabilisierung beiträgt. Obwohl die RNA zunächst einzelsträngig vorliegt, kann sie, ähnlich wie Proteine, über interne Basenpaarungen komplexe Sekundär- und sogar Tertiärstrukturen ausbilden. Die Enden von DNA und RNA verfügen in der Regel über ein Phosphat, welches am 5´ Ende (am 5ten C-Atom der Ribose) angehängt ist (5´Phosphat) und eine OH-Gruppe, welche am 3´Ende (am 3ten C-Atom der Ribose) angehängt ist (3´OH).

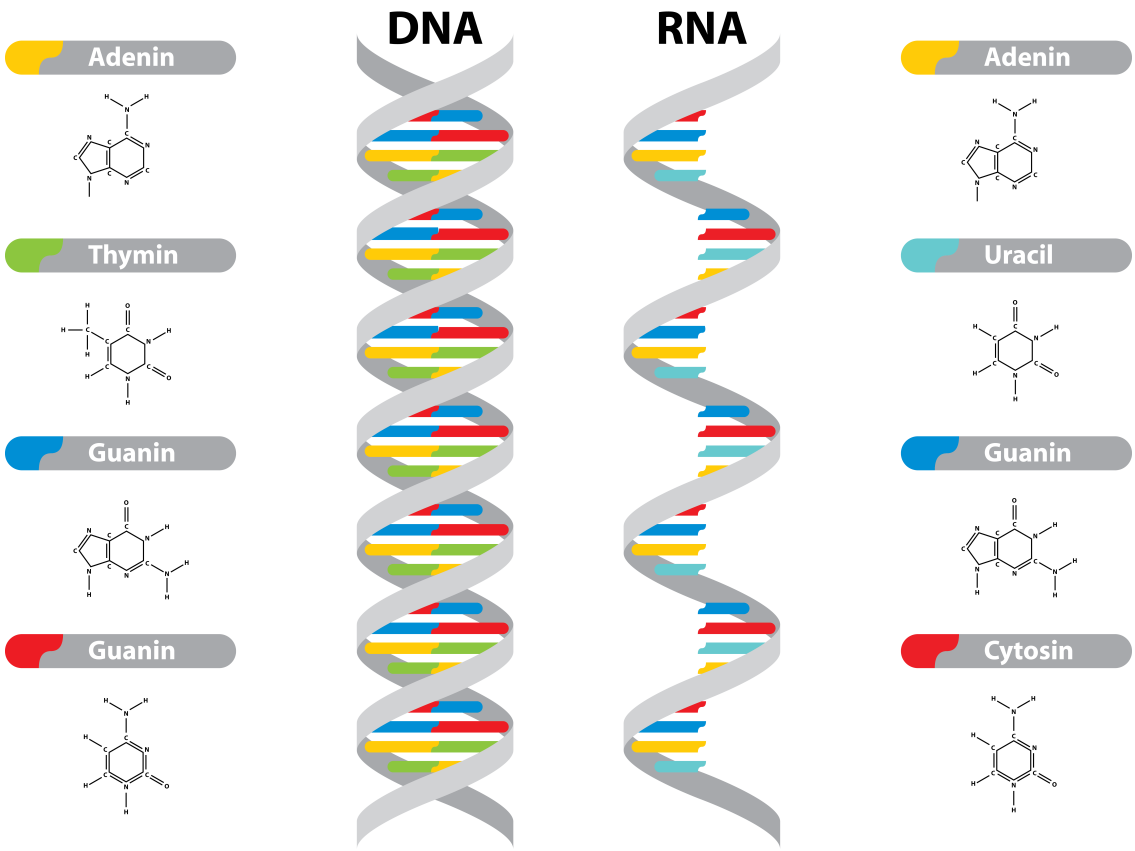


Abbildung 2.3 Unterschied zwischen DNA und RNA. DNA und RNA bestehen aus den Basen Adenin, Guanin und Cytosin. Die DNA verfügt zudem noch über die Base Thymin, die RNA über die Base Uracil. Während die DNA als Doppelhelix in der Zelle vorliegt und sehr stabil ist, ist die RNA zunächst einzelsträngig, kann aber mich sich selbst basenpaaren und ist instabiler als die DNA.

2.3.1 Die mRNA – Aufbau und Funktion des 5´UTRs

RNAs, welche die genetische Information für Proteine tragen, werden als *messenger* RNA (mRNA) bezeichnet. Es gibt grundlegende Unterschiede zwischen prokaryotischen und eukaryotischen mRNAs. Die eukaryotische prä-mRNA besteht aus dem 5´untranslatierten Bereich, einem codierenden Bereich mit Introns und Exons und dem 3´untranslatierten Bereich, welchem ein PolyA-Schwanz angehängt ist. Introns werden zwar transkribiert, jedoch aus der prä-mRNA herausgeschnitten, da sie keine Protein-codierenden Sequenzen tragen. Die eukaryotische mRNA ist zudem am 5´Ende mit einer Cap-Struktur versehen. Bei dieser Struktur handelt es sich meist um ein modifiziertes Guanin-Nukleotid. Das 5´Cap schützt die mRNA vor dem Abbau, ermöglicht den Transport der RNA aus dem Zellkern ins Cytosol mit Hilfe des Cap-Bindekomplexes (CBC) und erleichtert die Rekrutierung des 43S-Prä-Initationskomplexes (Furuichi et al. 1977; Kiledjian 2018; Hinnebusch 2014). Die so prozessierte mRNA wird auch als reife mRNA bezeichnet. An das 5´Cap

bindet der eukaryotische Translationsinitiationsfaktor eIF4E, welcher zusammen mit dem CBC die Initiation der Translation durch das Ribosom einleitet. Dem 5´Cap folgt der 5´ untranslatierte Bereich (5´UTR). Der prokaryotische 5´UTR verfügt über kein 5´Cap, das Ribosom bindet hier direkt an die Ribosomenbindestelle (RBS), welche 3-10 Basenpaare stromaufwärts des Startcodons liegt und als Shine-Dalgarno-Sequenz (SD) (AGGAGGU) bezeichnet wird (Saito et al. 2020). Im Gegensatz zu prokaryotischen mRNAs gibt es auf der eukaryotischen mRNA keine Ribosomenbindestellen. Der Initationskomplex, welcher aus der großen Untereinheit des Ribosoms und den zugehörigen Initationsfaktoren (eIFs) besteht, scannt hier den gesamten 5´UTR der mRNA nach einem Startcodon in einem geeigneten Kontext ab (Kozak 2002) (Altmann & Linder 2010; Hinnebusch 2011). Die Länge des 5´UTRs unterscheidet sich von Organismus zu Organismus. In Prokaryoten beträgt sie oft nur bis zu 10 Nukleotiden (nt), wohingegen eukaryotische 5´UTRs über 1000 nt lang sein können. Die Länge ist auch hier sehr variabel und kann sich selbst innerhalb eines einzigen Organismus von Gen zu Gen stark unterscheiden (Lin & W.-H. Li 2012). Die mittlere Länge des 5´UTRs in *Saccharomyces cerevisiae* beträgt zum Beispiel 96,5 ± 116,8 bp. Zu den wichtigsten regulatorischen Elementen im 5´UTR zählen Sekundärstrukturen, Bindestellen für RNA-bindende Proteine sowie *upstream* AUGs und *upstream* ORFs (*open reading frame*). Kurze und lange eingestreute Elemente (SINEs und LINEs) sowie einfache Sequenzwiederholungen (SSRs), Minisatelliten und Makrosatelliten sind in den eukaryotischen UTRs häufig (Pesole et al. 2001). Auch Sekundärstrukturen im 5´UTR fungieren als wichtiges Werkzeug der Genregulation. So konnten bei humanen mRNAs mit einer schlechten Translationseffizienz (wie Transkriptionsfaktoren, Protoonkogene, Wachstumsfaktoren) in der Mehrheit stabile Sekundärstrukturen in der Nähe des 5´Caps gefunden werden (Davuluri et al. 2000). Abbildung 2.3.1 zeigt den schematischen Aufbau einer eukaryotischen mRNA.

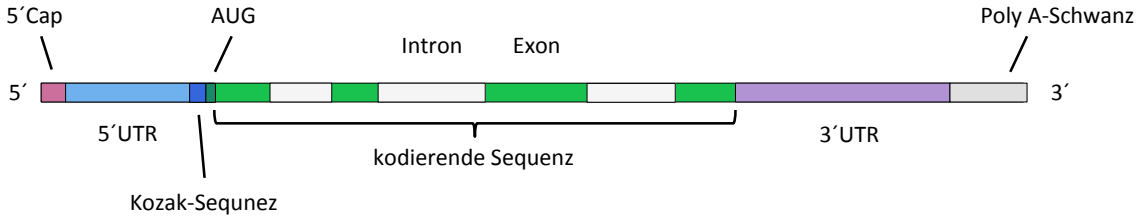


Abbildung 2.3.1 Aufbau einer eukaryotische mRNA. Die mRNA beginnt am 5´Ende und ist dort mit einer Cap-Struktur versehen, darauf folgt der 5´untranslantierte Bereich (5´UTR) mit der Kozak-Sequenz und dem Startcodon (AUG). Die kodierende Sequenz ist in Introns und Exons unterteilt, wobei nur die Exons die Information für ein Protein tragen. Die Introns sind in der reifen mRNA nicht mehr vorhanden, sie werden herausgespleißt. Durch alternatives Spleißen können durch das Zusammensetzen verschiedener Exons unterschiedliche Variationen eines Proteins entstehen. Nach der kodierenden Sequenz folgt schließlich der 3´untranslantierte Bereich (3´UTR) und der Poly A-Schwanz. Die mRNA endet am 3´Ende.

Wichtig für die Initiation der Translation und zudem ein Merkmal eukaryotischer mRNAs ist die Kozak-Sequenz, welche sich unmittelbar um das Startcodon herum befindet (Kozak 1986). Bezeichnet man das Adenin des Startcodons als Position +1, erstreckt sich die Kozak-Sequenz von Position -6 bis +6. Die optimale Kozak-Sequenz für hochexprimierte Gene in *S. cereviseae* wurde von Hamilton *et al.* (Hamilton et al. 1987) wie folgt definiert (J. Li et al. 2017):

(A/T)A(A/C)A(A/C)AATGTC(T/C)

Es konnte zudem gezeigt werden, dass der 5´UTR in *S. cerevisiae* eher reich an Adenin und arm an Guanin ist, vor allem in der Umgebung des Startcodons (Hamilton et al. 1987; Dvir et al. 2013). Mehrere Guanine haben einen hemmenden Einfluss auf die Proteinsynthese, wenn sie sich im Bereich der Kozak-Sequenz befinden (Valenzuela et al. 1979; Kniskern et al. 1986; J. Li et al. 2017). Durch Punktmutationen in diesem Bereich konnte jedoch gezeigt werden, dass auch bis zu 15 nt vor dem Startcodon die Translation negativ oder positiv beeinflusst werden kann. Es wurde zudem in dieser Studie herausgearbeitet, dass ein besonderes Augenmerk auf den Nukleotiden an den Positionen -11 bis – 14 liegt, da hier Mutationen die Translationseffizienz stark beeinflussten (J. Li et al. 2017).

2.3.2 Regulatorische RNAs

Neben der direkten Beteiligung an der Proteinbiosynthese übernehmen RNAs in der Zelle auch regulatorische Funktionen. Man unterscheidet zwischen *cis*-wirkenden RNAs und *trans*-wirkende RNAs. Die *cis*-wirkenden RNAs funktionieren als eigenständige RNA-Transkripte. So sind zum Beispiel die eukaryotischen miRNAs (*micro*RNAs) und siRNAs (*small interfering* RNAs) an einem regulativen Mechanismus beteiligt, den man als RNA-Interferenz (RNAi) bezeichnet. Bei den exogen in die Zelle eingebrachten siRNAs führt eine perfekte Basenpaarung mit dem 3‘UTR von mRNAs zu deren Abbau. miRNAs werden dagegen in der Zelle gebildet und führen durch eine imperfekte Basenpaarung mit dem 3‘UTR der mRNA zu einer Repression der Translation oder deren Abbau (Agrawal et al. 2003). Eine weitere Gruppe kleiner RNAs bilden die snoRNAs. Diese *small nucleolar* RNAs liegen, wie sich vom Namen ableiten lässt, meist im Kern vor und sind mit Riboproteinen assoziiert. Sie sind unerlässlich für das Funktionieren des Ribosoms und zudem an der Methylierung der rRNA und am Spleißen beteiligt (Falaleeva et al. 2017).

In prokaryotischen Zellen wirken *cis*-regulierende sRNAs (*small* RNAs) ebenfalls über Basenpaarung mit ihrer Ziel-RNA. Von ihnen werden unter anderem Stoffwechsel, Replikation oder die Stabilität

von Plasmiden reguliert (Brantl 2007). Den *cis*-wirkenden regulatorischen RNAs stehen die *trans*-wirkenden regulatorischen RNAs gegenüber, welche in der RNA, die sie regulieren, integriert sind. Diese *trans*-wirkenden RNAs werden in den folgenden Kapiteln genauer beschrieben.

2.4 Natürliche Riboswitche

Riboswitche wurden zunächst synthetisch entwickelt, bevor ihre natürliche Herkunft 2002 nachgewiesen werden konnte (Breaker 2012). Riboswitche sind RNA-basierte regulatorische Elemente, welche auf einen externen Metaboliten reagieren können. Riboswitche kommen in allen Domänen des Lebens vor, werden jedoch besonders häufig in Prokaryoten gefunden. Oft handelt es hier um RNA-Strukturen, welche zelluläre Metabolite binden können und hierdurch die zu diesen Metaboliten gehörenden Stoffwechselwege, zum Beispiel die von Vitaminen, Aminosäuren und Nukleotidanaloga, kontrollieren (Breaker 2012; Serganov & Nudler 2013). Durch die Bindung des Metaboliten an die RNA wird die Genexpression beeinflusst. Als Liganden dieser natürlich vorkommenden Riboswitche dienen unter anderem Anionen, Metalle, Purine und ihre Derivate, Cofaktoren und Vitamine sowie Aminosäuren. Durch das Binden der Liganden an die RNA-Strukturen wird die Genexpression beeinflusst und ein metabolisches Gleichgewicht kann aufrechterhalten werden. Ein schematisches Bild eines Riboswitches, welcher aus Aptamerdomäne und Expressionsplattform besteht, ist in Abbildung 2.4.A dargestellt. Die regulatorische Aktivität der Riboswitche basiert auf der ligandenabhängigen Bildung von zwei unterschiedlichen, sich gegenseitig ausschließenden, RNA-Konformationen. Die RNA-Strukturen können als Rho-unabhängige und Rho-abhängige Terminator-Haarnadeln die Transkription beeinflussen (Hollands et al. 2012). Die Translation hingegen kann durch eine RNA-Struktur-bedingte Sequestrierung der Ribosomenbindungsstelle gesteuert werden.

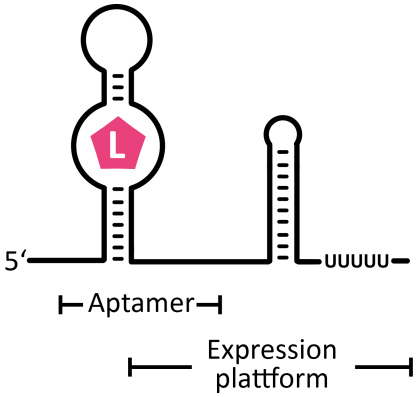


Abbildung 2.4.A Riboswitch-Schema. Das Aptamer ist im 5´UTR einer mRNA platziert und hat den Liganden (L) gebunden. Das Aptamer wirkt als Sensor-Domäne, welches die Information des Bindeereignisses an die Expressionsplattform weitergibt.

Auch über einen RNA-spaltenden Mechanismus kann die Translation beeinflusst werden. Das *glm*S-Ribozym interagiert mit Glucosamin-6-phosphat, welches nach Bindung die *glm*S-mRNA innerhalb der Riboswitch-Sequenz spaltet. Die gespaltene mRNA wird anschließend von der RNase J abgebaut und so die Translation der mRNA verhindert (Collins et al. 2007).

In eukaryotischen Zellen können Riboswitche die Genexpression durch alternatives Spleißen regulieren, wie dies zum Beispiel durch TPP-Riboswitche in Pilzen erfolgt (Cheah et al. 2007). Ohne TPP in der Zelle findet ein normaler Spleißvorgang statt und es entsteht ein Protein in voller Länge. Wenn dagegen TPP in einer bestimmten Schwellenkonzentration in der Zelle vorhanden ist, bindet es an den sensorischen Teil des Riboswitches und macht eine zuvor verdeckte Spleißstelle zugänglich. Die alternativ gespleißten mRNAs können zum Beispiel Stoppcodons enthalten, was zu einer vorzeitigen Termination führen kann. Bei Pflanzen wird durch den TPP-Riboswitch, welcher sich hier im 3´UTR befindet, die alternative Polyadenylierung und mRNA-Stabilität reguliert (Bocobza et al. 2007; Xue et al. 2013; Wachter et al. 2007).

In Säugetierzellen konnte man den bekannten Riboswitchen ähnliche RNA-Switche nachweisen (Venkata Subbaiah et al. 2019). Diese Switche werden *compound protein-directed RNA switches* (PDRS) genannt und können ebenfalls die Genexpression regulieren. Es handelt sich dabei um recht komplexe Systeme, welche durch sich gegenseitig ausschließende Interaktionen von mindestens zwei Sätzen RNA-bindender Proteine oder Protein-Komplexe mit denselben oder benachbarten *cis*-wirkenden RNA-Elementen angetrieben werden. So kann zum Beispiel PTBP/hnRNPI an eine CU-reiche Sequenz binden und damit die lokale Sekundärstruktur, zum Beispiel eine Stammschleife, stören (Abbildung 2.4.B). Dadurch wird die RNA dem für den miRNA-induzierenden *Silencing*-Komplex erkennbar und diesem ausgesetzt (Xue et al. 2013).

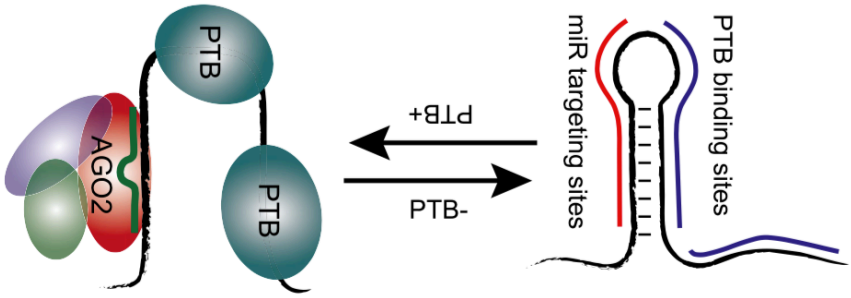


Abbildung 2.4.B Vorgeschlagenes Modell zur PTB-vermittelten Öffnung der Stammschleife. Die Abbildung wurde aus Xue *et al.* übernommen (Xue et al. 2013). PTB scheint die Exposition der mikroRNA Zielstelle durch die Bindung an pyrimidinreiche Bereiche der RNA zu durch Öffnung der Stammschleife zu induzieren, Ago2 kann dann an die Zielsequenz binden und die RNA schneiden.

Erwähnenswert für diese Doktorarbeit ist ein natürlich vorkommender Tandem-Riboswitch, der Glycin-Riboswitch, welcher bei mindestens 350 verschiedenen Bakterienarten vorkommt (Kazanov et al. 2007). Der Glycin-Riboswitch existiert meist in einer Tandem-Struktur, bei der zwei benachbarte, homologe Aptamere durch eine kurze Linkerregion verbunden sind, gefolgt von einer Expressionsplattform. Beide Glycin-Aptamere binden Glycin im mikromolaren Bereich. Der Tandem-Riboswitch dient in der Zelle als Sensor für Glycin. Glycin wird als Folge hoher Glycin-Konzentrationen in der Zelle abgebaut bzw. importiert, wenn die Konzentration in der Zelle zu niedrig wird. Funktional hängen die beiden Glycin-Aptamere, welche nicht exakt die gleiche Struktur, wohl aber die gleiche Bindetasche, aufweisen, voneinander ab. Ein Modell hierfür wurde von Ruff *et al.* 2016 vorgeschlagen: Im Liganden-ungebundenen Zustand ist die Dimerisierung der beiden Aptamerdomänen eher ungünstig. Der P1-Stamm des zweiten Aptamers wird hier nicht ausgebildet und interagiert mit der Expressionsplattform (Abbildung 2.4.C). Bei Zugabe von Glycin stabilisiert die Bindung des ersten Aptamers an Glycin den P1-Stamm des zweiten Aptamers, welcher die Expressionsplattform kontrolliert. Es wurde vorgeschlagen, dass sich die Dimerisierung der beiden Domänen thermodynamisch günstig auswirkt und zudem Energie liefert, um die veränderte Konformation der Expressionsplattform zu kompensieren (Ruff & Strobel 2014).

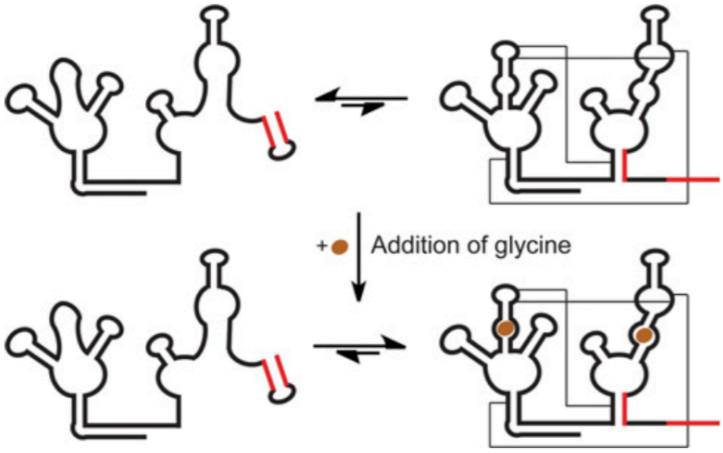


Abbildung 2.4.C Glycin-Tandem Riboswitch. Glycin-Tandem Riboswitch, Abbildung aus Ruff & Strobel *et al.* (Ruff & Strobel 2014). Im vorgeschlagenen Modell zur Funktion des Glycin-Tandem-Ribsowitches befinden sich die beiden Domänen in Abwesenheit von Glycin in einem energetisch ungünstigen Zustand. Der P1-Stamm des zweiten Aptamers wird hier zum größten Teil nicht gebildet, er faltet dagegen meist mit der nachfolgenden Expressionsplattform. In Anwesenheit von Glycin verschiebt sich das Gleichgewicht, jetzt wird der P1-Stamm des zweiten Aptamers ausgebildet, der durch die Dimerisierung der Aptamere stabilisiert wird.

2.5 Synthetische Schalter und Möglichkeiten der Genregulation auf RNA-Ebene

Ähnlich den zahlreich vorkommenden natürlichen Riboswitchen bieten synthetischen Riboswitche innerhalb der Synthetischen Biologie eine Vielzahl von Möglichkeiten zur Genregulation. Synthetische Schalter können sowohl in Prokaryonten als auch in Eukaryonten zum Einsatz kommen. Sie können als Transkriptionskontrolle oder Translationskontrolle und sowohl im 5´UTR als auch im 3´UTR auf unterschiedliche Weise wirken.

In dieser Arbeit geht es um Riboswitche, welche im 5´UTR einer mRNA zum Einsatz kommen. Hier ist der sensorische Teil eines Riboswitches das Aptamer, welches aus einer RNA-Struktur besteht. RNA kann eine komplexe dreidimensionale Strukturen annehmen und so die chemischen Seitenketten präzise präsentieren. Dies ist eine wesentliche Eigenschaft, damit sie als biologischer Katalysator, Regulator oder strukturelles Gerüst fungieren können (F. Groher & Suess 2014). Diese Fähigkeit ermöglicht es einem Aptamer, seinen Liganden in einer vorgeformten Bindungstasche mit hoher Affinität und Spezifität zu binden. In den folgenden Kapiteln sollen jedoch zunächst die unterschiedlichen Möglichkeiten aufgezeigt werden, wie die Genexpression in Pro- und Eukaryonten durch synthetische Schalter kontrolliert werden kann. Da Aptamere in vielen Fällen einen wesentlichen Bestandteil von Riboswitchen darstellen, wird zunächst auf die Selektion von Aptamere eingegangen und das Tetrazyklin-Aptamer, welches in dieser Arbeit hauptsächlich verwendet wurde,

detailliert beschrieben. Anschließend werden unterschiedliche Strategien verschiedener synthetischer Riboswitche aufgezeigt.

2.6 Selektion synthetischer Aptamere durch SELEX und das Engineering von Riboswitchen

Aptamere können prinzipiell gegen jeden beliebigen Liganden generiert werden. Es gibt zwei Möglichkeiten, um synthetische Aptamere zu generieren, entweder durch Modifikation natürlich vorkommender Aptamere oder durch einen *in vitro* Selektionsprozess, genannt SELEX (systematic evolution of ligands by exponential enrichment). Hier werden Aptamere aus einem Pool von 10^{12}-10^{15} randomisierten RNA-Molekülen gegen einen bestimmten Liganden selektiert. Die SELEX wurde von verschiedenen Arbeitsgruppen bereits 1990 unabhängig voneinander etabliert (F. Groher & Suess 2016; Ellington & Szostak 1990; Tuerk & Gold 1990). Die *in vitro* Selektionen gehen von einer initialen, chemisch synthetisierten und stark amplifizierten kombinatorischen Bibliothek von DNA-Oligonukleotiden aus, die zuerst in RNA transkribiert wird. Der Prozess selbst ist iterativ und geprägt von dem wiederholten Zusammenbringen der RNA mit dem Zielmolekül, dem anschließenden, Entfernen nicht-bindender Spezies durch mehrere Waschschritte und der spezifischen Elution der randomisierten RNA (Abbildung 2.6.B). Anschließend wird diese wieder amplifiziert und erneut mit dem Zielmolekül zusammengebracht. Hierdurch wird eine Anreicherung derjenigen RNAs erreicht, die spezifisch an das Zielmolekül binden. Um Aptamere zu erhalten, die ihr Zielmolekül mit der höchsten Affinität binden können, wird die Stringenz der Wasch- und Elutionsschritte graduell erhöht.

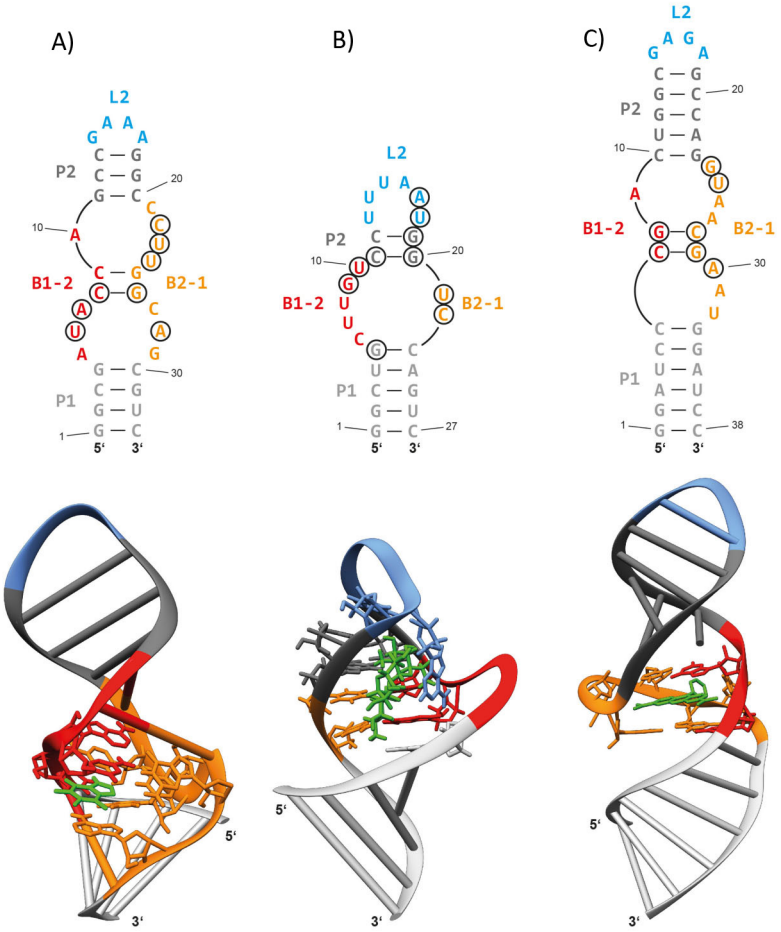


Abbildung 2.6.A Beispiel dreier Aptamere, welche häufig in Riboswitchen Verwendung finden. Zu sehen sind das **A)** Theophyllin-Aptamer, **B)** das Neomycin-Aptamer und **C)** das Malachitgrün-Aptamer in ihrer 2D und 3D Sekundärstruktur. Stämme werden mit einem P gekennzeichnet, Ausbuchtungen (*bulges*) sind in mit einem B gekennzeichnet und Schleifen (*loops*) sind mit einem L gekennzeichnet. Für die Bindung wichtige Nukleotide sind umrandet (2D) bzw. hervorgehoben (3D) und die Liganden sind in grün dargestellt. Die Abbildung wurde aus der Dissertationsschrift von Dr. Florian Groher übernommen und leicht abgeändert.

Obwohl inzwischen mehrere dutzend Aptamere erfolgreich selektiert wurden, die niedermolekulare Stoffe binden können (Stoltenburg et al. 2007), konnten nur wenige von ihnen als Riboswitche eingesetzt werden (Wittmann & Suess 2012). Häufig werden die Aptamere gegen Theophyllin, Neomycin (NEO) und Malachitgrün in Riboswitchen verwendet (Abbildung 2.6.A). Da eine Ligandenbindung immer mit Konformationsänderungen der Aptamerstruktur einhergeht, wird auch die Zielsequenz (auch Kontext genannt) in der mRNA beeinflusst. Umgekehrt beeinflusst der Kontext auch die Struktur des Aptamers und kann diese sogar komplett verändern und so inaktivieren. Anhand von NMR-spektroskopischen Analysen des NEO-Aptamers konnten genau solche Konformationsänderungen aufgezeigt werden (Duchardt-Ferner et al. 2010). Ein gut bindendes Aptamer muss nicht zwangsläufig mit einer Regulationsfähigkeit einhergehen. Dass ein Aptamer zwar

einen Liganden hochaffin binden kann, zeitgleich aber keine regulatorischen Eigenschaften aufweist, das heißt keine Konformationsänderung mehr in der Zielsequenz zeigt, konnte nach spektroskopischen Analysen des NEO-Riboswitches aufgezeigt werden (Weigand et al. 2010). Um aus einem generierten RNA-Pool also erfolgreich Riboswitche zu identifizieren, muss ein *in vivo* Screening der angereicherten RNAs folgen. Nur mit diesem, im Zielorganismus durchgeführten, Screening ist es möglich, Aptamere mit potentiell regulatorischen Eigenschaften zu identifizieren (Weigand et al. 2007). Für ein erfolgreiches Screening ist es jedoch notwendig, eine möglichst hohe Vielfalt an verschiedenen RNA-Sequenzen zu testen. Jedoch wirkt der durch die *in vitro* Selektion zwangsläufig aufgebaute Selektionsdruck eben dieser Vielfalt entgegen und in den meisten Fällen manifestieren sich gegen Ende der Selektion zwar gut bindende, aber nicht zwangsläufig gut regulierende RNAs. In einer Folgestudie wurde die Methode verbessert, indem sie um ein *next generation sequencing* erweitert wurde. Dies ermöglichte es, diejenigen angereicherten Selektionsrunden auszuwählen, welche noch eine möglichst große Vielfalt an RNA-Sequenzen aufweisen. So gelang es, ein Ciprofloxacin bindendes Aptamer zu finden (F. Groher et al. 2018).

Mit der Kombination der SELEX, des darauffolgenden Screenings und dem *next generation sequencing* ist es zwar prinzipiell möglich, gut funktionierende Riboswitche zu finden, jedoch ist der Aufwand enorm. Eine neue Methode schlägt eine Brücke zwischen SELEX und Screening. Dabei handelt es sich um eine einfache, schnellere Möglichkeit, nicht nur gut bindende, sondern auch strukturierte und umschaltende Aptamere zu finden, um schneller passende Riboswitche bilden zu können, die Capture-SELEX (Boussebayle, Torka, et al. 2019).

Bei einer herkömmlichen SELEX wird der Ligand an eine Säule immobilisiert. Dieser Vorgang bietet vor allem für kleinere Moleküle Nachteile. Wenn das Zielmolekül beispielsweise durch eine Biotinylierung an die Matrix gebunden wird, können einige chemische Gruppen des Liganden nicht mehr durch die RNAs erkannt und gebunden werden, da diese durch den Biotin-Linker behindert werden und so nicht zugänglich sind. Zudem besteht auch die Möglichkeit, dass die RNAs an den Biotin-Teil binden und nicht (nur) an den Liganden. Zudem können die zum Teil stark denaturierenden Bedingungen, unter denen der Immobilisierungsprozess durchgeführt wird, das Zielmolekül noch vor Beginn des iterativen SELEX-Prozesses beeinträchtigen. Bei der Capture-SELEX wird dagegen der RNA-Pool über eine kurze definierte Sequenz, die sogenannte Docking-Sequenz, über Watson-Crick-Interaktionen an ein komplementäres Oligonukleotid, das sogenannte Capture-Oligonukleotid, gebunden, welches in diesem Fall als eine Art Anker fungiert. Das Capture-Oligonukleotid ist mit einem Biotin-Tag versehen und dieser Komplex kann so auf Streptavidin-beschichteten Magnetkugeln angeheftet werden. Werden die Pool-RNAs nun mit dem in Puffer gelösten, nativen Liganden inkubiert, binden einige RNAs an den Liganden und werden in Folge dessen vom Capture-Oligonukleotid abgekoppelt. So kann nicht nun der Ligand in seiner

ursprünglichen Form verwendet werden, alle seine chemischen Gruppen sind frei zugänglich und das Aptamer erfährt zudem eine Umformung. Diese Umformung ist wichtig, denn es erleichtert den späteren Einsatz des Aptamers als Riboswitch, bei welchem durch die vom Liganden verursachte Veränderung der RNA die Genexpression gesteuert werden kann. Dabei ist es wichtig, dass die Bindung des Liganden relativ nah an der Docking-Sequenz stattfindet. So wird sichergestellt, dass ein Abkoppeln der RNA eine Folge der Ligandenbindung darstellt (Boussebayle, Groher, et al. 2019).

Es konnten kürzlich die Vorteile dieser relativ neuen Selektionsmethode aufgezeigt werden, indem ein gut funktionierender Riboswitch gegen Paromomycin (ein Aminoglykosid-Antibiotikum) entwickelt wurde, welcher zuvor mittels Capture-SELEX selektiert worden war. Abbildung 2.6.B zeigt die herkömmliche SELEX-Methode im Vergleich zur neueren Capture-SELEX-Methode.

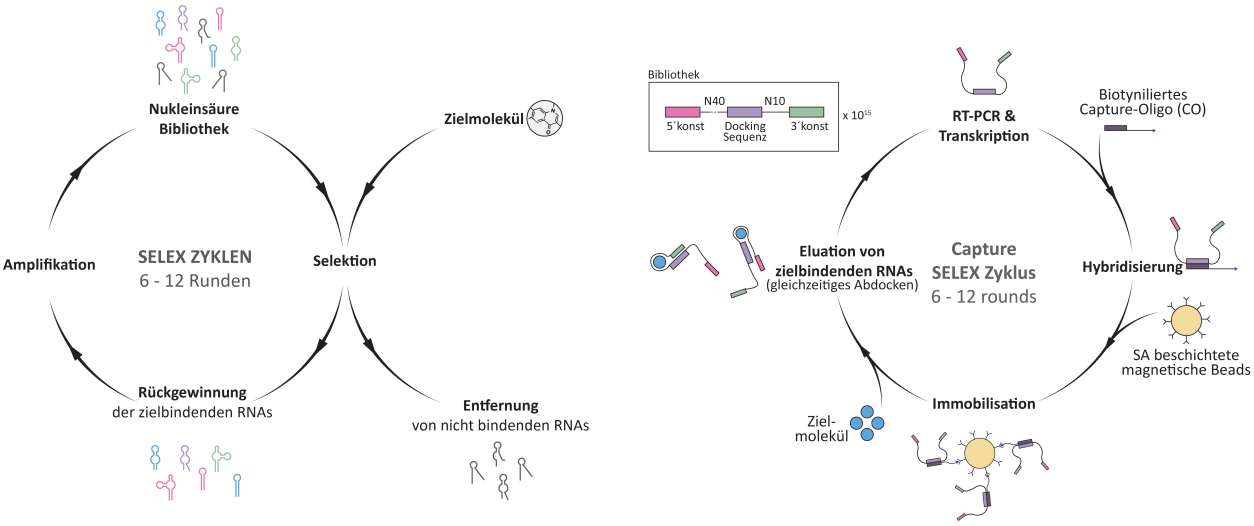


Abbildung 2.6.B Vergleich herkömmlicher SELEX mit der Capture-SELEX. In der Abbildung sind die Unterschiede der beiden SELEX-Arten dargestellt. Bei der herkömmlichen SELEX (links) ist der Ligand über einen Linker an der Säule fixiert. RNAs, die an den Liganden binden, verbleiben beim Waschen an diesem, während nicht bindende RNAs heruntergewaschen werden. Die an den Liganden gebundenen RNAs werden im nächsten Schritt eluiert, amplifiziert und erneut auf die Säule gegeben. Nach einigen Runden reichern sich die RNAs im Pool an, die den Liganden affin binden können. Im Unterschied dazu sind bei der Capture-SELEX (rechts) die RNAs über eine Dockingsequenz und ein Capture-Oligonukleotid an die Säule gebunden. Bei Zugabe des Liganden binden die Aptamere an diesen und entkoppeln sich so vom Capture-Oligonukleotid. Die zu dieser Abkopplung notwendige Konformationsänderung erleichtert den späteren Einsatz der Aptamere als Riboswitch. Abbildung wurde verwendet und leicht abgeändert mit freundlicher Genehmigung von Dr. Florian Groher.

2.7 Das Tetrazyklin-Aptamer

Das Tetrazyklin (TC)-Aptamer wurde 2001 durch SELEX gefunden. Bei dem Liganden des Aptamers, TC (Abbildung 2.7.A), handelt es sich um ein in Streptomyceten produziertes Antibiotikum, das die prokaryotische Translation hemmt, indem es die Bindung der Aminoacyl-tRNA an der ribosomalen A-Stelle stört. Es ist ein weit verbreiteter therapeutischer Wirkstoff von geringer Toxizität, der gegen viele Krankheitserreger wirksam ist und auch in der Tierernährung zur Stimulierung der Gewichtszunahme sowie zur prophylaktischen Krankheitsbekämpfung eingesetzt wurde (Berens et al. 2001).

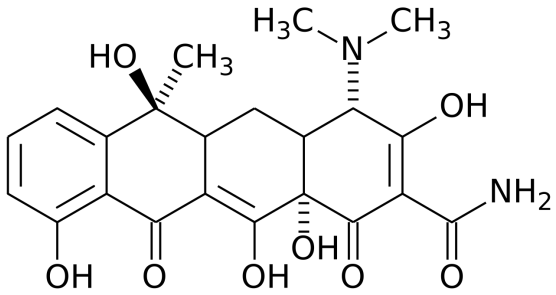


Abbildung 2.7.A Strukturformel von Tetrazyklin. Das Grundgerüst besteht aus 4 Kohlenstoff-Sechsringen. Von https://de.wikipedia.org/wiki/Tetracyclin

Der RNA-Pool, mit welchem das TC-Aptamer selektiert wurde, bestand aus randomisierten RNAs, welche 74 nt lang waren und von einer konstanten Region flankiert wurden. Unter den vielen verschiedenen Sequenzen, die gefunden wurden, gab es mehrere Klone, die sich nur in wenigen Basen unterschieden. Nachdem man die Klone mittels Affinitätssäule auf ihre Bindung getestet hatte, konnte ein Aptamer identifiziert werden, welches TC mit einem Kd von ca. 1 µM bindet.

Das in dieser Studie gefundene Aptamer besteht aus 3 Helices (im Weiteren auch Stämme genannt), P1-P3, zwei einzelsträngigen Regionen B1-2 und B2-3 und den Loops L2 und L3, welche die distalen Enden der Stämme P2 und P3 schließen (Abbildung). Die Regionen B1-2 und L3 sind für die Ligandenbindung verantwortlich. Die Stämme P1, P2 und P3 sind an der Ligandenbindung nicht beteiligt. Die Stämme P1 und P2 können verändert werden, ohne die Bindung des Liganden zu beeinträchtigen (Suess et al. 2003, Hanson et al. 2003).

Die Aufklärung der Kristallstruktur des TC-Aptamers im Jahre 2008 ermöglichte eine genaue Vorstellung des Aptamers und seiner Bindungseigenschaften. Die Bindungstasche des Aptamers liegt an der Kreuzung von zwei Helixstapeln, wobei die Faltung des TC-Aptamers der eines umgekehrten „h“ ähnelt. Die beiden Helices P1 und P3 stapeln aufeinander. Durch die Interaktion der Verbindungsbereiche B1/2 und B2/3 entsteht eine unregelmäßige Helix, auf die P2 aufbaut. Im Liganden-gebundenen Zustand des Aptamers, kommt es zu einer Wechselwirkung zwischen dem

Loop L3 und der kleinen „Einkerbung“ der unregelmäßigen Helix. Die Nukleotide (nt) aus den Verbindungsbereichen B1/2 und B2/3 bilden mit 3 Nukleotiden aus L3 ungewöhnliche Basenpaare, es bildet sich ein nicht kanonischer Pseudoknoten. Das Antibiotikum selbst bindet als Magnesiumchelat an die RNA des Aptamers, wobei die Bindungstasche des Aptamers aus genau diesem Bereich zwischen B1/2, B2/3 und L3 gebildet wird. Bei der Bindung des Aptamers an den Liganden spielt das Magnesiumion eine bedeutende Rolle. Generell spielen Magnesiumionen bei der Interaktion von RNA mit RNA sowie anderen Molekülen eine zentrale Rolle, da diese die negativen Ladungen kompensieren. Diesen Umstand findet man auch in den eukaryotischen Riboswitchen, der Thiaminpyrophosphat-bindenden Box und des Glucosamin-6-Phosphat-bindende glmS-Ribozyms (Xiao et al. 2008). Abbildung 2.7.B zeigt das TC-Aptamer in seiner 2-D und 3-D Struktur.

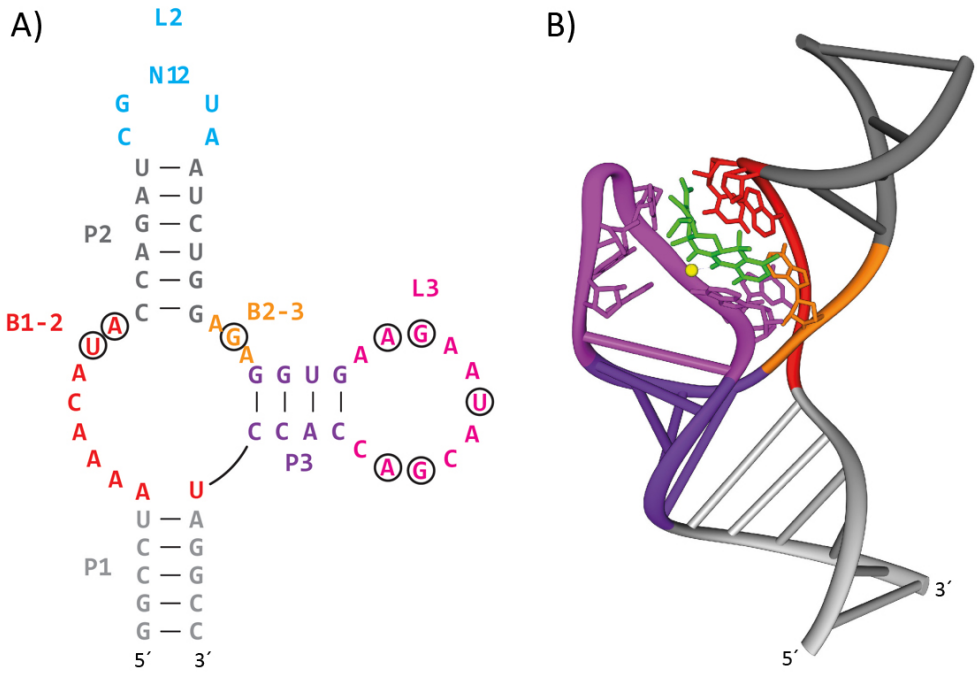


Abbildung 2.7.B 2D (A) und 3D (B) Struktur des TC-Aptamers. Die Stämme sind mit einem P gekennzeichnet, Ausbuchtungen (*bulges*) sind in mit einem B und Schleifen (*loops*) sind mit einem L gekennzeichnet. Die einzelsträngigen Bereiche B1-2 (rot), B2-3 (gelb und L3 (rosa) sowie die Helix P3 sind an der Ligandenbindung beteiligt. Für die Bindung wichtige nt (Nukleotide) sind umrandet (2D) bzw. hervorgehoben (3D) und der Ligand TC ist in grün dargestellt, Magnesiumion als gelber Ball. Die Abbildung wurde aus der Dissertationsschrift von Dr. Florian Groher übernommen und leicht abgeändert.

Mit Hilfe von Stopped-Flow-Messungen wurde die Bindungsdynamik des Aptamers mit seinem Liganden analysiert. Mit diesem Verfahren ist eine Überwachung dynamischer Prozesse im Millisekunden- bis Sekundenbereich möglich. So konnte festgestellt werden, dass TC, ähnlich wie

viele Proteine, in einem zweistufigen Prozess an das Aptamer bindet, bei dem eine Konformationsänderung induziert wird. Allerdings ist das TC-Aptamer innerhalb des mRNA-Kontext bereits vorstrukturiert und die Konformationsänderung bei Ligandenbindung ist gering. Der erste Schritt der Bindung verläuft reversibel, der zweite Schritt nahezu irreversibel. Dabei ist die Bindungsdynamik extrem schnell und nimmt mit zunehmender TC-Konzentration zu. Eine vollständige Ligandenbindung wird bereits unter 50 ms induziert (Förster et al. 2011). Eine Mutation der Basen A13 und A50 innerhalb der an der Bindung beteiligten Bereiche B1-2 und L3 führte zu einer starken Erhöhung der Dissoziationskonstante (KD), was auf eine starke Beteiligung dieser Basen am Bindungsprozess schließen lässt. Auch die Base A58 ist für eine schnelle Bindungskinetik und eine stabile Konformation, welche die Dissoziation von TC verhindert, unerlässlich. Ebenso ist die Base A9 von großer Bedeutung für die Bindungsdynamik, da sie eine Gerüstfunktion für die an der Bindung beteiligten Basen einnimmt und an der Bildung der Tertiärstruktur beteiligt ist. Eine Mutation dieser Base macht das Aptamer regulatorisch inaktiv. Bei der Mutation von A9 war es zu einer dramatischen Erhöhung der Rückreaktionsrate gekommen. Es wird diskutiert, dass dies der Fall ist, weil es durch die Mutation dieser Base zu einer Vielzahl an möglichen Aptamer-Strukturen kommt, bei welchen es zu keiner korrekten Ausbildung der Bindungstasche kommen kann. Auch Mutationen an den Positionen 13 in B1 und 49, 50 und 56-58 in L3 führen zu einem vollständigen Verlust der Regulation, unabhängig von dem eingeführten Nukleotidaustausch (Hanson et al. 2005). Es konnte bestimmt werden, das in einem stöchiometrischen Verhältnis von 1:1 die Affinität von Aptamer und Ligand 0,8 nM beträgt (Förster et al. 2011).

In den nächsten Kapiteln wird darauf eingegangen, in welch vielfältiger Weise Aptamere für die Kontrolle der Genexpression bereits eingesetzt wurden. Neben dem TC-Aptamer ist das Theophyllin-Aptamer das wohl bekannteste Aptamer. Ähnlich wie das TC-Aptamer wurde es sowohl als Riboswitch sowie auch als Ribozym zur Kontrolle der Genexpression eingesetzt.

2.8 RNA-basierte Translationskontrolle in Prokaryoten durch Riboswitche

In Prokaryoten markiert die SD-Sequenz die Ribosomenbindestelle (RBS) und somit den Startpunkt der Translation. Durch die unterschiedlichen Konformationen, die ein Riboswitch einnehmen kann, kann die SD-Sequenz maskiert oder freigesetzt werden und so die Zugänglichkeit des Ribosoms zur Ribosomenbindestelle (RBS) kontrolliert werden (Abbildung 2.8). Im Jahr 2004 konnte ein synthetischer Riboswitch basierend auf diesem Wirkmechanismus entwickelt werden (Desai & Gallivan 2004). Ein Theophyllin-Aptamer wurde hierfür stromaufwärts der SD-Sequenz eines Reportergens gesetzt. Im Theophyllin-ungebunden Zustand ist die RBS für das Ribosom zugänglich.

Bei Bindung des Liganden Theophyllin an das Aptamer ändert sich die Konformation der RNA und es bildet sich so ein doppelsträngiger Bereich auf der RNA aus, der die SD-Sequenz maskiert, die RBS wird unzugängig, das Ribosom kann nicht mehr an die RNA binden und die Translation wird inhibiert (Desai & Gallivan 2004).

Um die Transkriptionstermination in *E. coli* zu kontrollieren, wurde ein Theophyllin-Aptamer gefolgt von einer Uridin-reichen-, dem Aptamer und Terminator zum Teil komplementären Sequenz, verwendet (Wachsmuth et al. 2013). Im Theophyllin-ungebunden Zustand wird ein Terminator-Stamm ausgebildet. Bei Bindung des Aptamers kommt es wiederum zu einer Konformationsänderung, was zur Folge hat, dass das Aptamer Teile des Terminator-Stamms für sich beansprucht und die Termination verhindert wird (Abbildung 2.8 C).

Einen weiteren Ansatz der synthetischen RNA-basierten Translationskontrolle in Prokaryoten bietet die Regulation über *trans*-exprimierte RNAs, die mit Bereichen im 5'UTR der mRNA interagieren. Die Expressionsplattform beinhaltet die sogenannte crRNA (CRISPR-RNA), welche durch die Ausbildung einer Stammschleife über Watson-Crick-Basenpaarung die RBS sequestriert und somit die Ribosomenbindung an die mRNA verhindert. Die *trans*-exprimierte taRNA (trans-aktivierende RNA) ist komplementär zu Teilen der die RBS-blockierenden Stammschleife und kann über kanonische Basenpaarung die RBS freilegen und die Translationsinitiaton ermöglichen (Callura et al. 2012; Isaacs et al. 2004). Eine Weiterentwicklung dieses Riboregulators stellt der von Green *et al.* vorgestellte *toeholdswitch* dar (A. A. Green et al. 2014). Die *trans*-RNA, hier Trigger-RNA genannt, ist im Gegensatz zum herkömmlichen Riboswitch nicht länger komplementär zur RBS, sondern zu anderen, der RBS benachbarten Sequenzbereichen des 5´UTRs. Die RBS und das Startcodon sind hier in einzelsträngigen Bereichen einer Stammschleife verborgen (Abbildung 2.8).

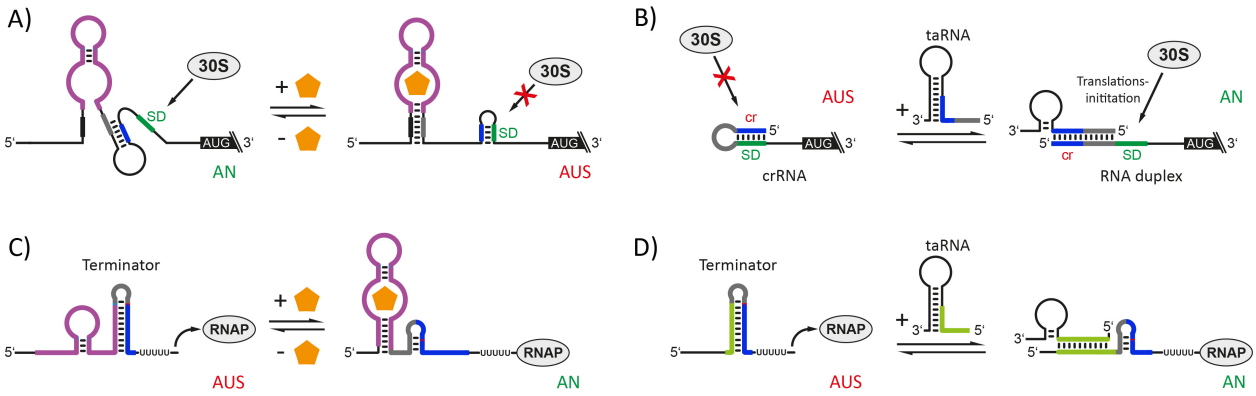


Abbildung 2.8 Mechanismen von Riboswitchen in porkaryotischer mRNA. A) In Abwesenheit des Liganden ist die SD-Sequenz frei zugänglich, das Ribosom kann an die mRNA binden und die Translation kann stattfinden. Bei Anwesenheit des Liganden bildet sich der Stamm des Aptamers aus, die zur SD-Sequenz zum Teil komplementäre Struktur ist nun frei und bildet mit der SD-Sequenz eine kleine Stammschleife aus, diese ist nun nicht mehr für die 30S Untereinheit des Ribosoms zugänglich, die Translation kann nicht statt finden. **B)** In Abwesenheit des Liganden bildet sich der Transkriptionsterminator aus, in Anwesenheit des Liganden kann sich die Struktur des Terminatorstamms dagegen nicht ausbilden, da ein Teil der Sequenz das Liganden bindende Aptamer bildet. Die Transkription wird nicht terminiert und die Expression des Gens kann fortgesetzt werden. **C)** In Abwesenheit der *trans*-regulierenden RNA taRNA ist die SD-Sequenz nicht für die 30S Untereinheit des Ribosoms zugänglich, da sie mit einer kurzen Sequenz, dem *cis-repressor* (cr) basenpaart. In Anwesenheit der taRNA bindet die Repressorsequenz mit der taRNA und die SD-Sequenz ist für die Bindung der 30S-Untereinheit des Ribosoms zugänglich. **D)** Die taRNA kann auch mit entsprechenden Sequenzen die Transkription regulieren. In Abwesenheit der taRNA bildet sich der Terminatorstamm aus, die Transkription wird unterbrochen. In Anwesenheit der taRNA kann sich der Terminatorstamm nicht aus weil die taRNA mit einem Teil der Stammsequenz basenpaart. Die Abbildung wurde aus der Dissertationsschrift von Dr. Florian Groher übernommen und leicht abgeändert.

2.9 Die Kontrolle der Genexpression durch Ribozyme

Aptazyme sind sich selbst spaltende, allosterisch regulierende Ribozyme. Eingefügt in den 3´UTR einer eukaryotischen mRNA kann die Genexpression durch Zugabe eines Liganden gesteuert werden. Durch rationales Design wurde ein ATP-Aptamer mit einem Hammerhead-Ribozym aus *Schistosoma mansoni* fusioniert und die Expression eines Reportergens in Mäusen kontrolliert. Dabei wurde das Ribozym direkt vor die Translationsstartstelle gesetzt. Über einen Linker-Bereich wurden Aptamer und Ribozym fusioniert, die Bindung des Liganden an das Aptamer führt zu einer Konformationsänderung des Ribozyms und ermöglicht so die Selbstspaltung. Die Funktionalität des Konstrukts hängt dabei vor allem von der Linker-Region, auch Kommunikationsmodul genannt, ab. Dieses kann über computergestützte Vorhersagen sowie durch *in vitro* und *in vivo* Versuche ermittelt werden. Unterschiedliche Kommunikationsmodule können eine Selbstspaltung des Ribozyms

entweder fördern oder inhibieren. An- oder Aus-Schalter können so erzeugt werden (Tang & Breaker 1997; Winkler et al. 2004; Soukup & Breaker 1999).

Auch in prokaryotischen Zellen können Aptamer-kontrollierte Ribozyme zum Einsatz kommen. Ein Ribozym wurde mit einem Theophyllin-Aptamer so fusioniert, dass die SD-Sequenz mit einer Anti-SD-Sequenz maskiert wurde. Nach der Zugabe des Liganden und der damit einhergehenden Selbstspaltung des Ribozyms, wurde der die SD-Sequenz beinhaltende, stromabwärtsgelegene Teil der mRNA abgespalten und die Translation konnte stattfinden (Wieland & Hartig 2008). Das TC-Aptamer, welches auch Gegenstand dieser Arbeit ist, wurde verwendet, um in Kombination des Ribozyms aus *S. mansonii* die Genexpression in *S. cerevisiae* zu steuern. Die Insertion von Aptazymen in die 3'UTR von Reportergenen kann zur Degradation der Ziel-mRNA in Abwesenheit oder Anwesenheit des Liganden führen (Abbildung 2.9.A). Für die Studie wurden drei unterschiedliche Bereiche innerhalb der Linker-Region randomisiert, um nach Sequenzen zu suchen, die eine Kommunikation zwischen katalytischer und regulatorischer Domäne ermöglichen. Dazu wurde mit einem Pool von 2,7 x 10^8 verschiedenen RNAs eine *in vitro* Selektion durchgeführt.

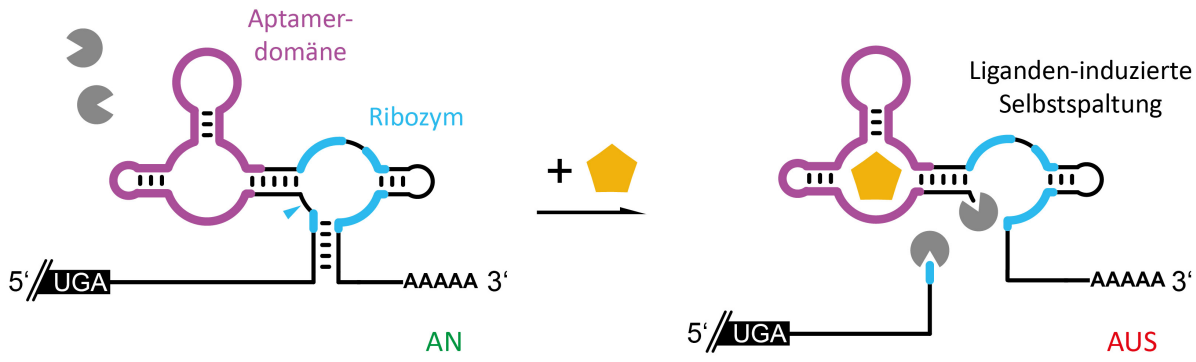


Abbildung 2.9.A Der Hammerhead-Ribozym-Wirkmechanismus. Ein selbstspaltendes Aptazym wird in die 5´ oder 3´ UTR einer eukaryotischen mRNA eingefügt. Die Selbstspaltung wird durch einen Liganden induziert, reduziert die mRNA-Stabilität und löst den RNA-Abbau über Exoribonukleasen aus (grau). Die Abbildung wurde aus der Dissertationsschrift von Dr. Florian Groher übernommen und leicht abgeändert.

Drei Konstrukte, die eine Spaltungsaktivität zeigten, wurden in einem Reportergen-Assay in Hefe untersucht. Dazu wurden sie in das 3´UTR eines GFP-Gens kloniert und mit CAA-Linkersequenzen flankiert, um RNA-RNA-Interkationen weitestgehend zu vermeiden. Die Reportergenexpression konnte mit diesen Konstrukten um 30-60% gesenkt werden (Wittmann & Suess 2011). Diesem auf dem TC-Aptamer basierendem OFF-Schalter, welches in Hefe in Anwesenheit von TC mRNA spaltete, folgte das rationale Design eines in Säugetierzellen funktionierendem ON-Schalters. Da die tertiäre Loop-Loop-Interaktion des Hammerhead-Ribozyms wichtig für die korrekte Faltung des katalytischen Zentrums des Ribozyms ist und ebenso essentiell für die Spaltungsaktivität unter zellulärer

Magnesiumkonzentration, wurde eben diese Interaktion durch die Integration von drei nt unterbrochen (Beilstein et al. 2014). Der Stamm P1 des TC-Aptamers wird durch die 3 nt so geschwächt, dass es in Abwesenheit von TC zu einer Spaltungsreaktion kommt, der PolyA-Schwanz wird abgetrennt und die RNA abgebaut. In Abwesenheit von TC dagegen wird der P1 Stamm stabilisiert, sodass die Spaltung verhindert wird. Eine Steigerung der Regulation wurde durch die Integration von 2 Ribozymen in das 5´UTR bzw. das 3´UTR erreicht. Auch bei einer stabilen Genomintegration konnte durch das System eine Regulation erzielt werden.

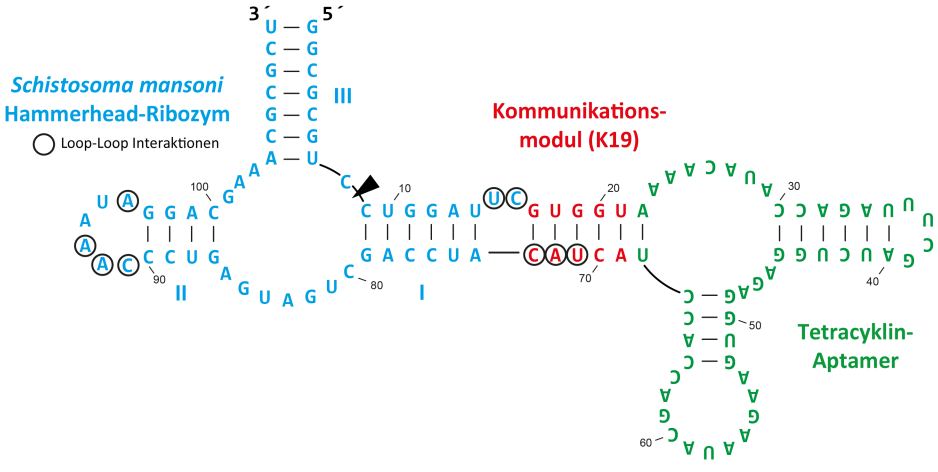


Abbildung 2.9.B Sekundärstruktur des Hammerhead-Ribozyms von Beilstein *et al* (Beilstein et al. 2014). Der ursprüngliche Ring zwischen den Hammerhead-Ribozym und dem TC-Aptamer wurde aufgelöst und in den P1-Stamm des TC-Aptamers integriert, welcher gleichzeitig als Linkerbereich dient. Die grün markierten Basen des P1-Stammes sowie die pink markierten Basen des Hammerhead-Ribozyms, interagieren in der Tertiärstruktur miteinander. Dies führt zu einer erhöhten Spaltungsaktivität. Für den umrandete Teil des P1-Stammes wurden verschiedene Sequenzen entwickelt, welche zu unterschiedlichen Schaltfaktoren führte (Abbildung mit freundlicher Genehmigung von Dr. Florian Groher, abgeändert nach Beilstein *et al.*)

2.10 RNA-Interferenz

In Eukaryoten bietet der RNAi-Weg (RNA-Interfernz) eine weitere Möglichkeit, die Genexpression zu beeinflussen. Die RNA-Interferenz (RNAi) ist ein konservierter biologischer Mechanismus, der in Eukaryonten vorhanden ist und zwei Klassen von kleinen RNAs, die bereits erwähnten, siRNAs und die miRNAs, umfasst. Nach der Prozessierung durch Drosha und Dicer werden die 21 bis 25 nt langen RNAs in den RNA *induced silencing complex* (RISC) überführt, um die Genexpression entweder durch mRNA-Abbau oder durch Translationsrepression zu inhibieren. Es wurden für diesen Stoffwechselweg synthetische Riboswitche entwickelt, welche die Drosha-Spaltung von pri-miRNAs und die Spaltung von exogen applizierten kleinen Haarnadel-RNAs (shRNAs) durch Dicer regulieren.

So wurde das Theophyllin-Aptamer in die terminale Region einer shRNA eingebaut (An et al. 2006). Die Bindung von Theophyllin an das Aptamer kann dann die Erkennung oder Spaltung durch Dicer stören. In Abwesenheit eines Liganden führte die shRNA-Prozessierung zu einer geringen und die Zugabe von Theophyllin zu einer erhöhten Reportergenexpression (Tuleuova et al. 2008).

2.11 RNA-basierte Translationskontrolle in Eukaryonten durch synthetische Riboswitche

In Eukaryoten scannt der 43S-Prä-Initiationskomplexes des Ribosoms nach Bindung an das 5´Cap den gesamten 5´UTR, um schließlich am Startcodon mit der Translation zu beginnen. Sowohl die Bindung des Prä-Initiationskomplexes an die mRNA als auch der Scanning-Mechanismus kann mit der Einbringung einer starken Stammschleife unterbunden werden (Hinnebusch 2011). Bei der Bindung des Liganden in die vorgeformte Bindetasche eines im 5´UTR vorkommenden Aptamers kommt es zu einer Stabilisierung der RNA-Struktur, was eine Störung der Translation der mRNA zu Folge hat. Der 43S-Prä-Initiationskomplex, welcher den 5´UTR der mRNA nach einem Startcodon scannt, wird durch die stabile, Liganden gebundene Konformation des Aptamers blockiert und die Translation wird so inhibiert (Abbildung 2.11).

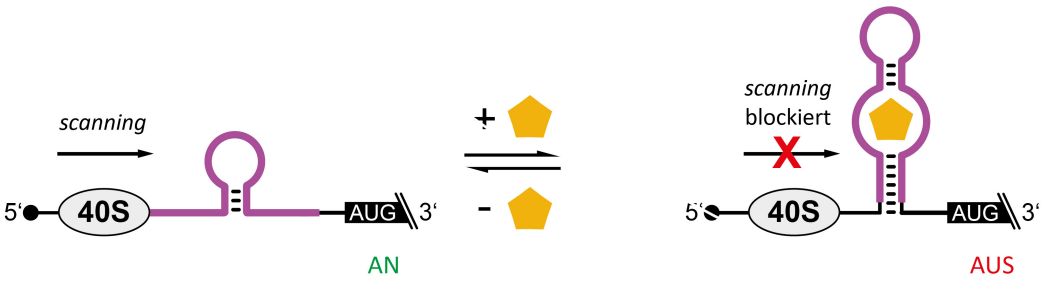


Abbildung 2.11 Wirkmechanismus eines Riboswitches im 5´UTR einer mRNA. Im ungebundenen Zustand kann die Translation stattfinden und wird durch die Stammschleife des Aptamers in ungebundener Form nur leicht inhibiert. Nach Ligandenbindung ist die Sekundärstruktur des Aptamers so stabil, dass der Scanning-Mechanismus des Ribosoms nach dem AUG nicht mehr stattfinden kann. Die Abbildung wurde aus der Dissertationsschrift von Dr. Florian Groher übernommen und leicht abgeändert.

Eine 10-fache Regulation der Genexpression in einer aus Hamster-Ovarien bestehenden Zelllinie (CHO-Zellen) konnte so mit einem in das 5´UTR eines Reportergenes eingebrachten Aptamers gegen den Hoechst-Farbstoff H33258 erzielt werden, wie in einer 1998 veröffentlichen Studie von Werstuck und Green gezeigt wurde (Werstuck & M. R. Green 1998). In weiteren Studien konnte diese Art der Riboswitche mit verschiedenen Aptameren etabliert werden.

2.12 Kontrolle des pre-mRNA-Spleißens

Ein Mechanismus, der Eukaryoten von Prokaryoten unterscheidet und breite Anwendungsmöglichkeiten von synthetischen Riboswitchen ermöglicht, ist die Prozessierung der mRNA durch das Spleißen. Eukaryotische Gene verfügen üblicherweise über ein oder mehrere Introns. Das Spleißen ist hier ein wichtiger, die Diversität des Proteoms erhöhender, Vorgang. Durch die Insertion der 3´ Spleißstelle (SS) eines Introns in das Theophyllin-Aptamer konnte die Menge an gespleißter prä-mRNA reguliert werden. Nach Zugabe des Liganden Theophyllin reduzierte sich die Menge der mRNA, da die Erkennung der 3´SS blockiert wurde. Die daraus folgende Intronretention führte zu einem schnellen mRNA-Abbau (F. Groher & Suess 2014) (Kim et al. 2005). Eine 16-fache Regulation konnte analog zum Theophyllin-Aptamer mit dem TC-Aptamers erreicht werden. Hierfür wurde mit dessen P1-Stamm die 5´SS durch intramolekulare Basenpaarung maskiert. Es wird angenommen, dass so der Zugang des U1-snRNP zur 5´SS blockiert wurde (F. Groher & Suess 2014) (Müller et al. 2006)(Abbildung 2.11.B).

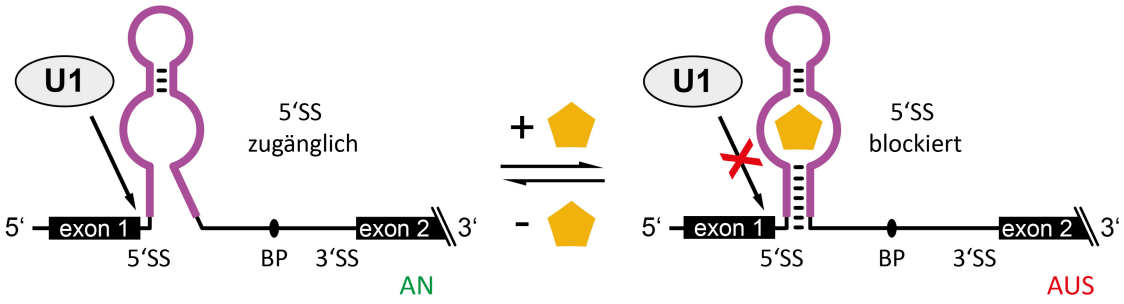


Abbildung 2.11.B Regulierung des prä-mRNA-Spleißens: Ein Aptamer wird in ein Intron einer eukaryotischen mRNA eingesetzt. Die Zugänglichkeit der 5´ ´SS wird nach Ligandenbindung verhindert. Die Abbildung wurde aus der Dissertationsschrift von Dr. Florian Groher übernommen und leicht abgeändert.

2.12.1 Der TC-Riboswitch reguliert die Genexpression auf Ebene der Translation

Durch die Translationsregulation, welche hauptsächlich auf der Ebene der Initiation reguliert, kann die Höhe des Genprodukts in der Zelle rasch verändert werden. Sie stellt einen wichtigen Punkt der post-transkriptionellen Kontrolle der Genexpression dar (Schneider & Suess 2015; Kaufman 1994). Im ungebundenen Zustand bietet die Aptamer-Struktur, die sich im 5´UTR befindet, selbst einen gewissen Widerstand und verschlechtert die Translationsrate. Eine 100%ige Translationsrate (Basalexpression) kann man beim Einbringen eines Aptamers in das 5´UTR nicht mehr beobachten. Meist liegt die Translationsrate des Reportergens im An-Zustand zwischen 80-10%. Im Liganden gebundenen Zustand stabilisiert sich die Struktur des Aptamers weiter und die Translation wird noch

mehr inhibiert. Der Riboswitch befindet sich dann im Aus-Zustand. Den Quotienten aus diesen beiden Werten, dem An- und den Aus-Zustand bezeichnet man als Schaltfaktor.

Bereits 2003 konnte man beobachten, dass das Regulationspotential eines Aptamers im 5´UTR positionsabhängig ist. Das TC-Aptamer wurde in dieser Studie erstmals verwendet, um die Expression eines Reportergens in Hefe zu kontrollieren (Hanson et al. 2003). Dass sich Stammschleifen je nach Position im 5´UTR unterschiedlich auf die Inhibierung der Translation auswirken, wurde bereits 1993 gezeigt (Vega Laso et al. 1993). Ein Minimer des TC-bindende Aptamer (genannt 32sh) (Berens et al. 2001) wurde in den 5´UTR eines konstitutiv exprimierten Glühwürmchen-Luciferase-Gens eingeführt. Das Aptamer wurde entweder nahe dem 5‘ Cap (neun nt dahinter) oder fünf nt vor dem AUG, also nahe dem Startcodon, eingefügt. Die Luziferase-Aktivität wurde für alle Konstrukte in Abwesenheit und Anwesenheit von 250 mM TC gemessen. Für das Cap-distal eingesetzte Konstrukt konnte eine 9-fache Abnahme der Luciferase-Aktivität beobachtet werden, für das Cap-proximal Konstrukt nur eine 3-fache Abnahme. Der Unterschied in der Regulationsfähigkeit ließ sich vor allem auf die Unterschiede in der Basalexpression, die Translation des Proteins ohne Ligand, herleiten. Die Basalexpression des Cap-proximalen Konstrukts war höher als das des AUG-nahen Konstrukts.

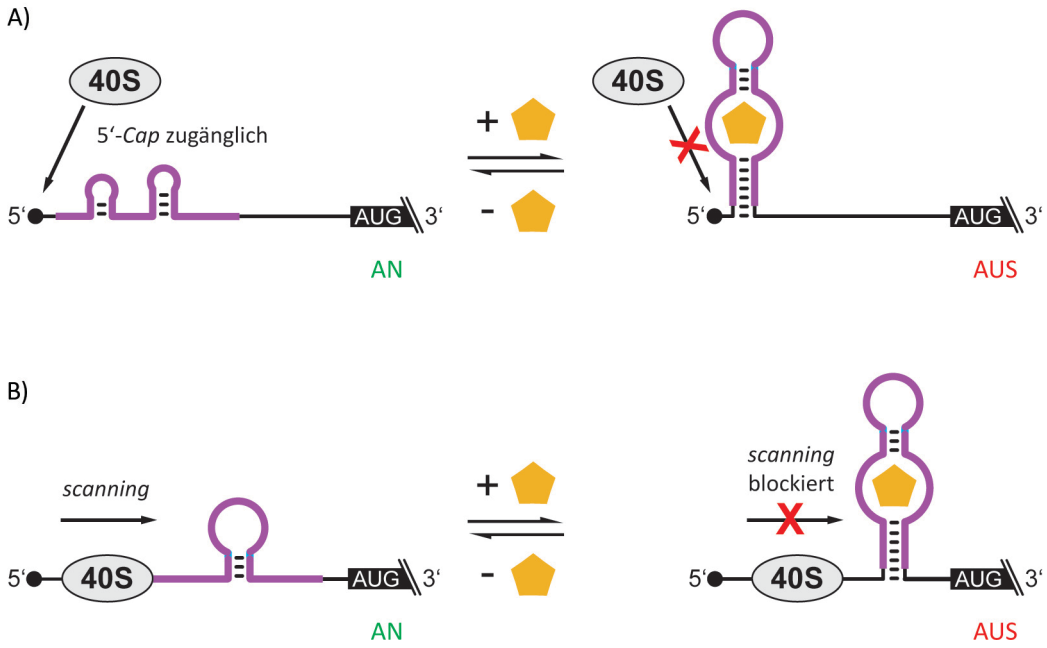


Abbildung 2.11.1 Verschiedene Positionen des Aptamers im 5´UTR. Das Aptamer kann in der Nähe des 5´Caps **A)** oder in der Nähe des Startcodons eingesetzt werden **B)**. Die Abbildung wurde aus der Dissertationsschrift von Dr. Florian Groher übernommen und leicht abgeändert.

Darüber hinaus konnte ebenfalls 2003 nachgewiesen werden, dass eine Veränderung der Sequenz des Stamms P1 des Aptamers zwar nicht die Ligandenbindung beeinflusst, wohl aber einen Einfluss auf die regulatorischen Eigenschaften des Riboswitches hat. Der Stamm wurde durch die Einführung

zweier zusätzlicher Basenpaare stabilisiert, was eine geringere GFP-Expression aber eine erhöhte Regulation zur Folge hatte. Eine Destabilisierung des Stammes erhöhte die GFP-Expression dagegen und verringerte die Regulation (Suess et al. 2003).

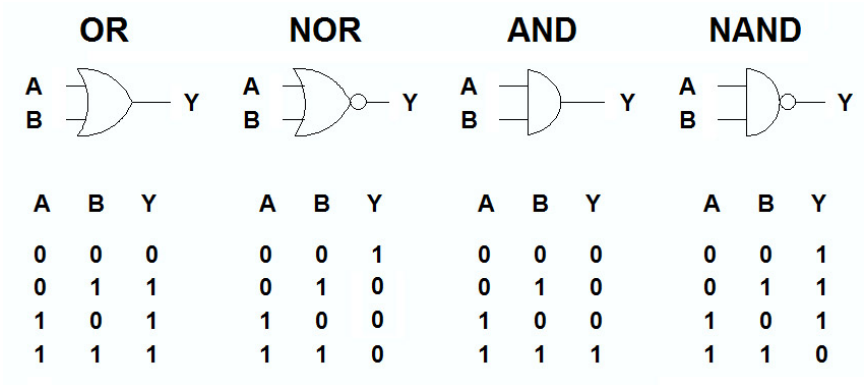
Die Regulationsfähigkeit des TC-Aptamers lässt sich durch das Einfügen mehrerer Kopien in das 5´UTR eines Gens verbessern. Es sind wenige natürliche Riboswitche bekannt, bei welchen zwei Aptamerdomänen eine einzige Expressionsplattform steuern (Mandal et al. 2004; Breaker 2011). In einer 2005 durchgeführten Studie wurde so die Regulation des Riboswitches von 8-fach über 21-fach mit zwei Kopien bis hin zu 37-fach mit 3 Kopien des TC-Aptamers erhöht. Zudem wurden fünf essentielle Gene der Hefe (NEP1, NOP8, NOP14, PGI1 und SEC1) als *proof-of-concept*-Studie für die TC-abhängige Translationsregulation ausgewählt und durch das Einfügen der Aptamere über Insertionskassetten eine Regulation dieser erreicht (Kötter 2009).

2.13 Translationsbasierte logische Gatter

In der Biotechnologie werden Zellen in vielfältiger Weise genutzt, um beispielsweise Verbindungen aus nachwachsenden Rohstoffen herzustellen (Chappell et al. 2015). Zentral ist hier der Entwurf und die Synthese funktioneller, streng regulierter Schaltkreise, welche aus einem Netzwerk regulatorischer Moleküle und ihrer dazugehörigen genetischen Steuerungselemente bestehen. Teilelemente dieser Schaltkreise sind logische Gatter, welche auch in Zellen ihre Anwendung finden können. Logische Gatter sind Teil des Werkzeugkastens, aus dem Schaltkreise entwickelt werden können. Damit logische Gatter in Zellen als Bausteine für synthetische Schaltkreise verwendet werden können, sollten sie so störungsunanfällig wie möglich sein, um nicht homogene Expressionsmuster, das sogenannte Hintergrundrauschen (*noise*), möglichst niedrig zu halten. Genetische Schaltelemente reagieren sehr stark auf unterschiedliche Umweltbedingungen (Wachstumsbedingungen) und sollten daher sehr robust sein (Brophy & Voigt 2014). Die Nutzbarkeit von RNA innerhalb logischer Schaltkreise hat sich als vorteilhaft erwiesen. RNA ist flexibel, modular und vielseitig einsetzbar. Auf der translatorischen Ebene lässt sich sehr effektiv die Protein-Expression steuern. Jedoch muss bei RNA- oder DNA-basierten Steuerungselementen dem genetischen Kontext eine besondere Beachtung geschenkt werden. Mittlerweile gibt es eine Vielzahl an synthetischen RNA-Schaltkreiselementen, darunter Riboschalter und -regulatoren sowie aus CRISBR abgeleitete Elemente (Chappell et al. 2015; McKeague et al. 2016; Berens et al. 2015).

2.13.1 Boolesche Logik

Angelehnt an die logischen Gatter der elektronischen Schaltungen, werden auch die Gatter in der Synthetischen Biologie als boolesche Funktionen implementiert. Abbildung 2.12. zeigt vier verschiedene Schemen Boolescher Logik-Gatter.



A	B	Y
0	0	0
0	1	1
1	0	1
1	1	1

A	B	Y
0	0	1
0	1	0
1	0	0
1	1	0

A	B	Y
0	0	0
0	1	0
1	0	0
1	1	1

A	B	Y
0	0	1
0	1	1
1	0	1
1	1	0

Abbildung 2.12 Boolesche Logik-Gatter. Alle Gatter integrieren zwei Eingangssignale zu einem Ausgangssignal. Bei einem OR-Gatter muss mindestens ein Signal vorhanden sein, um ein Signal im Ausgang zu erzeugen. Dagegen darf bei einem NOR-Gatter kein Signal im Eingang vorhanden sein. Ein Signal im Eingang erzeugt hier kein Signal im Ausgang. Ein AND-Gate braucht zwei gleichzeitige Eingangssignale für ein Ausgangssignal. Ein NAND-Gate erzeugt nur mit zwei Eingangssignalen kein Ausgangssignal. Abbildung übernommen von: https://i.stack.imgur.com/U92yS.png.

Während zum Beispiel ein NOR-Gatter mit einem von zwei Eingangssignalen ausgeschaltet ist, braucht ein AND-Gatter zwei Eingangssignale, um angeschaltet zu sein. Als Beispiel für ein Riboswitch-basiertes NOR-Gatter kann der TC-NEO-Riboswitch in *S. cereviseae* angeführt werden (Schneider & Suess 2015). Die Translation des genetisch integrierten Reportergens kann nur stattfinden, wenn beide Liganden fehlen. Sobald ein Ligand an ein Aptamer im Riboswitch bindet, wird die Translation weitestgehend unterbunden. Ein NAND- und AND-Gatter konnte mit einem Theophyllin-TPP-Riboswitch in Bakterien generiert werden. Durch doppelte genetische Selektion wurden Riboswitche isoliert, welche in der Lage waren, entweder ein AND- oder ein NAND- Gatter zu bilden (Sharma et al. 2008). Durch den Einsatz von Aptazymen konnte die Modularität von RNA-Schaltern in *S. cereviseae* weiter erhöht werden, denn diese können für verschiedene RNA-Klassen eingesetzt werden und sind nicht nur auf mRNA beschränkt (Klauser et al. 2012). So gelang es AND-, NOR und ANDNOT- Gatter zu entwerfen, welche auf Theophyllin und TPP reagieren.

2.14 *Machine learning* und *deep learning* in der Biologie

Angesichts der heutzutage zum Teil riesigen Datenmengen, die innerhalb der Biologie generiert werden, sind *machine learning* und *deep learning* nützliche Werkzeuge, um diese Daten auszuwerten und nutzbar zu machen. Sie bieten neue Wege der Analyse, des Verständnisses und der Optimierung. Mittlerweile gibt es viele Beispiele für eine erfolgreiche Anwendung von *machine learning* innerhalb der Biologie, zum Beispiel Genom-Annotierung (Yip et al. 2013) oder die Vorhersage der Sequenzspezifitäten von DNA- und RNA-bindenden Proteinen (Alipanahi et al. 2015). Ein Ziel des *machine learnings* ist es, auf Basis der verarbeiteten Daten und eines zugrundeliegenden Algorithmusses ein statistisches Modell zur Vorhersage zu entwickeln (Camacho et al. 2018).

Der *machine learning* Algorithmus wird zunächst mit Daten „gefüttert". Bei diesen Daten handelt es sich zum einen um das Ergebnis von Messungen, die auch als *features* (Merkmale) bezeichnet werden. Auch *labels* (Kennzeichnungen) gehören zu den Eingabedaten des Lernalgorithmus. Sie ergeben sich aus den Daten und sind auch gleichzeitig das, was das Modell vorhersagen soll, also die Ausgabe des statistischen Modells. Auf Basis der eingegebenen Daten wird ein Modell trainiert und schließlich eine Vorhersage getroffen. Die verschiedenen Methoden des *machine learnings* werden in zwei Klassen eingeteilt: das überwachte und das unüberwachte Lernen. Für das unüberwachte Lernen braucht man lediglich ein Ausgangsdatenset, auf Basis dessen der Algorithmus klassifiziert. Für das überwachte Lernen braucht man dagegen zwei Datensätze, ein Eingangs- und ein Ausgangsdatenset. Die Basis des Lernens beschreibt den Prozess, den optimalen Satz an Parametern zu finden (Camacho et al. 2018) (A.-C. Groher et al. 2018). Der *random forest* ist ein Algorithmus auf Basis des überwachten Lernens. In diesem Algorithmus werden sogenannte Entscheidungsbäume (*decision trees*) zufällig erstellt, diese *decision trees* bestehen jeweils aus mehreren Verzweigungen, die entstehen, indem die Daten auf Grund der Eigenschaften und auf Basis bestimmter Regeln klassifiziert werden. Über die Verzweigungen sind die Knoten der einzelnen Bäume mit den nächsten Knoten verbunden. An jedem Knoten des Baums wird eine binäre Entscheidung getroffen, welcher Knoten als nächstes angesteuert wird. Das wird so lange wiederholt, bis das Ende des Baums erreicht ist und ein Stoppkriterium die Erstellung des Baums beendet.

Deep learning ist ein weiteres Teilgebiet des *machine learnings* und verwendet künstliche neuronale Netzwerke als Basis des Lernens. Ein Beispiel dafür ist der CNN (*Convolutional Neural Network*), ein „faltendes neuronales Netzwerk". Es besteht, wie andere neuronale Netzwerke, aus mehreren Schichten, wobei mindestens die erste Schicht eine Faltungsschicht ist. Diese Schicht kann Merkmale aus den Eingangsdaten erkennen und Muster extrahieren. Die Muster werden dann in der nächsten Faltungsschicht gespeichert und weiterverarbeitet. Die Faltungsschichten ermöglichen es, die Daten aus verschiedenen Perspektiven (Filtern) zu untersuchen. In den weiteren Ebenen lernt der CNN aus diesen Mustern. Dazu werden die Daten in den Ebenen immer wieder neu untersucht und gefiltert.

Da ein CNN ausgezeichnet Muster erkennen und darauf basierend lernen kann, wird diese Art des künstlichen neuronalen Netzwerks häufig in der Bildverarbeitung eingesetzt. CNNs werden in der Biologie zum Beispiel für das Auswerten und das Lernen auf Basis von DNA-Sequenzen oder medizinischer und mikroskopischer Bilder verwendet (Ching et al. 2018).

2.15 Zielsetzung

Die Generierung synthetischer Aptamere gelingt durch neue Methoden inzwischen immer besser. Ein Nadelöhr stellt aber nach wie vor die Anwendung dieser Aptamere als funktionelle Riboswitche dar, welche die Genexpression kontrollieren und als logische Gatter eingesetzt werden können. Neue Riboswitche sind ihrer anfänglichen Schaltleistung oft schwach und es bedarf zeit- und kostenintensiver Optimierungsstrategien, um diese zu verbessern.

Ziel dieser Arbeit ist es daher zum einen, durch die neuen *in silico*-Anwendungen des *machine learnings* einen einträglichen Weg zur Optimierung von Riboswitchen zu generieren und zum anderen, durch die Ergebnisse des *machine learning*-Prozesses weitere Erkenntnisse über die Funktionsweise von Riboswitchen, vor allem hinsichtlich ihrer biophysikalischen Parameter und ihrer Sequenz, zu gewinnen. Im letzten Abschnitt der Arbeit soll auf Basis dieser Ergebnisse ein Hybrid-Riboswitch aus zwei verschiedenen Aptameren zu einem funktionalen logischen Gatter zusammengesetzt werden.

3 Ergebnisse

Im folgenden Kapitel wird zunächst das Design der *machine learning* Runden beschrieben und anschließend die Messergebnisse (Basalexpressionen und Schaltfaktoren) sowie biophysikalische Parameter und die Sequenzanalysen aufgearbeitet. Anschließend folgt die Analyse eines sehr guten Riboswitches (R4-G8) im Hinblick auf seine biophysikalischen Parameter und seine Sequenz. Im letzten Teil dieses Kapitels wird die Konstruktion eines Hybrid-Riboswitch beschrieben, der aus einem TC-Aptamer und einem Tobramycin-Aptamer besteht und als NOR-Gate funktioniert.

3.1 *Machine learning* und *deep learning* mit dem TC-Dimer – Runden-Design

3.1.1 Ausarbeitung von Design-Parametern für das *machine learning* auf Basis des TC-Dimers

Ein Ziel dieser Arbeit war die Entwicklung einer Methode zur Verbesserung der Riboswitch-Regulation durch *machine learning*. Zu diesem Zweck wurde ein TC-Riboswitch ausgewählt, der aus zwei TC-Aptameren besteht (TC-Dimer), welche in den 5' UTR eines GFP-Reportergens eingefügt wurden (vgl. Abbildung 3.1.1). Das TC-Dimer wurde von Lara Gorini im Rahmen einer Bachelorarbeit designt. Das 5´ TC-Aptamer ist 41 nt nach dem Transkriptionsstartpunkt platziert und das 3´ TC-Aptamer endet fünf nt vor dem AUG, direkt an der Kozak-Sequenz. Die beiden Aptamere sind durch einen 27 nt langen CAA-Spacer getrennt, welcher am 5´ Ende ein zusätzliches Uracil enthält (Abbildung 3.1.1). Das Uracil ist Teil einer AgeI-Schnittstelle, welche den Austausch des Aptamers ermöglicht. Der CAA-Spacer ergibt ein unstrukturiertes Isolationsmodul, das es den beiden Aptameren ermöglichen soll, sich unabhängig voneinander richtig zu falten und den Liganden zu binden. Die Aptamere sind weitestgehend identisch und unterscheiden sich nur in der Sequenz ihrer P1-Stämme (Sequenz siehe Kapitel 5.3.3).

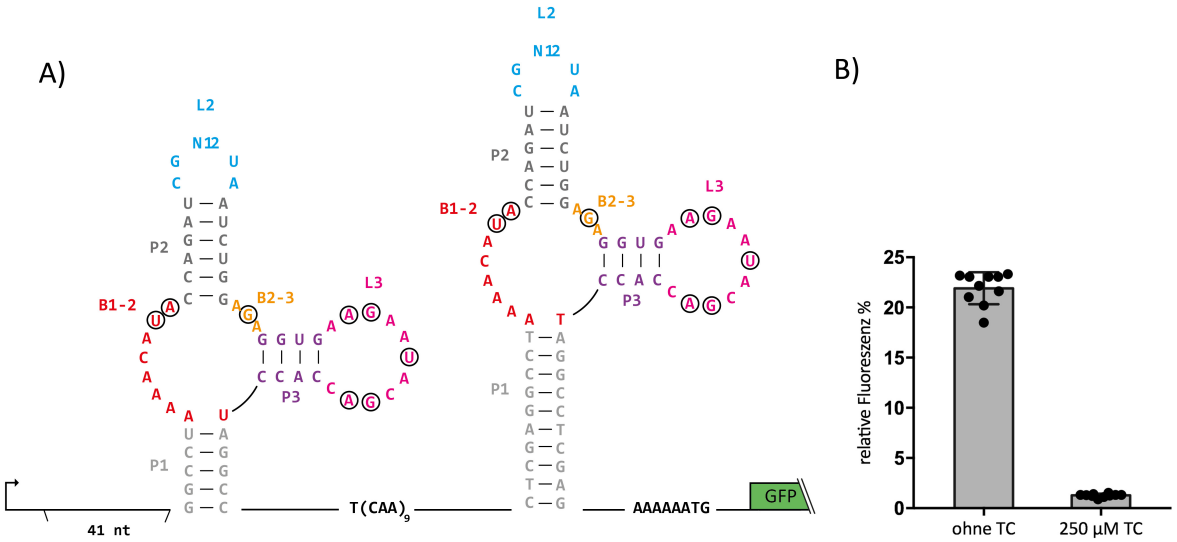


Abbildung 3.1.1. **5´UTR mit TC-Dimer LG3**. **A)** Die beiden Aptamere sind als 2D-Vorhersage dargestellt. Die Bindung des Liganden findet mit den Bereichen B1-2 (rot), B2-3 (gelb), P3 (lila) und L3 (rosa) statt. Für die Bindung wichtige Nukleotide sind schwarz umrandet. Die Stämme P1 der Aptamere weisen eine unterschiedliche Länge und Sequenz auf und bedingen so ein unterschiedliches Expressionsmuster und einen unterschiedlichen Schaltfaktor. Der Stamm des 5´Cap proximalen Aptamers (5´) besteht aus 5 Basenpaare (bp), der Stamm des zweiten Aptamers (3´) umfasst 10 bp. **B)** Vergleich der relativen Fluoreszenz von LG3 ± 250 µm TC (Mittelwert ± SD, n=10).

Frühere Studien zeigten, dass die Sequenz von P1 einen signifikanten Einfluss auf den Schaltfaktor hat (Suess et al. 2003). Das Ziel war es, die optimale Stammkomposition zu finden, die zu einem hohen Schaltfaktor mit einem hohen initialen Expressionsniveau (Basalexpression) führt. Bei dem Ausgangs- und Vergleichskonstrukt, genannt LG3, handelt es sich um einen sehr guten Riboswitch, der im 5´UTR eines Reportergenes einen Schaltfaktor von durchschnittlich 18-fach mit einer Basalexpression von 22% erreicht. Dieser Schalter sollte im Hinblick auf sein basales Expressionsmuster und seine Schaltfähigkeit verbessert werden. Dies sollte alleine durch die Veränderung eines der beiden P1-Stämme erreicht werden. Der P1-Stamm des 3' Aptamers wurde für die Mutationen ausgewählt, weil bereits bekannt war, dass Aptamere, wenn sie näher am AUG platziert werden, einen stärkeren Einfluss auf die Regulation haben (Suess et al. 2003). Die restliche Sequenz des Dimers sollte unverändert bleiben.

3.1.2 Generierung von Datenpunkten für das *machine learning* und Festlegung von biophysikalischen Parametern – Runde 1

Für ein *machine learning*-Programm werden zunächst Datenpunkte benötigt, mit welchen das Programm lernen kann. Daher sollte eine Messung von 96 Tandem-Konstrukten, die sich im 3´P1-Stamm unterscheidenden, als Ausgangsdatenset für das anschließende *machine learning* dienen. Um

gut funktionierende Riboschalter zu erhalten, wurden neben der Stammlänge des zweiten Aptamers weitere Kriterien festgelegt. In früheren Experimenten konnte gezeigt werden, dass ein ΔG zwischen -20 und -30 kcal/mol zu einer Basalexpression im Bereich zwischen 20% bis 70% führt (Dissertation Christopher Schneider). Daher wurde dieser ΔG als angestrebter Parameter für die zukünftigen Riboswitche ausgewählt.

Um eine Reduktion des Lösungsraumes für die Tandem-Riboswitche zu erreichen, wurde nur eines der beiden Aptamere verändert und das 5' Aptamer mit einer Stammlänge von fünf nt konstant gehalten. Der P1-Stamm des 3'-Aptamers wurde zufällig auf Stammlängen zwischen sechs und zehn nt geändert. Wie bereits erwähnt, hat eine Veränderung der Stammlänge und Sequenz des P1-Stamms keinen Einfluss auf die Liganden-Bindungseigenschaften des Aptamers.

Damit ein initialer Datensatz für das *machine learning*-Programm erstellt werden konnte, wurden in einer ersten Runde die Sequenzen der Stämme nach dem Zufallsprinzip generiert. Dieser Randomisierungsalgorithmus (sowie das später verwendete Programm zum *machine learning*) wurden von Dr. Sven Jager (AG Hamacher) entwickelt. Dabei wurden die zuvor genannten Kriterien für das 3´ Aptamer berücksichtigt. Die nt-Positionen des P1-Stamms wurden für die Länge $L_i \in \{6,7,8,9,10\}$ randomisiert. Um festzustellen, ob eine "Fehlfaltung" des Dimers (z.B. durch den Sequenz-Kontext) auftritt, wurde die Sekundärstruktur für jede Sequenz vorhergesagt (der randomisierte Stamm zuzüglich der Sequenzen 40 nt stromaufwärts und 150 nt stromabwärts des Aptamers). Weiterhin wurde getestet, ob der ΔG der Aptamere in dem Bereich zwischen -20 und -30 kcal/mol liegt. Um eine hohe Vielfalt in der Basenpaarzusammensetzung des Stammes zu erreichen, wurde der Levenshtein-Abstand (*Levenshtein-Distance:* LD) von den Stämmen untereinander berechnet und dann geclustert. Dazu wurden die vorhergesagten Stämme in möglichst unterschiedliche Gruppen eingeteilt, um so die Sequenz-Diversität zu erhöhen. Die einzelnen Sequenzen wurden dann aus diesen Gruppen zufällig und gleichverteilt ausgelost. Der LD ist definiert als die kleinste Anzahl von Einfügungen, Streichungen und Substitutionen, die erforderlich sind, um eine Zeichenkette (hier Sequenz) in eine andere umzuwandeln.

Auf Basis der festgelegten Kriterien wurde das erste Set an Stämme vorhergesagt. Von ursprünglich 96 geplanten Konstrukten konnten 79 kloniert und gemessen werden. Die Fluoreszenzaktivität dieser Konstrukte wurde mit und ohne Zugabe von 250 µM TC im Medium und einer Inkubationszeit von 24 h getestet. Die Klonierungen und Messungen erfolgten bis einschließlich zur zweiten Runde durch Christopher Schneider (Dissertationsschrift Christopher Schneider). Das Schaltverhalten der Konstrukte wird durch ihren Schaltfaktor (Schaltfaktor/x-fach, Quotient aus An-Zustand (hier: ohne Ligand) und Aus-Zustand (hier: mit Ligand) und die Basalexpression (Basalexpression, %, Fluoreszenz des An-Zustandes im Verhältnis zur Fluoreszenz eines Konstrukts ohne Aptamer-Insertion) beschrieben.

3.1.3 Die Vorhersage mit dem *random forest* - 2. und 3. Runde

Die 2. Runde wurde mit dem *random forest* auf der Grundlage der Daten aus den ersten Messungen vorhergesagt. Für die 2. und 3. Runde wurden die Riboswitche in vier Klassen eingeteilt: Klasse eins -> hohe Basalexpression und niedriger Schaltfaktor; Klasse zwei: hohe Basalexpression und hoher Schaltfaktor; Klasse drei -> niedrige Basalexpression und niedriger Schaltfaktor und Klasse vier; niedrige Basalexpression und hoher Schaltfaktor (Abbildung im Anhang 8.5.A). Es wurde angestrebt, Schalter der Klasse zwei vorherzusagen (hohe Basalexpression und hoher Schaltfaktor).

Als Klassifikator wurde ein *random forest* gewählt, weil der Prozess zur Erstellung der einzelnen *trees* (Entscheidungsbäume) effizient ist, er leicht zu implementieren ist, die *trees* sehr komplexe Beziehungen abbilden können und auch ohne spezielle Vorverarbeitung mit verschiedenen Daten umgehen können (Louppe 2015; Biau 2012). Abbildung 3.1.3 zeigt die schematische Darstellung eines *random forest*.

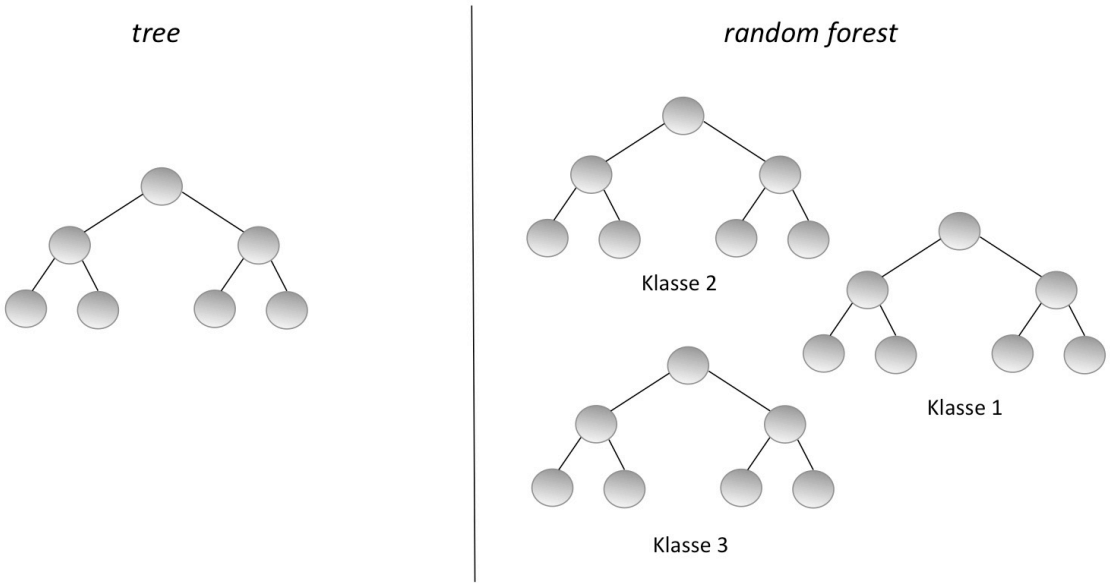


Abbildung 3.1.3 Schema eines einzelnen Entscheidungsbaums und einem *random forest*. Der *random forest* besteht aus vielen *trees*, welche alle unterschiedlich trainiert werden. Für die endgültige Klassifizierung werden jedoch alle berücksichtig und die Klasse mit den meisten Stimmen entscheidet über die endgültige Klassifizierung. Bild wurde Abgeändert nach https://miro.medium.com/max/700/0*YEwFetXQGPB8aDFV.

Die Daten aus der 1. Runde wurden verwendet, um diesen *random forest* mit biophysikalischen Parametern zu trainieren, die aus der Sequenz berechnet werden können: die minimale freie Energie (MFE oder ΔG) des 3´Aptamers, die Länge des P1-Stamms und dessen GC-Gehalt, die Entropie (Shannon) des Stammes sowie die freie Basenpaare des 3´Aptamers und an der Wasserstoffbrückenbindung beteiligte Basenpaare (H-Bindung), die Schmelztemperatur von P1 (Tm

P1) und die Schmelztemperatur des kompletten Schalters (Tm). Im weiteren Verlauf dieser Arbeit werden die Parameter ΔG sowie der Tm des Stammes, die Entropie und der GC-Gehalt tiefer analysiert. Im *random forest* wachsen alle *trees* während des Lernprozesses zufällig. Für eine Klassifizierung trifft jeder *tree* in diesem *random forest* eine Entscheidung für eine Klasse und die Klasse mit den meisten Stimmen entscheidet über die endgültige Klassifizierung. Die Vorhersage selbst läuft über einen Vektor, der die biophysikalischen Merkmale als Einträge speichert. Für jede neu erzeugte Sequenz wird dieser Vektor berechnet und der *random forest* kann im Folgenden vorhersagen, ob die Sequenz zu einem hohen oder niedrigen Schaltfaktor führen wird (A.-C. Groher et al. 2018). Schließlich wurde die 3. Runde auf Basis der Datenpunkte der ersten beiden Runden vorhergesagt.

3.1.4 *Deep learning*, Implementierung der Sequenz und weitere Anpassungen der 4. Runde

Nach der 3. Runde wurde eine Testrunde durchgeführt. In dieser Runde wurden Konstrukte mit niedrigen Schaltfaktoren und Konstrukte mit hohen Schaltfaktoren verglichen und analysiert. Auf Grund dieser Analyse und mit Hilfe von rationalem Design wurden zehn neue Riboswitche konstruiert. Diese Riboswitche zeigten eine überdurchschnittlich gute Leistung und ähnelten sich in ihrer Sequenz.

Um die Sequenz in die Vorhersage zu integrieren, wurde als zweiter Klassifikator ein *convolutional neural network* (CNN) in Kombination mit einem mehrschichtigen neuronalen Netz (Perzeptron) eingeführt. Dieser CNN sollte mit einer Kombination von Informationen aus der Sequenz und Sekundärstrukturvorhersage lernen und sollte den Zufallswald nicht ersetzten, sondern als zweiten Schritt ergänzen. Dazu wurde die Sekundärstruktur des Aptamers in Punkt-Klammer-Annotation in ungerichtete Graphen umgewandelt. Für einen *motif mining*-Ansatz wurde jeder dieser Graphen in alle möglichen Untergraphen mit k-Knoten in einem Bereich von 3-7 unterteilt (Abbildung 3.1.4.A) (Gawronski & Turcotte 2014). Vorteilhaft an dieser Methode ist, dass auch Sequenzmotive auftreten können, die nur durch Sekundärstrukturen miteinander verbunden sind. Die Auftrittswahrscheinlichkeit dieser Motive wurde binär kodiert und anschließend als Merkmalsvektor genommen.

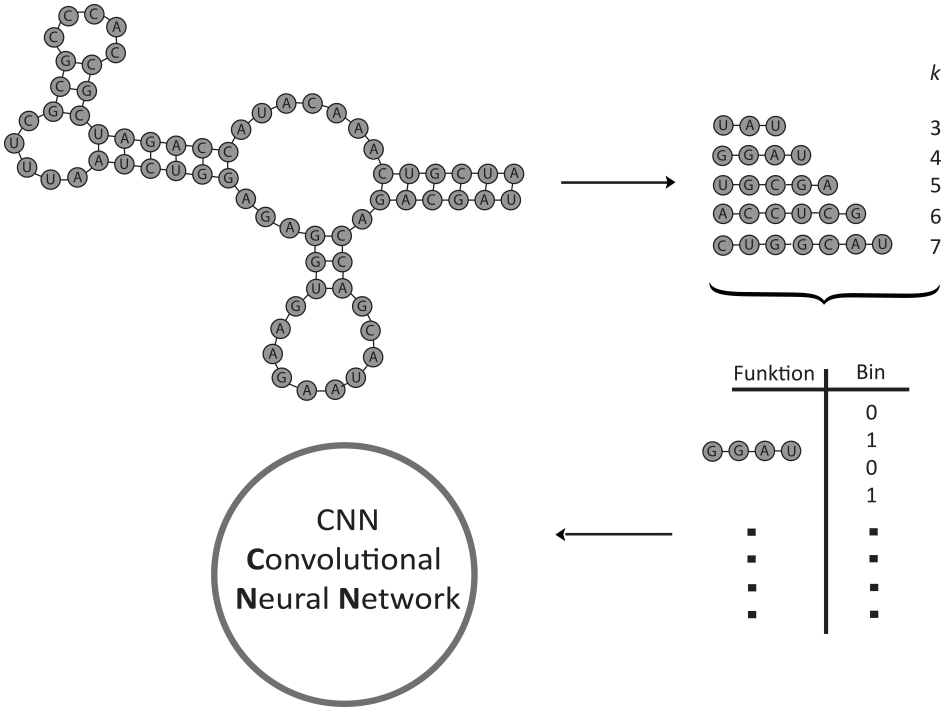


Abbildung 3.1.4.A: Workflow für die Erstellung von Subgraph-Features für den CNN durch vorherige Einteilung des Graphen in Sequenzmotive. Abgeändert nach (A.-C. Groher et al. 2018).

Eine weitere Anpassung war das Verkleinern des Lösungsraums. Die bisherigen vorhergesagten Stämme umfassten eine Stammlänge von sechs bis zehn nt. Die Möglichkeiten für verschiedene Stammkompositionen sind bei dieser Anzahl an nt sehr groß und kann durch eine Stammlänge von acht nt erheblich eingegrenzt werden. Die Wahl fiel auf eine Stammlänge von acht nt, weil bei dieser die meisten Riboswitche mit einem hohen Schaltfaktor gefunden werden konnten (Daten in den jeweiligen Kapiteln der Runden). Zusätzlich wurden die Schalter zur Vorhersage der 4. Runde nicht mehr in vier Klassen eingeteilt, sondern nur noch in zwei: Schalter mit niedrigem und Schalter mit hohem Schaltfaktor (Vergleich Abbildung 8.5.A, Anhang). Hierbei wurde versucht, Schalter mit hohem Schaltfaktor vorherzusagen.

Zudem folgte eine Anpassung des Hyperparameters, den man für die Steuerung des Trainingsalgorithmus verwendet. Für die ersten drei Runden war dieser Wert auf einen Schaltfaktor von 12,5 festgelegt worden. Für die 4. Runde sollte dieser Wert neu bestimmt werden. Um zu überprüfen, an welchem Schwellenwert die Riboswitche in die verschiedenen oben beschriebenen Klassen eingeteilt werden, wurde der Schwellenwert mit Intervallen von 0,01 variiert und anschließend der AUC (***A**rea **u**nder the ROC-**c**urve*) des trainierten Modelles bestimmt. Ziel war des den AUC zu maximieren, um den Punkt zu finden, an dem das Modell die größte Trennschärfe zwischen den Klassen hat. Der AUC misst den gesamten zweidimensionale Bereich unter einer ROC-Kurve (R*eceiver Operating Characteristic*), welche selbst zwei Parameter angibt: Echte positive Rate

und falsche positive Rate bei verschiedenen Klassifizierungsschwellen. Wird der Klassifizierungsschwellenwert verringert, werden mehr Elemente als positiv klassifiziert und sowohl die echte positive Rate als auch die falsche positive Rate wird erhöht (Abbildung 8.5.B im Anhang). Die Ergebnisse dieser Berechnungen sind in Abbildung 3.1.4.B dargestellt. Der ursprüngliche Hyperparameter lag bei einem Schaltfaktor von 12,5-fach. Hier zeigt sich, dass die Vorhersagekraft für diesen Schaltfaktor vergleichsweise niedrig ist. Für die 4. Runde wurde der Hyperparameter daher auf 14,6-fach festgelegt, denn hier ist die Vorhersagekraft recht hoch, die Streuung dieser gering und es liegen immer noch genug Datenpunkte für die Vorhersage vor.

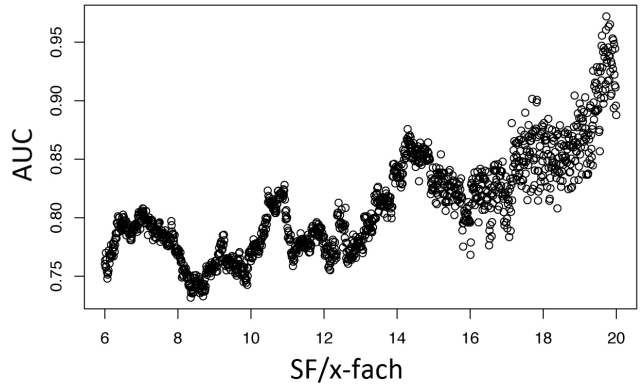


Abbildung 3.1.4.B Maßstab des Hyperparameters. Der AUC wird gegen den Hyperparameter aufgetragen. Abbildung übernommen und leicht abgeändert aus (A.-C. Groher et al. 2018).

Für die Runde 4 wurden die Stämme zuerst durch *random forest* und anschließend mit dem CNN klassifiziert. Schließlich wurden mit Hilfe des *k-means* und der Levenshtein-Distanz 100 Cluster gebildet und aus jedem dieser Cluster wurde eine Sequenz zur Klonierung ausgelost (A.-C. Groher et al. 2018).

3.2 Ergebnisse der randomisierten Runde 1

In der 1. (randomisierten) Runde wurden 79 Konstrukte kloniert und gemessen. Die Konstrukte unterschieden sich nur in ihrem 3´P1-Stamm. Die Stammlänge der unterschiedlichen P1-Stämme der getesteten Konstrukte lag zwischen sechs und zehn nt. Abbildung 3.2.A zeigt die Basalexpression in Bezug auf den Schaltfaktor dieser Runde. Die Randomisierung des Stammes P1 führte zu einer breiten Verteilung des Schaltfaktor mit einem Mittelwert von 7,9-fach. In dieser Runde wurde bereits ein Riboschalter gefunden, welcher in einem ähnlichen Bereich schaltet wie das Ausgangskonstrukt LG3. Dieser Riboschalter schaltet 18,6-fach und hat eine Basalexpression von etwa 25%. Drei Konstrukte haben eine sehr geringe Basalexpression und einen sehr geringen Schaltfaktor.

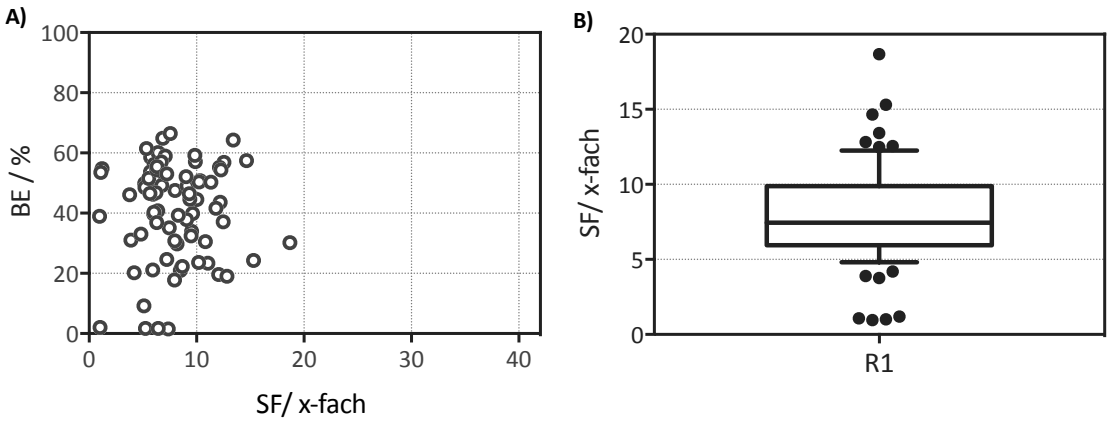


Abbildung 3.2.A Ergebnisse der Messungen aus der randomisierten Runde, Basalexpression und Schaltfaktor A) Basalexpression in % in Bezug auf den Schaltfaktor (SF), jeder Punkt zeigt den Mittelwert aus 2 unabhängigen Messungen. **B)** Boxplot der Schaltfaktoren der Konstrukte von Runde 1. Whisker stellen P10 und P90 dar.

Abbildung 3.2.B zeigt, in welchem Bereich sich der Tm des P1-Stammes in der 1. Runde befindet und welcher Schaltfaktor welchem Tm zuzuordnen ist. Die meisten Stämme weisen einen Tm zwischen 15°C und 50°C auf. Die P1-Stämme der drei Konstrukte mit den höchsten Schaltfaktoren befinden sich in einem Bereich zwischen 40°C und 50°C. Die Konstrukte, welche hier den niedrigsten Tm des P1-Stammes aufweisen, haben auch niedrige Schaltfaktoren. Vier Konstrukte mit sehr geringen Schaltfaktoren haben allerdings einen Tm zwischen 15°C und 40°C.

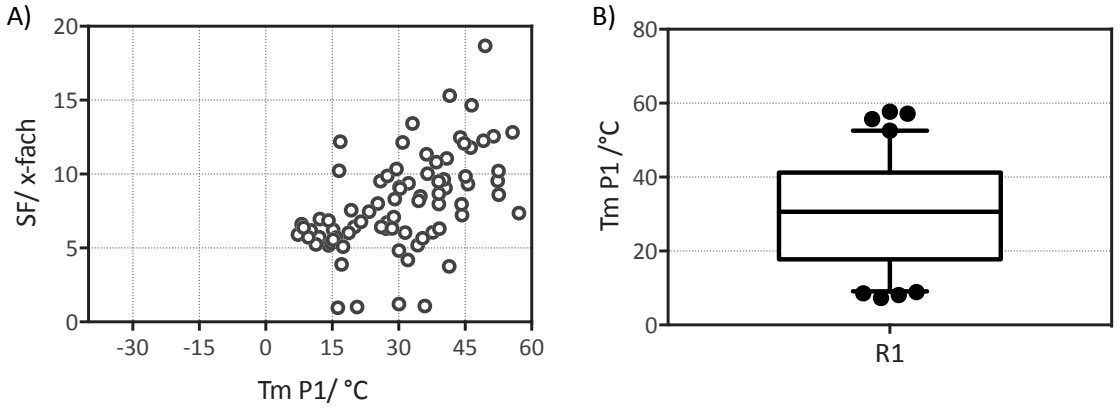


Abbildung 3.2.B Ergebnisse der Messungen aus der randomisierten Runde, Schaltfaktor und Tm des P1-Stamms in °C A) Tm in °C in Bezug auf den Schaltfaktor (SF), jeder Punkt zeigt den Mittelwert aus 2 unabhängigen Messungen. **B)** Boxplot des Tm P1 der Konstrukte von Runde 1. Whisker stellen P5 und P95 dar.

Abbildung 3.2.C zeigt die Verteilung des biophysikalischen Parameters ΔG. Dieser liegt in einem Bereich zwischen -30 und -17 kcal/mol. Riboswitche, die einen höheren Schaltfaktor haben, liegen im Bereich zwischen -27 kcal/mol und -23 kcal/mol. Vier Konstrukte mit einem niedrigen Schaltfaktor

liegen in einem Bereich zwischen -24 und - 18 kcal/mol. Ein Riboswitch, der in einem ΔG Bereich zwischen -30 kcal/mol und -25 kcal/mol liegt, hat hier keinen Schaltfaktor der schlechter als 6-fach ist.

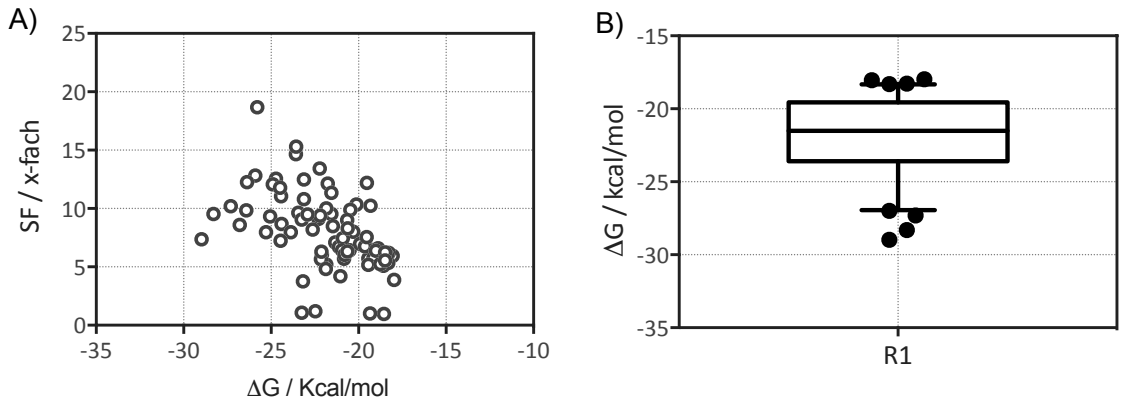


Abbildung 3.2.C Ergebnisse der Messungen aus der randomisierten Runde, Schaltfaktor und ΔG in kcal/mol A) ΔG in kcal/mol in Bezug auf den Schaltfaktor (SF), jeder Punkt zeigt den Mittelwert aus 2 unabhängigen Messungen. **B**) Boxplot des ΔG-Wertes der Konstrukte von Runde 1. Whisker stellen P5 und P95 dar.

Abbildung 3.2.D zeigt die Verteilung des GC-Gehalts von P1 in Prozent und die Verteilung in Bezug auf den Schaltfaktor. Der GC-Gehalt ist breit verteilt und befindet sich in einem Bereich zwischen 20% und 80%. Auch hier sind die vier niedrig schaltenden Konstrukte zu finden, zwei der vier Konstrukte liegen übereinander mit einem Schaltfaktor von 0,96 und 1,01 und einem GC-Gehalt von 43%. Die zwei anderen Konstrukte haben einen GC-Gehalt von 30% bzw. 57%. Zwischen einem GC-Gehalt des P1-Stamms von 57% und 75% liegen hier die Riboswitche mit den höchsten Schaltfaktoren.

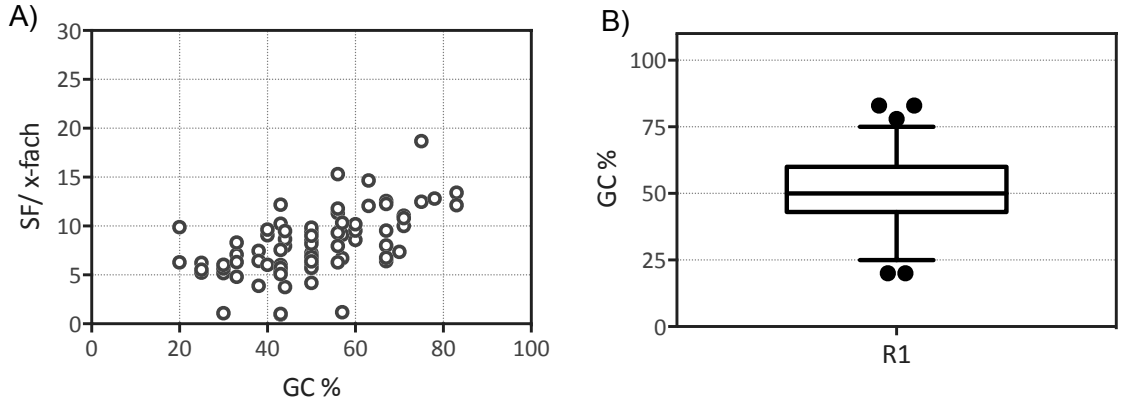


Abbildung 3.2.D Ergebnisse der Messungen aus der randomisierten Runde, Schaltfaktor und GC-Gehalt in % A) GC-Gehalt von P1 in % in Bezug auf den Schaltfaktor (SF), jeder Punkt zeigt den Mittelwert aus 2 unabhängigen Messungen. **B** Boxplot des GC-Gehalts von P1 in % von Runde 1. Whisker stellen P5 und P95 dar.

Abbildung 3.2.E zeigt die Verteilung der Entropie in bit des P1-Stamms und in Bezug auf den Schaltfaktor. Die Entropie beschreibt die Sequenzvielfalt des Stamms. Die Entropie liegt hier zwischen 0,6 und 1,4. Die vier nicht schaltenden Konstrukte liegen breit verteilt, eine Gemeinsamkeit lässt sich nicht erkennen. Eine höhere Entropie ab 0,8 zeigt höhere Schaltfaktoren, jedoch befinden sich in diesem Bereich auch Konstrukte mit einem niedrigen Schaltfaktor.

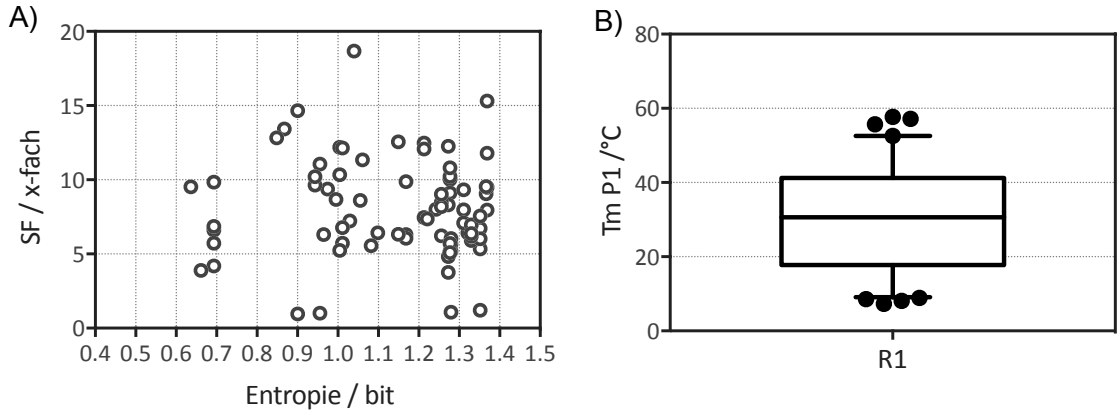


Abbildung 3.2.F Ergebnisse der Messungen aus der randomisierten Runde, Schaltfaktor und Entropie in bit A) Entropie in bit in Bezug auf den Schaltfaktor (SF), jeder Punkt zeigt den Mittelwert aus 2 unabhängigen Messungen. **B** Boxplot der Entropien in bit von Runde 1. Whisker stellen P5 und P95 dar.

Abbildung 3.2.G zeigt ein Histogramm der Stammlänge in Bezug auf die Häufigkeit der jeweiligen Schaltfaktoren. Stämme mit der Stammlänge sechs nt waren in Runde 1 weniger häufig vorhanden als andere Stammlängen. Die höchsten Schaltfaktoren wurden mit Stämmen der Länge acht nt erreicht.

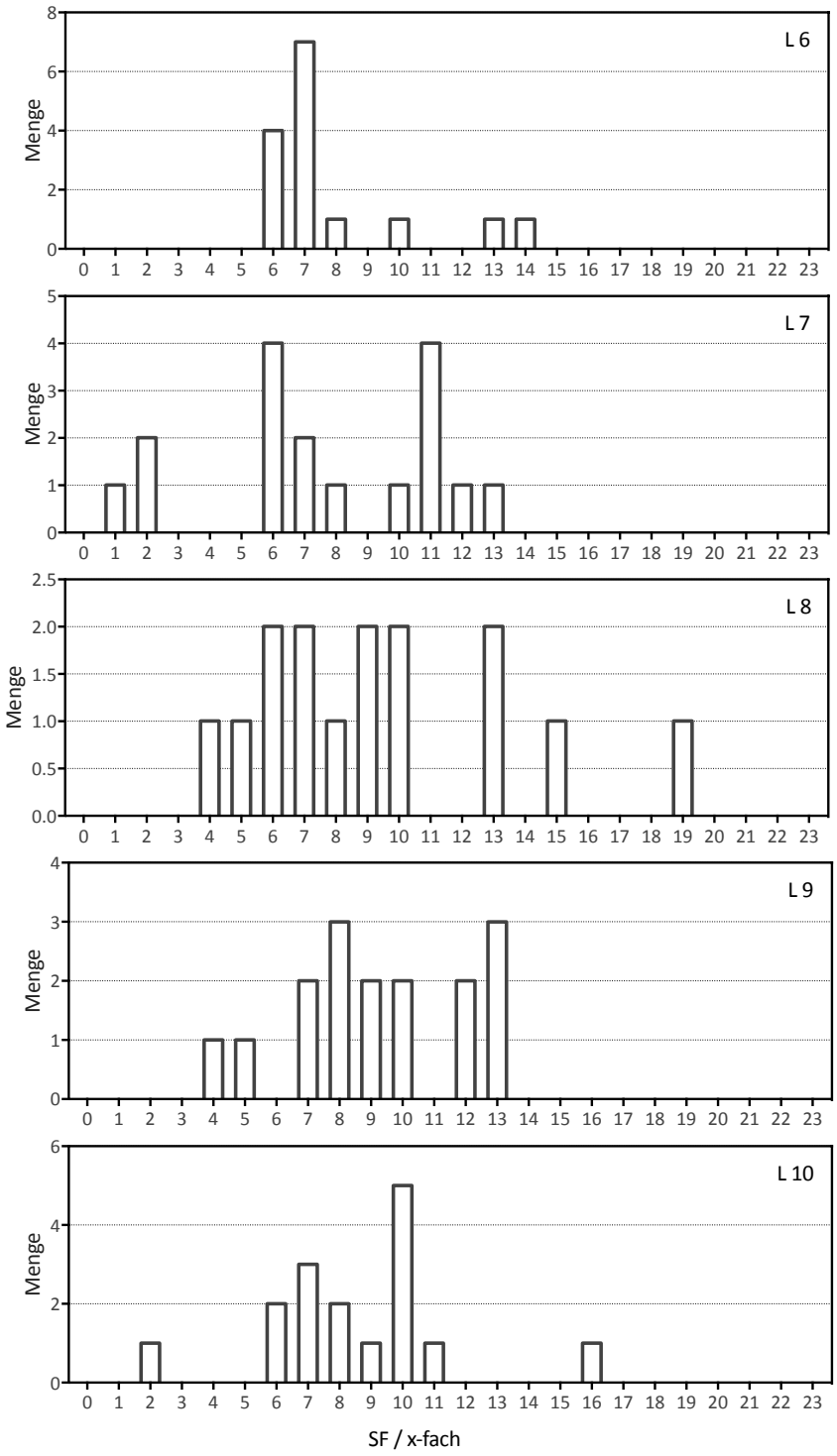


Abbildung 3.2.G Histogramme des Schaltfaktors der ersten Runde, nach Stammlänge geordnet. Die Menge zeigt an, wie häufig ein bestimmter Schaltfaktor (SF) bei einer Stammlänge von sechs-zehn bp (L6-L10) vorkam.

3.3 Ergebnisse der *machine learning* Runde 2

In der 2. Runde, welche die erste war, die auf *machine learning*-basierend vorhergesagt wurde, wurden 81 Konstrukte kloniert und gemessen. Die Stammlänge der unterschiedlichen P1-Stämme der getesteten Konstrukte lag zwischen sechs und zehn nt. Abbildung 3.3.A zeigt die Basalexpression in Bezug auf den Schaltfaktor der 2. Runde. Die Veränderung des Stammes P1 führte, ähnlich wie bei der 1. Runde, zu einer breiten Verteilung des Schaltfaktor mit einem Mittelwert von 7,7-fach. Der Mittelwert der Runde ist etwas geringer, als der Mittelwert der 1. Runde, jedoch konnte hier ein Riboswitch mit einem Schaltfaktor von 19,2-fach gefunden werden. Dieser Riboswitch hat den bis zu dieser Runde höchsten Schaltfaktor. Zudem konnte ein weiterer Riboswitch mit einem ähnlich hohen Schaltfaktor gefunden werden, dieser liegt bei 18,5-fach. Die Verteilung der Basalexpression ist in dieser Runde noch breiter als in der 1. Runde. Man kann erkennen, dass Konstrukte mit einer höheren Basalexpression eher einen niedrigeren Schaltfaktor haben. Auch hier gibt es wieder Konstrukte, welche einen Schaltfaktor nahe eins haben und kaum eine Basalexpression zeigen.

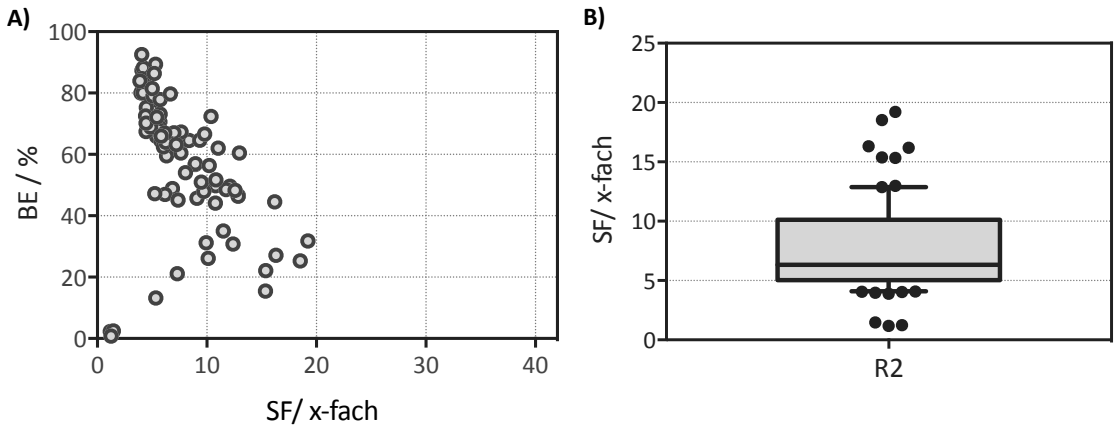


Abbildung 3.3.A Ergebnisse der Messungen aus der 2. Runde, Schaltfaktor und Basalexpression A) Basalexpression in % in Bezug auf den Schaltfaktor (SF), jeder Punkt zeigt den Mittelwert aus 2 unabhängigen Messungen). **B)** Boxplot des Schaltfaktors der Konstrukte von Runde 2. Whisker stellen P10 und P90 dar.

Abbildung 3.3.B zeigt, in welchem Bereich sich der Tm des P1-Stammes in der 2. Runde befindet und welcher Schaltfaktor welchem Tm zuzuordnen ist. Der Tm ist breiter verteilt als in der 1. Runde und geht von -30°C bis 60°C. Die P1-Stämme der drei Konstrukte mit den höchsten Schaltfaktoren befinden sich in einem Bereich zwischen 45°C und 50°C. Die Konstrukte, welche hier den niedrigsten Tm des P1 Stammes aufweisen, haben auch sehr niedrige Schaltfaktoren. Drei Konstrukte mit sehr geringen Schaltfaktoren haben allerdings einen Tm zwischen 40°C und 60°C.

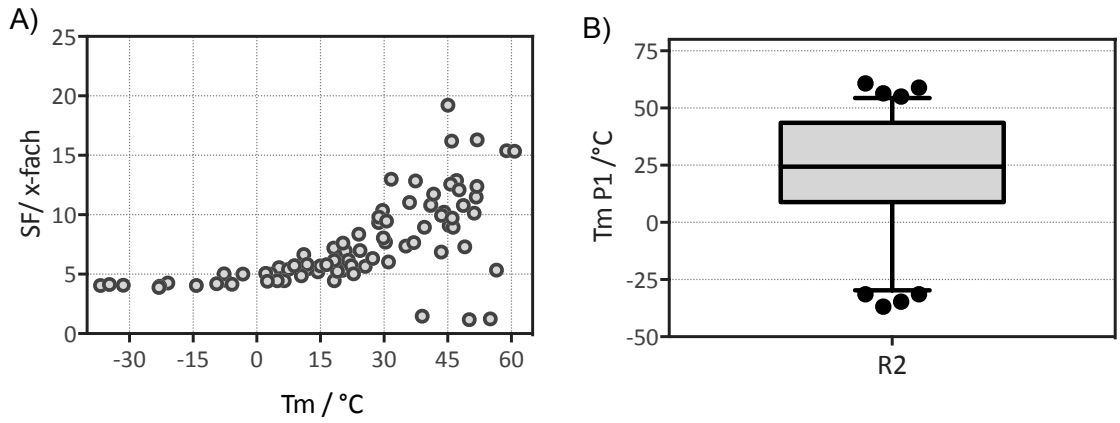


Abbildung 3.3.B Ergebnisse der Messungen aus der 2. Runde, Schaltfaktor und Tm des P1-Stamms in °C A) Tm in °C in Bezug auf den Schaltfaktor (SF), jeder Punkt zeigt den Mittelwert aus 2 unabhängigen Messungen. **B)** Boxplot des Tm P1 der Konstrukte von Runde 2. Whisker stellen P5 und P95 dar.

Abbildung 3.3.C zeigt die Verteilung des biophysikalischen Parameters ΔG der 2. Runde. Dieser liegt in einem Bereich zwischen -30 und -12 kcal/mol. Riboswitche, die einen höheren Schaltfaktor haben, liegen im Bereich zwischen -27 kcal/mol und -25 kcal/mol, also sehr ähnlich der 1. Runde. Riboswitche mit einem höheren ΔG haben vermehrt einen niedrigeren Schaltfaktor. Drei Konstrukte, mit einem sehr niedrigen Schaltfaktor liegen in einem Bereich zwischen -30 und - 23 kcal/mol.

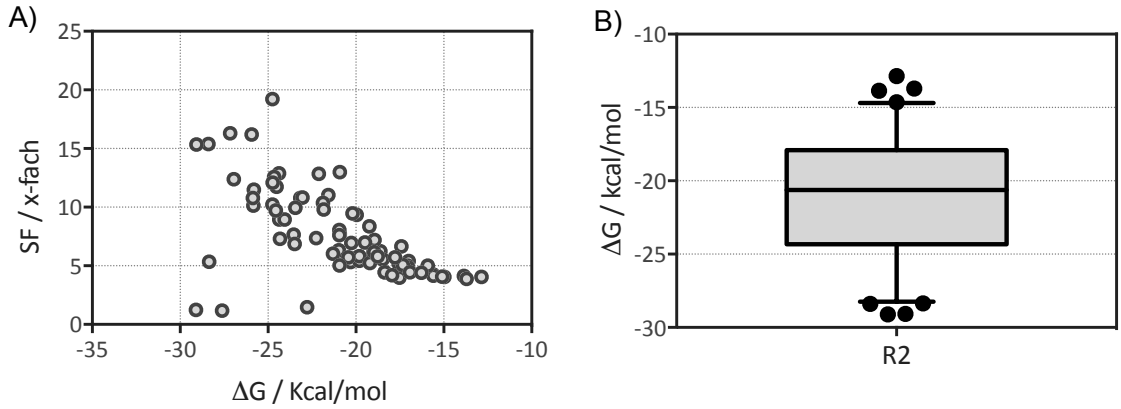


Abbildung 3.3.C Ergebnisse der Messungen aus der 2. Runde, Schaltfaktor und ΔG des P1-Stamms in kcal/mol A) ΔG in kcal/mol in Bezug auf den Schaltfaktor der 2. Runde (SF), jeder Punkt zeigt den Mittelwert aus zwei unabhängigen Messungen). **B)** Boxplot des ΔG-Wertes der Konstrukte von Runde 2. Whisker stellen P5 und P95 dar.

Die Verteilung des GC-Gehalts der P1-Stämme in % und die Verteilung in Bezug auf den Schaltfaktor der 2. Runde ist in Abbildung 3.3.C gezeigt. Der GC-Gehalt ist sehr breit verteilt und befindet sich in einem Bereich zwischen 5% und 90%. Höher schaltende Konstrukte befinden sich in einem Bereich

von 57% bis 78%. Auch hier sind die drei niedrig schaltenden Konstrukte zu finden. Sie weißen einen GC-Gehalt von 40% bis 60% auf. Bei einem GC-Gehalt von unter 20% finden sich keine Konstrukte, welche einen Schaltfaktor höher 5-fach aufweisen.

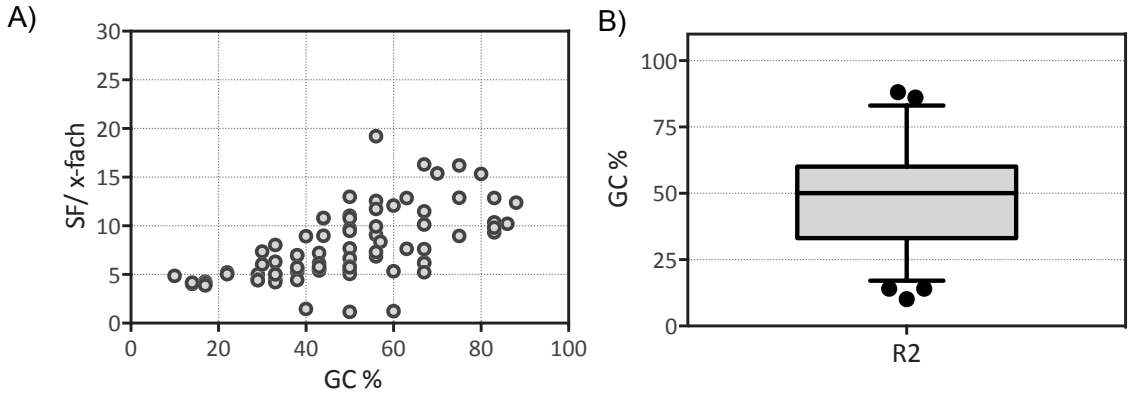


Abbildung 3.3.C Ergebnisse der Messungen aus der zweiten Runde, Schaltfaktor und GC-Gehalt in % A) GC-Gehalt von P1 in % in Bezug auf den Schaltfaktor (SF), jeder Punkt zeigt den Mittelwert aus zwei unabhängigen Messungen). **B)** Boxplot des GC-Gehalt von P1 in % von Runde 2. Whisker stellen P5 und P95 dar.

Abbildung 3.3.D zeigt die Verteilung der Entropie in bit des P1-Stamms und in Bezug auf den Schaltfaktor der 2. Runde. Die Entropie liegt, wie auch in der 1. Runde, zwischen 0,6 und 1,4. Eine höhere Entropie ab 0,9 zeigt höhere Schaltfaktoren, jedoch befinden sich in diesem Bereich auch Konstrukte mit einem niedrigen Schaltfaktor. Die drei Konstrukte mit einem sehr niedrigen Schaltfaktor, liegen in etwa bei einer Entropie von 1,3, wobei zwei der drei Konstrukte in der Abbildung übereinander liegen. Sie haben beide eine Entropie von 1,27 und einen Schaltfaktor von 1,1-fach bzw. 1,2-fach.

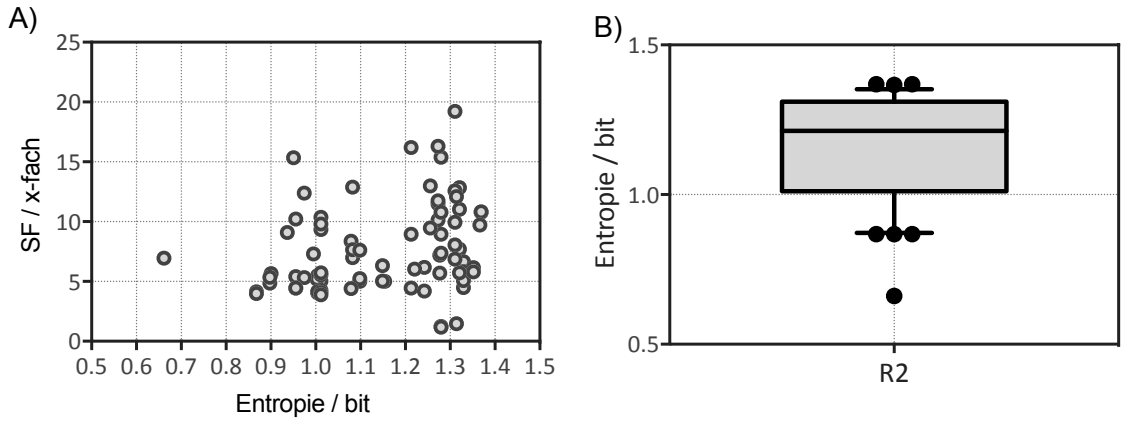


Abbildung 3.3.D Ergebnisse der Messungen aus der 2. Runde Schaltfaktor und Shannon-Entropie in bit A) Entropie in bit in Bezug auf den Schaltfaktor (SF), jeder Punkt zeigt den Mittelwert aus zwei unabhängigen Messungen). **B)** Boxplot der Entropie in bit von Runde 2. Whisker stellen P5 und P95 dar.

In Abbildung 3.3.E ist ein Histogramm der Stammlänge in Bezug auf die Häufigkeit der jeweiligen Schaltfaktoren dargestellt. Die Stammlängen wurden etwa gleichhäufig vorhergesagt. Am seltensten kommen Stämme mit der Länge sieben nt vor. In dieser Runde wurden die höchsten Schaltfaktoren mit Stämmen der Länge neun nt erreicht.

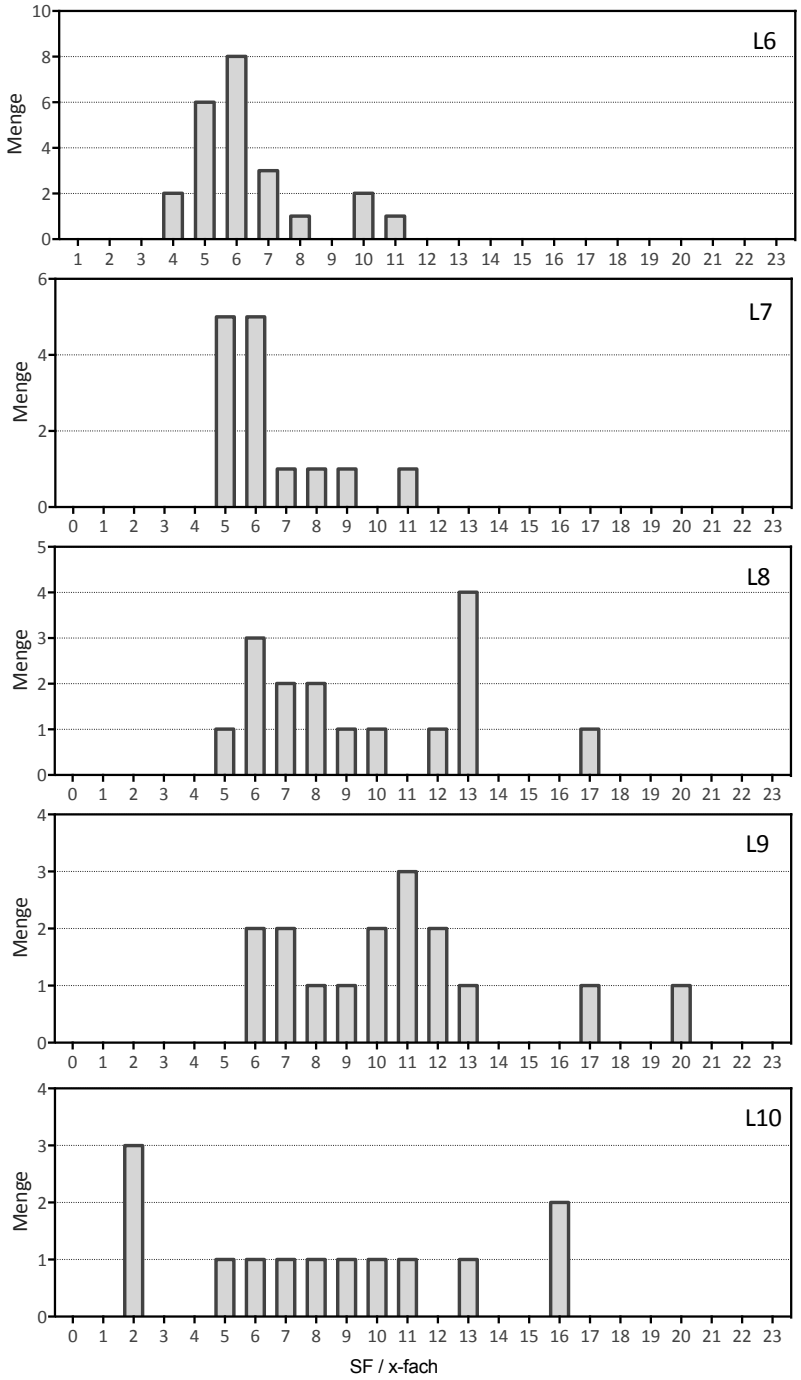


Abbildung 3.3.E Histogramme des Schaltfaktors der 2. Runde, nach Stammlänge geordnet. Die Menge zeigt an, wie häufig eine bestimmter Schaltfaktor in x-fach bei einer Stammlänge von 6-10 bp (L6-L10) vorkam.

3.3.1 Fazit nach der 2ten - *machine learning* - Runde

Mit der 2. Runde, welche erstmals auf *machine learning* basierte, konnte noch keine Erhöhung des mittleren Schaltfaktors erreicht werden. Dennoch konnte ein Riboswitch (R2-E7) gefunden werden, der einen höheren Schaltfaktor aufweist als das Ursprungskonstrukt LG3. In der 1. und der 2. Runde wurden zudem sechs Konstrukte gefunden, welche keine oder nur eine sehr geringe Schaltfähigkeit aufzeigten. Eine Analyse dieser Riboswitche nach der 2. Runde ergab, dass die P1-Stämme der Konstrukte entweder ein AUG enthielten oder mit dem vorangehenden CAA-Spacer ein AUG bilden (5´->3´-Sequenzen: R1-B5 **(A)UG**AGCAC; R1-E3 AGA**AUG**G; R1-F3 **(A)UG**UCUGU; R1-E12 UA**AUG**UCUCA; R2-A10 **(A)UG**CGAAAGUU; R2-H9 **(A)UG**AGCCAAGA und R2-G10 **(A)UG**GCUGGAUC). Das AUG an dieser Stelle des Konstruktes stört die Basalexpression maßgeblich, insbesondere, wenn dieses nicht *in frame* mit dem eigentlichen AUG des GFP-Gens ist. In der dritten Runde wurden deshalb Konstrukte, welche ein AUG enthielten aus der Vorhersage ausgeschlossen.

3.4 Ergebnisse der *machine learning* Runde 3

In der 3. Runde (wie die 2. Runde mittels *machine learning* vorhergesagt) wurden 83 Konstrukte kloniert und gemessen. Die Stammlänge der unterschiedlichen P1-Stämme der getesteten Konstrukte lag zwischen sechs und zehn nt. Abbildung 3.4.A zeigt die Basalexpression in Bezug auf den Schaltfaktor der 2. Runde. Die Randomisierung des Stammes P1 führte, ähnlich wie bei der ersten und zweiten Runde, zu einer breiten Verteilung des Schaltfaktors und der Basalexpression. Allerdings kann hier eine Verbesserung des mittleren Schaltfaktors verzeichnet werden, dieser steigt auf 11,3-fach an. Der Mittelwert dieser Runde ist also viel höher als der Mittelwert der ersten beiden Runden. Zudem konnten in dieser Runde vier Riboswitche mit einem Schaltfaktor über 20-fach gefunden werden.

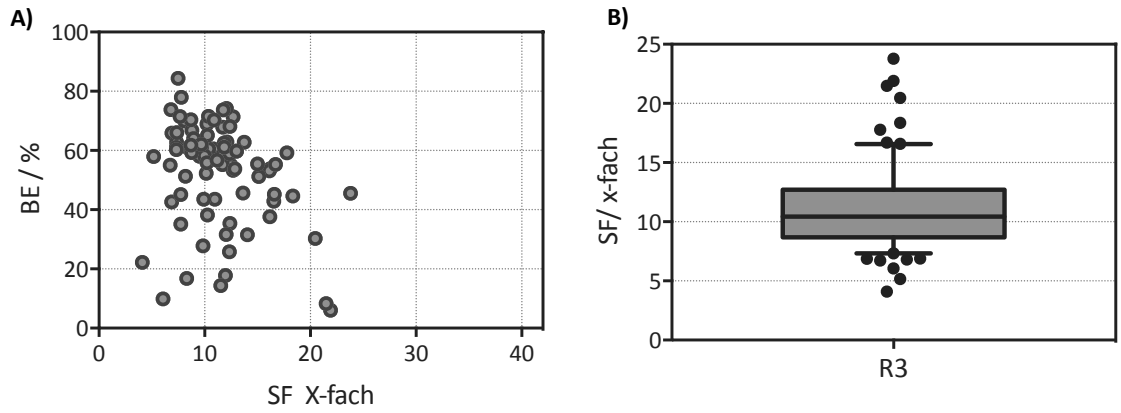


Abbildung 3.4.A Ergebnisse der Messungen aus der 3. Runde, Schaltfaktor und Basalexpression A) Basalexpression (BE) in % in Bezug auf den Schaltfaktor (SF), jeder Punkt zeigt den Mittelwert aus 2 unabhängigen Messungen. **B)** Boxplot des Schaltfaktors der Konstrukte von Runde 3. Whisker stellen P10 und P90 dar.

Im Gegensatz zu den ersten beiden Runden zeigt sich der Tm in Runde 3 nun in einer weniger breiten Verteilung (Abbildung 3.4.B). Der Bereich ist eingegrenzter und befindet sich nun zwischen 15°C und 55°C. Besonders stark ist hier der Unterschied zu Runde 2. Die P1-Stämme der vier Konstrukte mit den höchsten Schaltfaktoren befinden sich in einem Bereich zwischen 45°C. Die Konstrukte, welche hier den niedrigsten Tm des P1-Stammes aufweisen, haben auch niedrige Schaltfaktoren. Bei einem Tm in der Nähe von 45 °C schaltet kein Konstrukt schlechter als 9-fach.

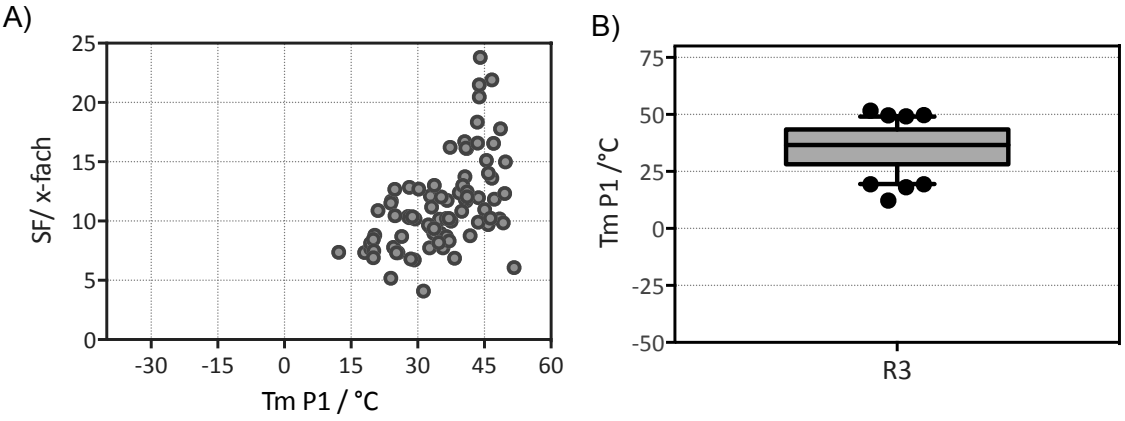


Abbildung 3.4.B Ergebnisse der Messungen aus der 3. Runde, Schaltfaktor und Tm des P1-Stamms in °C A) Tm in °C in Bezug auf den Schaltfaktor (SF), jeder Punkt zeigt den Mittelwert aus 2 unabhängigen Messungen. **B)** Boxplot des Tm P1 der Konstrukte von Runde 3. Whisker stellen P5 und P95 dar.

Abbildung 3.4.C zeigt die Verteilung des biophysikalischen Parameters ΔG von Runde 3. Die ΔG-Werte der 3. Runde befinden sich in einem weitaus zentrierten Bereich als in den ersten beiden Runden und liegen zwischen -27 und -18 kcal/mol. Die Riboswitche mit dem höchsten Schaltfaktor liegen alle bei etwa -25 kcal/mol. In diesem ΔG Bereich befindet sich auch kein Riboswitch mit einem

Schaltfaktor geringer als 9-fach. Konstrukte mit einem höheren ΔG haben auch hier häufiger einen niedrigeren Schaltfaktor.

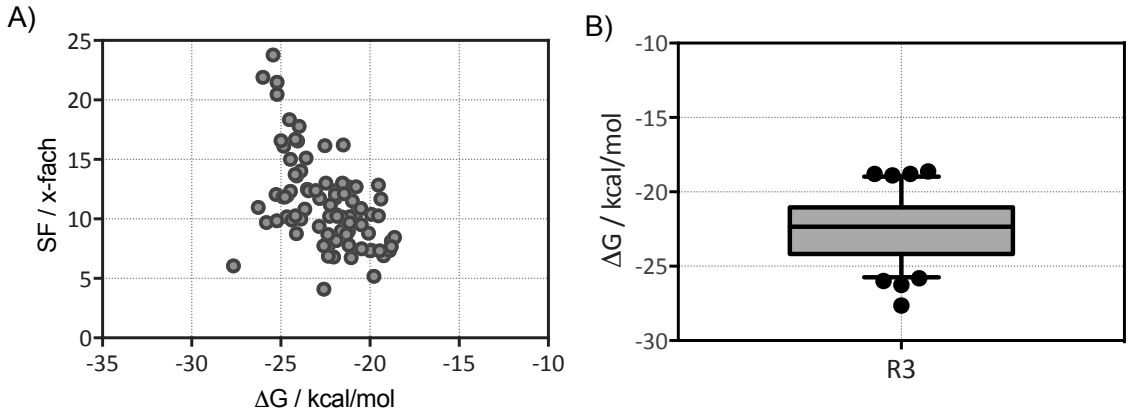


Abbildung 3.4.C Ergebnisse der Messungen aus der 3. Runde, Schaltfaktor und ΔG des P1-Stamms in kcal/mol A) ΔG in kcal/mol in Bezug auf den Schaltfaktor (SF) der 3. Runde, jeder Punkt zeigt den Mittelwert aus 2 unabhängigen Messungen. **B)** Boxplot des ΔG-Wertes der Konstrukte von Runde 3. Whisker stellen P5 und P95 dar.

Die Verteilung des GC-Gehalts der P1-Stämme in % und die Verteilung in Bezug auf den Schaltfaktor der 3. Runde ist in Abbildung 3.4.D gezeigt. Der GC-Gehalt ist hier nun weniger breit verteilt, verzeichnet jedoch auch zwei Stämme mit einem GC-Gehalt von 100%. Die vier Riboswitche mit den höchsten Schaltfaktoren, liegen in einem Bereich von 50% - 63%. Besser schaltende Konstrukte befinden sich in einem Bereich von 57% bis 67%.

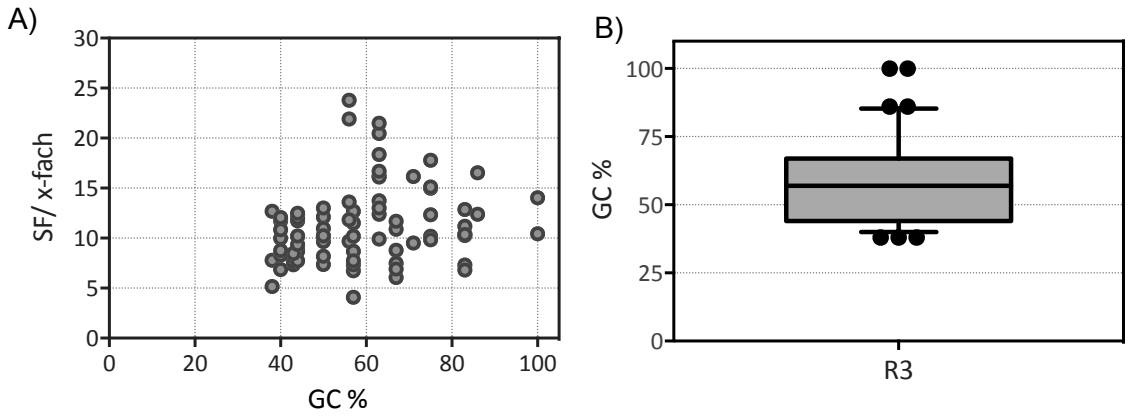


Abbildung 3.4.D Ergebnisse der Messungen aus der 3. Runde, Schaltfaktor und GC-Gehalt in % A) GC-Gehalt von P1 in % in Bezug auf den Schaltfaktor (SF), jeder Punkt zeigt den Mittelwert aus 2 unabhängigen Messungen. **B)** Boxplot des GC-Gehalt von P1 in % von Runde 3. Whisker stellen P5 und P95 dar.

Abbildung 3.4.E zeigt die Verteilung der Entropie in bit des P1-Stamms und in Bezug auf den Schaltfaktor der 3. Runde. Die Entropie liegt, ähnlich wie in den Runden zuvor, zwischen 0,55 und 1,4.

Drei der vier hoch schaltenden Riboswitche liegen mit ihren P1-Stämmen in einem Entropiebereich von 1,3-1,4. Insgesamt ist die Verteilung der Entropie der P1-Stämme der 3. Runden den Verteilungen der ersten beiden Runden sehr ähnlich.

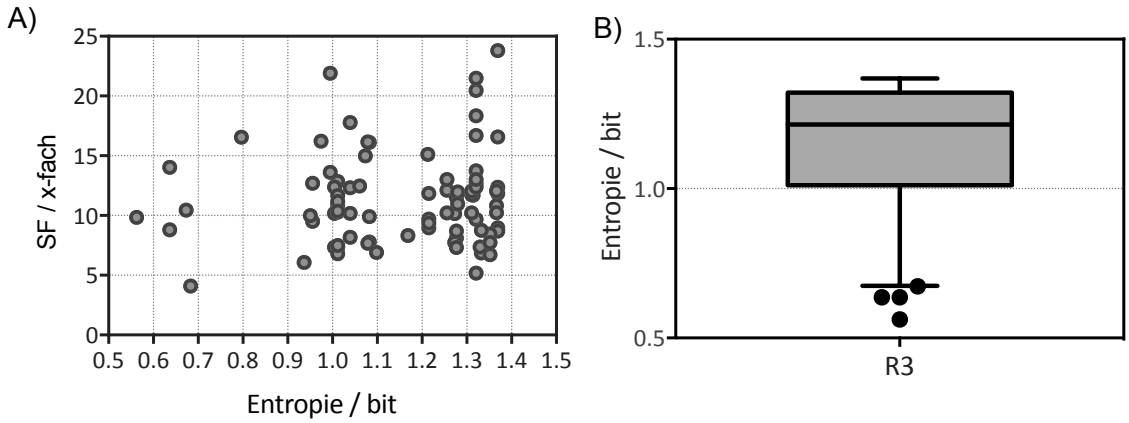


Abbildung 3.4.E Ergebnisse der Messungen aus der 3. Runde, Schaltfaktor und Shannon-Entropie in bit A) Entropie in bit in Bezug auf den Schaltfaktor (SF), jeder Punkt zeigt den Mittelwert aus 2 unabhängigen Messungen). **B** Boxplot der Entropie in bit von Runde 3. Whisker stellen P5 und P95 dar.

Abbildung 3.4.F zeigt ein Histogramm der Stammlänge in Bezug auf die Häufigkeit der jeweiligen Schaltfaktoren der 3. Runde. Stämme mit der Länge acht nt wurden am häufigsten vorhergesagt, Stämme mit der Stammlänge zehn nt wurden am seltensten vorhergesagt. Die restlichen Stammlängen wurden etwa gleichhäufig vorhergesagt. Die höchsten Schaltfaktoren konnten mit Stämmen der Länge acht nt und neun nt erreicht werden.

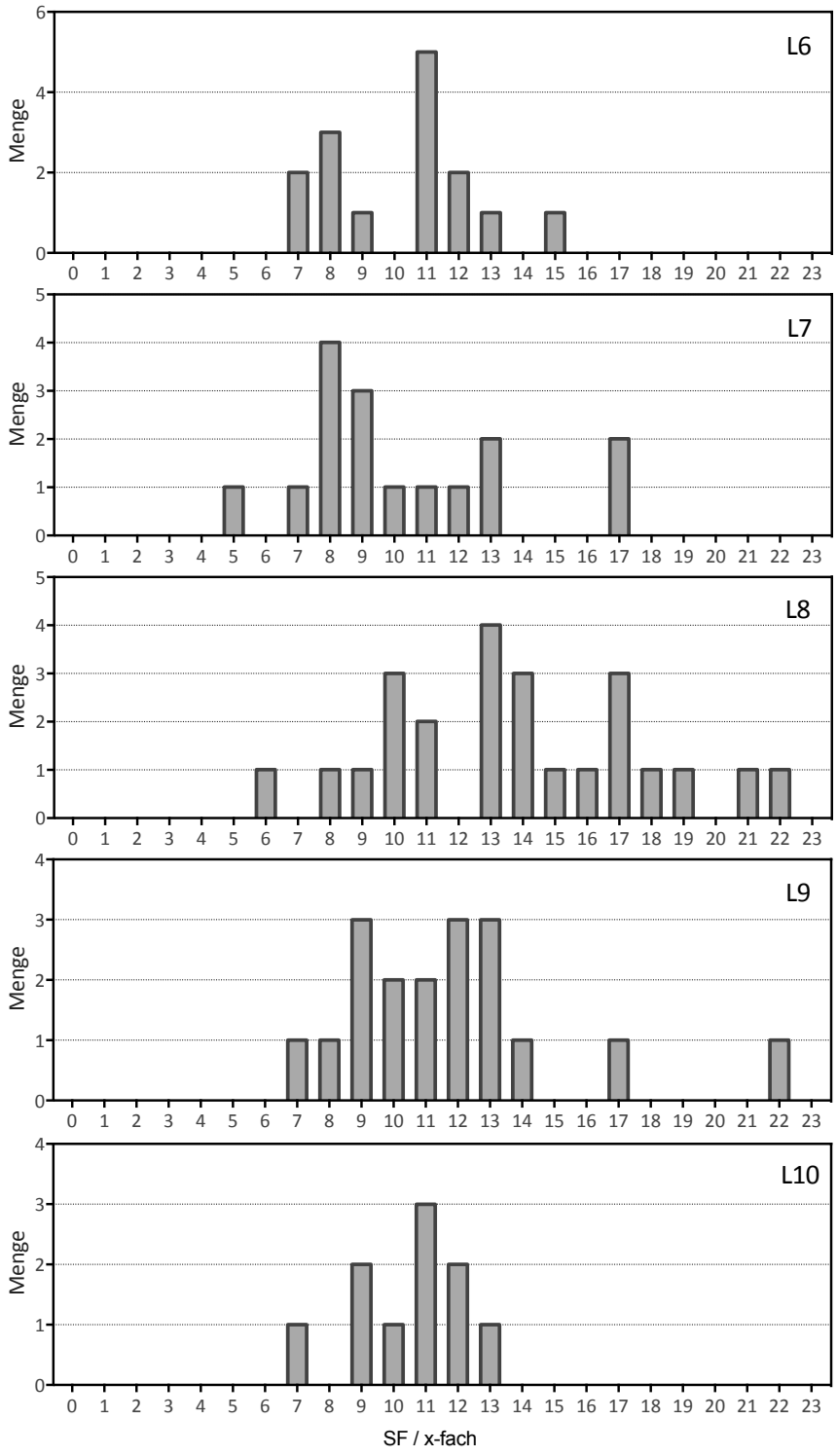


Abbildung 3.4.F Histogramme des Schaltfaktors der 3. Runde, nach Stammlänge geordnet. Die Menge zeigt an, wie häufig eine bestimmter Schaltfaktor (SF) in x-fach bei einer Stammlänge von 6-10 bp (L6-L10) vorkam.

3.4.1 Fazit nach der 3. Runde des *machine learnings*

Mit der 3. Runde konnte der mittlere Schaltfaktor der Runde im Vergleich zu den ersten beiden Runden erhöht werden. Zudem konnten einige sehr gut schaltende Riboswitche gefunden werden: R3-B7 (5´-> 3´Sequenz AUCGGUGAC) mit einem Schaltfaktor von fast 24-fach, R3-D2 (5´-> 3´Sequenz: UCCCACAUC) mit einem Schaltfaktor von ca. 22-fach, R3-E4 (5´-> 3´Sequenz: AGGGCAUC) mit einem Schaltfaktor von 21,5-fach und R3-D10 (5´-> 3´Sequenz: AGGCAUCC) mit einem Schaltfaktor von 20,5-fach. Die Riboswitche R3-D2 und R3-E8 fallen zudem durch ihre niedrige Basalexpression auf, diese liegt mit 6% bzw. 8% deutlich unter dem Mittelwert der Basalexpression der 3. Runde. Beide P1-Stämme verfügen an der gleichen Stelle über eine CAU-Sequenz, das im komplementären Teil des Stammes ein AUG ausbildet, welches mit dem Startcodon *in frame* ist. Deshalb ist möglicherweise dieser Sequenzabschnitt für die niedrige Basalexpression der beiden Riboswitche verantwortlich. Das Konstrukt R3-D10 enthält ebenfalls eine CAU-Sequenz, jedoch ist diese nicht mit dem Startcodon *in frame*, die Basalexpression dieses Konstruktes liegt bei etwa 30%.

3.5 Vergleichsrunde: Rationales Design und *machine learning*

Nach der 3. Runde wurde eine Vergleichsrunde durchgeführt. In dieser wurde die Leistung des *machine learning*-Programms gegen ein rationales Design-Konzept getestet. Es sollte die Frage beantwortet werden, ob eine analytische Herangehensweise im Schalter-Design mit der Vorhersagekraft des *random forest* verglichen werden kann. Getestet und miteinander verglichen wurden zehn Konstrukte, konstruiert durch rationales Design (RD) und zehn Konstrukte, welche mit dem *random forest* aus den bisher gesammelten Daten vorhergesagt wurde.

Für die auf rationalem Design basierende Vorhersage wurden die 8 Konstrukte mit den höchsten bisher erreichten Schaltfaktoren aller drei Runden gegen die 8 Konstrukte mit den niedrigsten Schaltfaktoren dieser Runden miteinander verglichen und analysiert (Tabelle 3.5.A und Tabelle 3.5.B).

Tabelle 3.5.A P1-Stämme mit den dazugehörigen biophysikalischen Parametern, Basalexpression (BE), Expression im Aus-Zustand (Aus) und dem Schaltfaktor (SF) aus Runde 1-3 mit einem hohem Schaltfaktor (SF) .

Name	Stamm P1 [5´-3´]	SF [x-fach]	BE [%]	Aus [%]	Entropie [bit]	Tm [°C]	dG [kcal/mol]	CG-Gehalt [%]
R3_B7	ATCGGTGAC	23,79	45,53	1,93	1,3689	44,0616	-25,43	56
R3_D2	TCCCACATC	21,90	6,07	0,28	0,9950	46,6857	-26	56
R3_E4	AGGGCATC	21,49	8,29	0,39	1,3209	43,8476	-25,21	63
R3_D10	AGGCATCC	20,47	30,27	1,50	1,3209	43,8476	-25,21	63
R2_E7	GGATAACCC	19,22	31,75	1,65	1,3108	45,0644	-24,75	56
R1_D6	GGTGTGCC	18,68	20,24	1,08	1,0397	49,5387	-25,79	75
R3_E8	ATTGGGCC	18,35	44,59	2,48	1,3209	43,3993	-24,51	63
R3_F3	CCTGGGTG	17,78	59,19	3,35	1,0397	48,6463	-23,97	75

Tabelle 3.5.B P1-Stämme mit den dazugehörigen biophysikalischen Parametern, Basalexpression (BE), Expression im Aus-Zustand (Aus) und dem Schaltfaktor (SF) aus Runde 1-3 mit einem niedrigen Schaltfaktor.

Name	Stamm P1 [5´-3´]	SF [X-fach]	BE [%]	Aus [%]	Entropie [bit]	Tm [°C]	dG [kcal/mol]	CG-Gehalt [%]
R2_H12	GAATAT	3,88	83,95	21,92	1,0114	-23,0284	-13,71	17
R1_G5	GAAAAAGG	3,89	37,09	9,16	0,6616	17,1408	-17,97	38
R2_D12	ATAAGA	3,98	80,01	20,08	0,8676	-22,8731	-17,53	17
R2_C11	CAAAAT	4,04	92,51	22,96	0,8676	-36,8339	-12,86	17
R2_A5	CTATTAA	4,05	84,70	20,91	1,0042	-14,2314	-14,97	14
R2_G12	AAATTC	4,08	87,35	21,45	1,0114	-31,4478	-15,09	17
R3_C10	GAGGAGA	4,10	22,19	5,41	0,6829	31,2970	-22,58	57
R2_D11	ACTTTT	4,13	81,35	19,69	0,8676	-34,7090	-13,86	17

Die Konstrukte mit hohem und niedrigen Schaltfaktoren unterscheiden sich in allen biophysikalischen Parametern. Die meisten Konstrukte mit niedrigen Schaltfaktor haben auch einen relativ niedrigen GC-Gehalt. Ein Konstrukt (R3-C10) mit einem höheren GC-Gehalt hat eine niedrige Entropie, im Stamm sind nur Purine enthalten. Der Tm des Stamms ist bei den Konstrukten mit einem niedrigen Schaltfaktor niedriger, der ΔG dagegen ist höher. Keines der guten Konstrukte hat weniger als acht Basenpaare. Die Konstrukte mit hohem Schaltfaktor unterscheiden sich von denen mit niedrigem Schaltfaktor auch in ihrer Sequenz. Es fällt auf, dass alle hoch schaltenden Konstrukte aus den ersten drei Runden über ein oder zwei Cytosine am Stammende verfügen. Auch verfügen viele der besseren Konstrukte über zwei Guanine hintereinander, die sich meist am Anfang des Stamms befinden. Bei sechs der acht guten Konstrukte befinden sich unter den ersten drei Nukleotiden des Stamms mindestens zwei Guanine oder Cytosine, also Basen, welche drei Wasserstoffbrücken ausbilden. Dagegen finden sich in der Mitte des Stamms häufiger Adenine oder Uracile, welche nur zwei Wasserstoffbrücken ausbilden. Anhand dieser Stamm-Analyse wurden zehn neue Konstrukte designt

(Tabelle 3.5.C). Die Konstrukte, die anhand der vorherigen, biophysikalischen Parameter mit dem *random forest* vorhergesagt wurden, sind in Tabelle 3.5.D zu finden.

Tabelle 3.5.C: Auf Grund von rationalem Design vorgeschlagene P1-Stämme mit biophysikalischen Parametern, Basalexpression (BE), Expression im Aus-Zustand (Aus) und dem Schaltfaktor (SF).

Name	Stamm P1 [5´-3´]	SF [x-fach]	BE [%]	Aus [%]	Entropie [bit]	Tm [°C]	dG [kcal/mol]	CG-Gehalt [%]
ML-RD_A1	AGGCTACG	12,35	36,27	2,95	1,3591	40,87	-22,81	63
ML-RD_A2	AGGCGATC	12,33	53,75	4,38	1,3591	41,24	-24,86	63
ML-RD_A3	AGGCCATC	15,24	10,66	0,71	1,3591	43,85	-25,21	63
ML-RD_A4	ACCCTGTAC	9,74	33,37	3,45	1,3624	47,23	-25,69	56
ML-RD_A5	ACCCTGTTC	11,03	38,30	3,47	1,3624	45,95	-25,69	56
ML-RD_A6	GCTGAGCC	15,87	38,84	2,45	1,3527	49,20	-26,61	75
ML-RD_A7	TCGCGTAC	12,17	57,32	4,74	1,3591	39,11	-23,78	63
ML-RD_A8	ACCGGCTAAC	14,67	14,07	0,96	1,3619	53,73	-27,74	60
ML-RD_A9	AGCGCATC	10,34	15,64	1,50	1,3591	40,58	-24,82	63
ML-RD_A10	TCGGCAAC	14,54	50,04	3,45	1,3591	40,44	-24,13	63

Tabelle 3.5.D: P1-Stämme, die auf Basis der vorherigen drei Runden vom *random forest* vorhergesagt wurden mit ihren biophysikalischen Parametern, Basalexpression (BE), Expression im Aus-Zustand (Aus) und dem Schaltfaktor (SF).

Name	Stamm P1 [5´-3´]	SF [x-fach]	BE [%]	Aus [%]	Entropie [bit]	Tm [°C]	dG [kcal/mol]	CG-Gehalt [%]
ML-RD_S1	CGCGTGAAT	9,08	64,40	7,11	1,3624	39,83	-22,58	56
ML-RD_S2	AGAGGTA	3,67	83,45	23,02	1,3613	22,06	-20,28	43
ML-RD_S3	GAACGCA	8,18	61,12	7,49	1,3596	25,56	-20,61	57
ML-RD_S4	CCTAAAG	4,90	75,44	15,40	1,3613	11,31	-17,38	43
ML-RD_S5	GGTCCAA	9,42	77,90	8,28	1,3596	29,72	-20,68	57
ML-RD_S6	ACCGGAT	4,70	59,39	12,67	1,3596	29,56	-20,35	57
ML-RD_S7	GTACTGT	6,62	72,22	10,92	1,3613	18,72	-18,37	43
ML-RD_S8	TCGCATA	5,16	45,04	8,72	1,3613	18,37	-19,5	43
ML-RD_S9	CTTACCA	6,79	71,39	10,56	1,3613	16,48	-19,08	43
ML-RD_S10	CGACGTCT	12,54	46,47	3,70	1,3591	37,42	-22,11	63

Die auf rationalem Design basierten Konstrukte zeigen im Durchschnitt einen höheren GC-Gehalt, einen niedrigeren ΔG und einen höheren Tm als die durch *machine learning* vorhergesagten. Mehrheitlich sind die rationales Design Konstrukte über acht Basenpaare lang und orientieren sich in ihrer Sequenz an den acht besten Konstrukten der ersten beiden Runden. Die durch *machine learning* vorhergesagten Konstrukte zeigen meist eine höhere Basalexpression und einen niedrigen Schaltfaktor, als die durch rationales Design entworfenen Konstrukte (Abbildung 3.5). Im Mittel zeigen die durch rationales Design entworfenen Konstrukte einen Schaltfaktor von ca. 13-fach, während der Schaltfaktor der durch *machine learning* vorhergesagten Konstrukte im Mittel bei

sieben liegt. Dieser Unterschied ist signifikant. Mit den Konstrukten, welche durch rationales Design konstruiert wurden, wurden demnach bessere Ergebnisse erzielt.

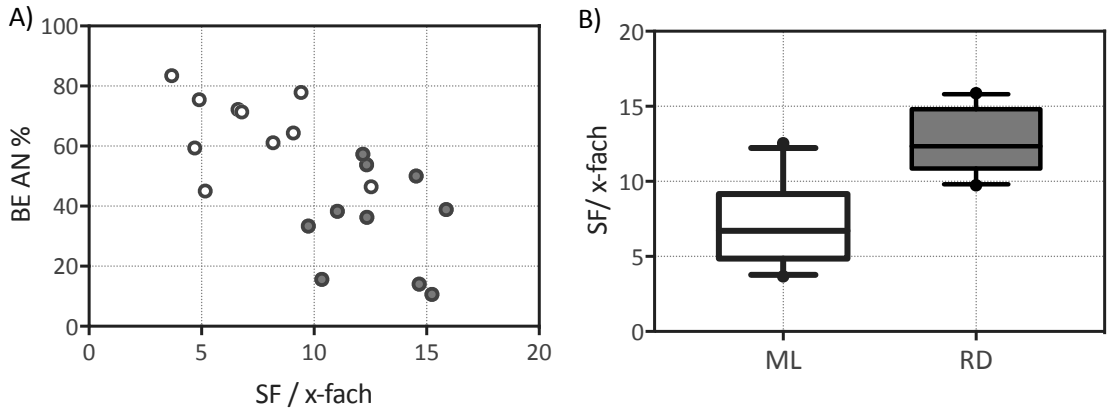


Abbildung 3.5 Ergebnisse *machine learning* – rationales Design Runde A) Basalexpression (BE) in % in Bezug auf den Schaltfaktor (SF) (jeder Punkt zeigt den Mittelwert aus 2 unabhängigen Messungen), weiß: Mittelwerte der *machine learning* Messung, grau: Mittelwerte der rationales Design Messung. **B)** Boxplot des Schaltfaktors (SF) der Konstrukte der *machine learning* – rationales Design Runde, weiß *machine learning* (ML); grau: rationales Design (RD). Whisker stellen P10 und P90 dar.

3.6 Ergebnisse der *machine learning* kombinierten *deep learning* Runde 4

In der 4. Runde wurde das *machine learning* Programm, auf Grund der Ergebnisse der durchgeführten Vergleichsrunde, unter anderem um ein *deep learning* Programm erweitert. Kapitel 3.1.4 beschreibt dazu die Herangehensweise und weitere Veränderungen des Programms. Von den 96 vorhergesagten Konstrukten wurden 94 kloniert und gemessen. Die Stammlänge der unterschiedlichen P1-Stämme der getesteten Konstrukte lag konstant bei acht nt. Abbildung 3.6.A zeigt die Basalexpression in Bezug zum Schaltfaktor der 4. Runde. Sowohl die verschiedenen Schaltfaktoren als auch die Basalexpressionen treten hier wesentlich zentrierter auf als in den Runden zuvor. Zudem kann hier wieder eine Verbesserung des mittleren Schaltfaktors verzeichnet werden, dieser steigt auf 13,3-fach an. Der Mittelwert der Schaltfaktoren dieser Runde ist nochmal höher als der Mittelwert der dritten Runde. Zudem konnten in dieser Runde ein Riboswitch mit einem Schaltfaktor nahe 40-fach gefunden werden.

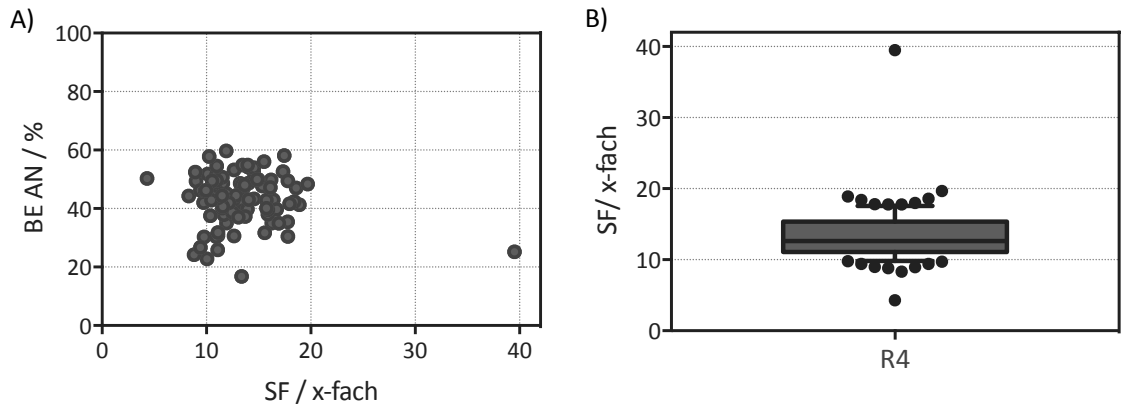


Abbildung 3.6.A Ergebnisse der Messungen aus der 4. Runde, Schaltfaktor und Basalexpression A) Basalexpression (BE) in % in Bezug auf den Schaltfaktor (SF), jeder Punkt zeigt den Mittelwert aus 3 unabhängigen Messungen. **B)** Boxplot des Schaltfaktors (SF) der Konstrukte von Runde 4. Whisker stellen P10 und P90 dar.

Der Tm des P1-Stammes der Konstrukte in Runde 4 ist stark eingegrenzt und begrenzt sich nur noch auf einen Bereich von 40°C - 50°C (Abbildung 3.6.B). Eine kleine „Einkerbung“ lässt sich bei 45°C erkennen, hier schaltet kein Konstrukt schlechter als 9-fach. Ein Konstrukt, das einen Tm von 44°C hat, schaltet nur 4-fach. Es ist das Konstrukt mit dem niedrigsten Schaltfaktor dieser Runde. Das Konstrukt mit dem höchsten Schaltfaktor hat im P1-Stamm ebenfalls einen Tm von 44°C.

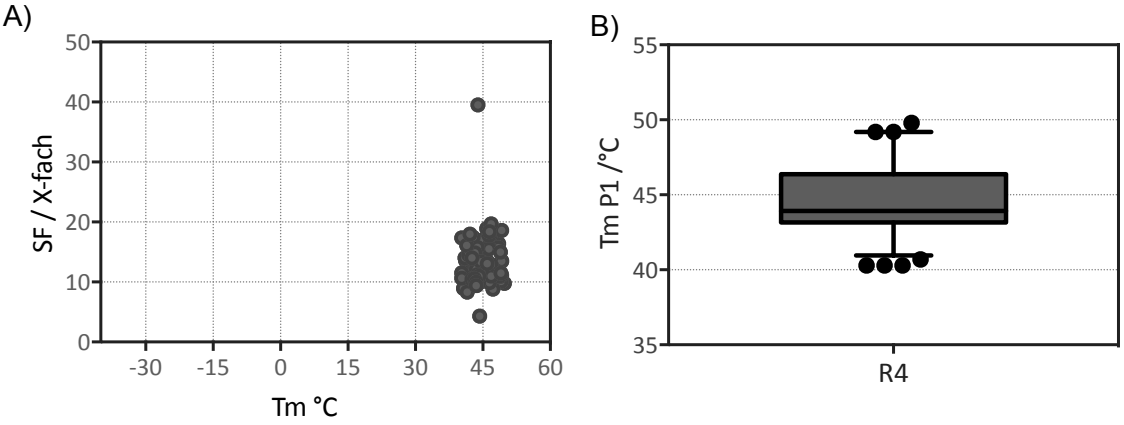


Abbildung 3.6.B Ergebnisse der Messungen aus der 4. Runde, Schaltfaktor und Tm des P1-Stamms in °C A) Tm in °C in Bezug auf den Schaltfaktor (SF), jeder Punkt zeigt den Mittelwert aus 3 unabhängigen Messungen. **B)** Boxplot des Tm P1 der Konstrukte von Runde 4. Whisker stellen P5 und P95 dar.

Abbildung 3.6.C zeigt die Verteilung des ΔG-Wertes der 4. Runde. Auch hier zeigt sich eine eingeschränkte Verteilung, die von -27 kcal/mol bis -23 kcal/mol geht. Genau wie in Runde 3 weißt der Riboswitch mit dem höchsten Schaltfaktor einen ΔG von -25 kcal/mol auf. Das Konstrukt dieser Runde mit dem niedrigsten Schaltfaktor hat einen dG von -24 kcal/mol.

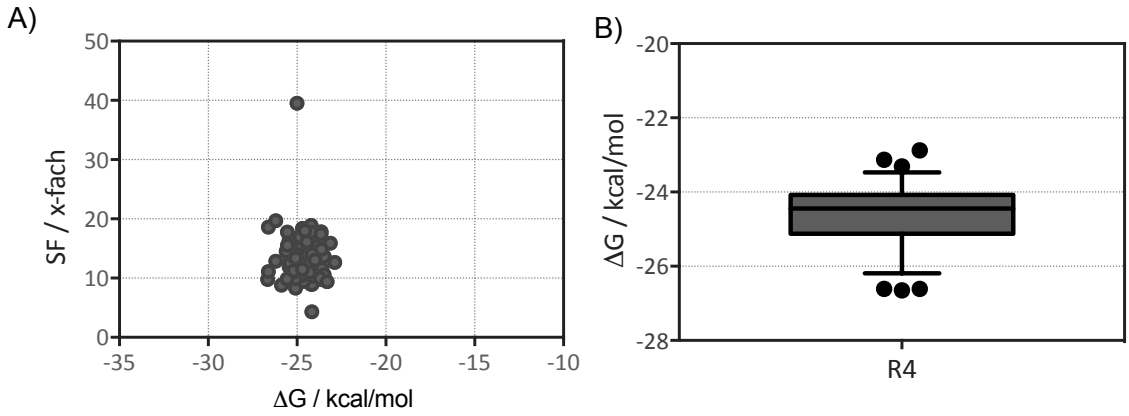


Abbildung 3.6.C Ergebnisse der Messungen aus der 4. Runde, Schaltfaktor und ΔG des P1-Stamms in kcal/mol A) ΔG in kcal/mol in Bezug auf den Schaltfaktor (SF) der 4. Runde, jeder Punkt zeigt den Mittelwert aus drei unabhängigen Messungen. **B)** Boxplot des ΔG-Wertes P1 Konstrukte von Runde 4. Whisker stellen P5 und P95 dar.

Auch bei der Verteilung des GC-Gehalts des P1-Stamms in der 4. Runde zeigt sich ein stark eingeschränktes Bild (Abbildung 3.6.D). Es wurden hier nur noch Stämme mit einem GC-Gehalt von 63% bzw. 75% vorhergesagt. Das Konstrukt mit dem höchsten Schaltfaktor hat einen GC-Gehalt von 63%, das Konstrukt mit dem niedrigsten Schaltfaktor ebenfalls.

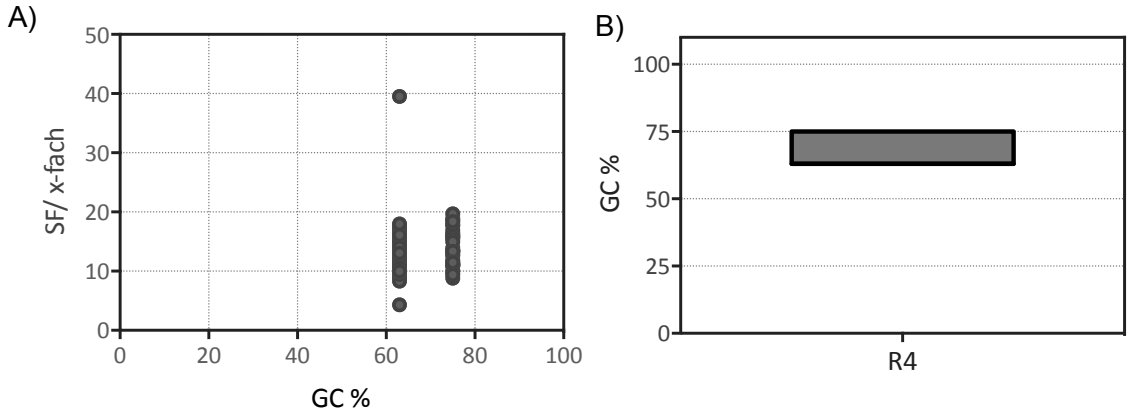


Abbildung 3.6.D Ergebnisse der Messungen aus der 4. Runde, Schaltfaktor und GC-Gehalt des P1-Stamms in % A) GC-Gehalt von P1 in % in Bezug auf den Schaltfaktor (SF) der vierten Runde, jeder Punkt zeigt den Mittelwert aus drei unabhängigen Messungen. **B)** Boxplot des GC-Gehalt von P1 in % von Runde 4.

Auch die Verteilung der Entropie beschränkt sich nur noch auf zwei Bereiche, 1,255 bit und 1,320 bit (Abbildung 3.6.E). Dies hängt auch mit der Länge des P1-Stammes zusammen, der in dieser Runde auf acht nt begrenzt wurde. Der Riboswitch mit dem höchsten Schaltfaktor weist in seinem P1-Stamm eine Entropie von 1,32 bit auf, genauso wie der P1-Stamm des Riboswitches mit dem niedrigsten Schaltfaktor.

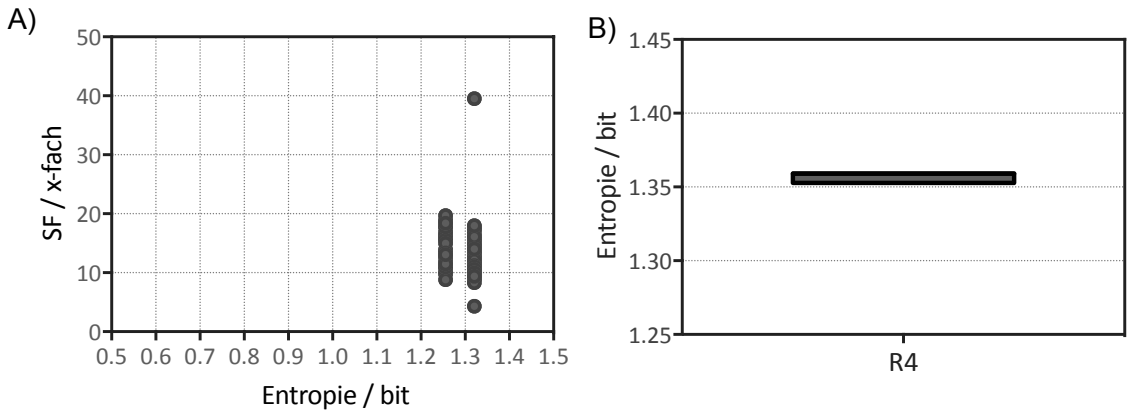


Abbildung 3.6.E Ergebnisse der Messungen aus der 4. Runde, Schaltfaktor und Shannon-Entropie des P1-Stamms in bit A) Entropie in bit in Bezug auf den Schaltfaktor (SF) der 4. Runde, jeder Punkt zeigt den Mittelwert aus drei unabhängigen Messungen. **B)** Boxplot der Entropie in bit von Runde 4.

In Runde 4 wurden das *machine learning*-Programm dahingehend verändert, dass nur noch Stämme der Länge acht nt vorhergesagt wurden. Das Histogramm in Abbildung 3.6.F zeigt, dass am häufigsten der Schaltfaktor um 12-fach erreicht werden konnte.

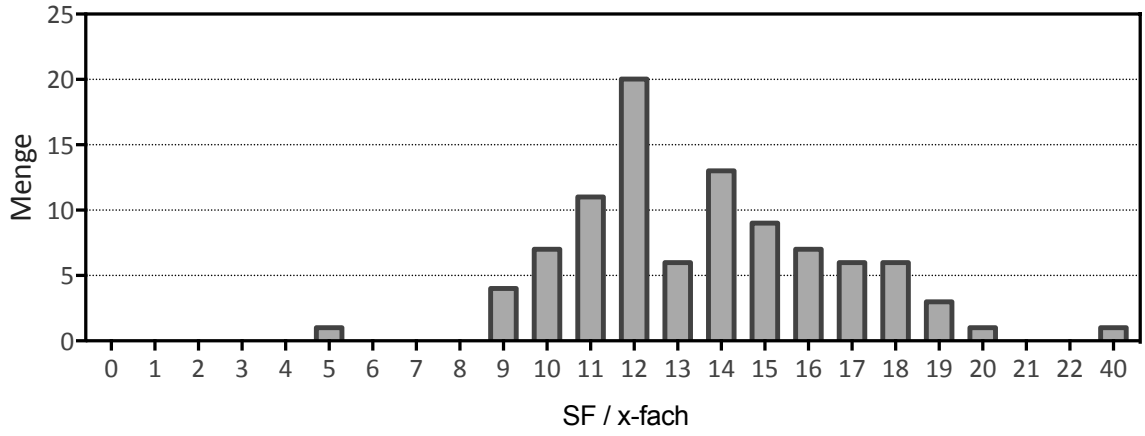


Abbildung 3.6.F Histogramme des Schaltfaktors der 4. Runde. Die Menge zeigt an, wie häufig ein Schaltfaktor (SF) in x-fach bei einer Stammlänge von acht bp vorkam.

3.7 Parameter-Vergleich aller Runden

In den vorangegangenen Kapiteln wurden die Ergebnisse der vier Runden im Einzelnen besprochen. In diesem Kapitel sollen diese nun direkt miteinander verglichen werden.

Abbildung 3.7.A zeigt den Verlauf der Schaltfaktoren und der Basalexpression der vier Runden. Während sich der mittlere Schaltfaktor von Runde 1 auf Runde 2 nicht verbesserte, konnte sowohl von Runde 2 auf Runde 3 als auch von Runde 3 auf Runde 4 eine signifikante Verbesserung des mittleren Schaltfaktors erreicht werden. Auch die Verteilung der Basalexpression änderte sich über die 4 Runden. Während der mittlere Schaltfaktor im Verlauf der Runden zunahm, nahm die Basalexpression ab. Ihren höchsten Wert erreicht sie in Runde 2, in dieser Runde ist der mittlere Schaltfaktor mit 7,7-fach am niedrigsten. Den niedrigsten Wert erreichte die Basalexpression mit der 4. Runde.

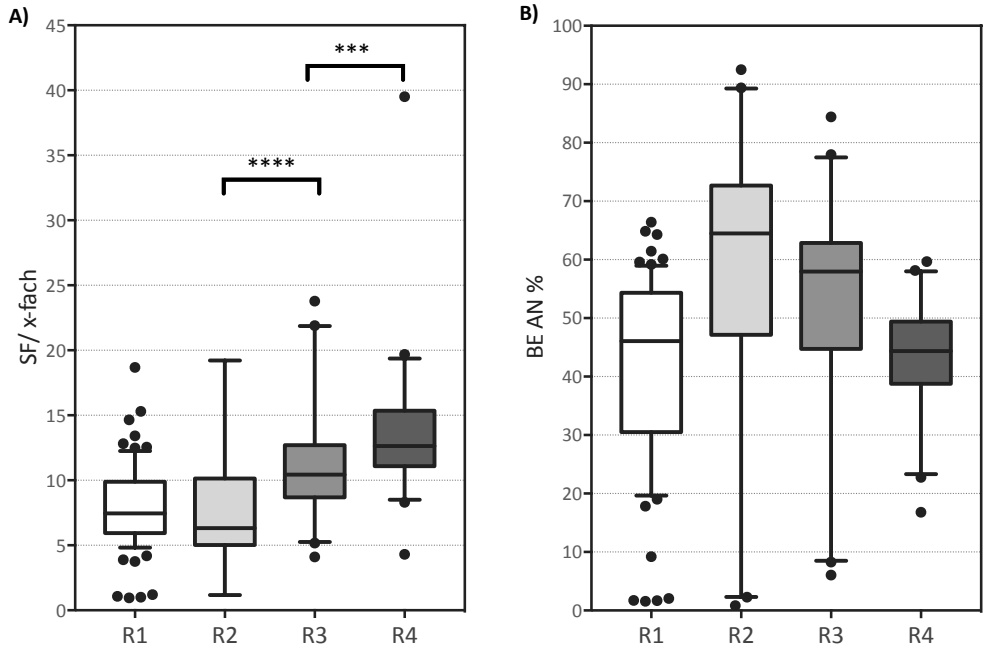


Abbildung 3.7.A Vergleich des Schaltfaktors und der Basalexpression aller vier Runden: A) Boxplot des Schaltfaktor (SF) von Runde eins bis vier. Die univariate Statistik mittels t-Test berechnet (gepaarter, zweigeteilter t-Test; *** = p value < 0,001; ; **** = p value < 0,0001). Whisker stellen P10 und P90 dar. **B)** Boxplot der Basalexpression von Runde 1 - 4. Whisker stellen P10 und P90 dar.

Auch die Verteilung der biophysikalischen Parameter ändert sich im Verlauf der vier Runden (Abbildung 3.7.B). Besonders stark kann man eine Veränderung zur 4. Runde erkennen, diese ist bei allen vier Parametern signifikant. Am stärksten verändert sich die Verteilung des Tm. Auch der ΔG erfährt eine starke Änderung. Der GC-Gehalt ändert sich erst ab Runde 3 signifikant. Am wenigsten

verändert sich die Entropie, hier wird die Veränderung erst ab Runde 4 signifikant. Dass die Sequenz in dieser Runde das erste Mal mitberücksichtigt wurde, könnte hier ein möglicher Einflussfaktor sein.

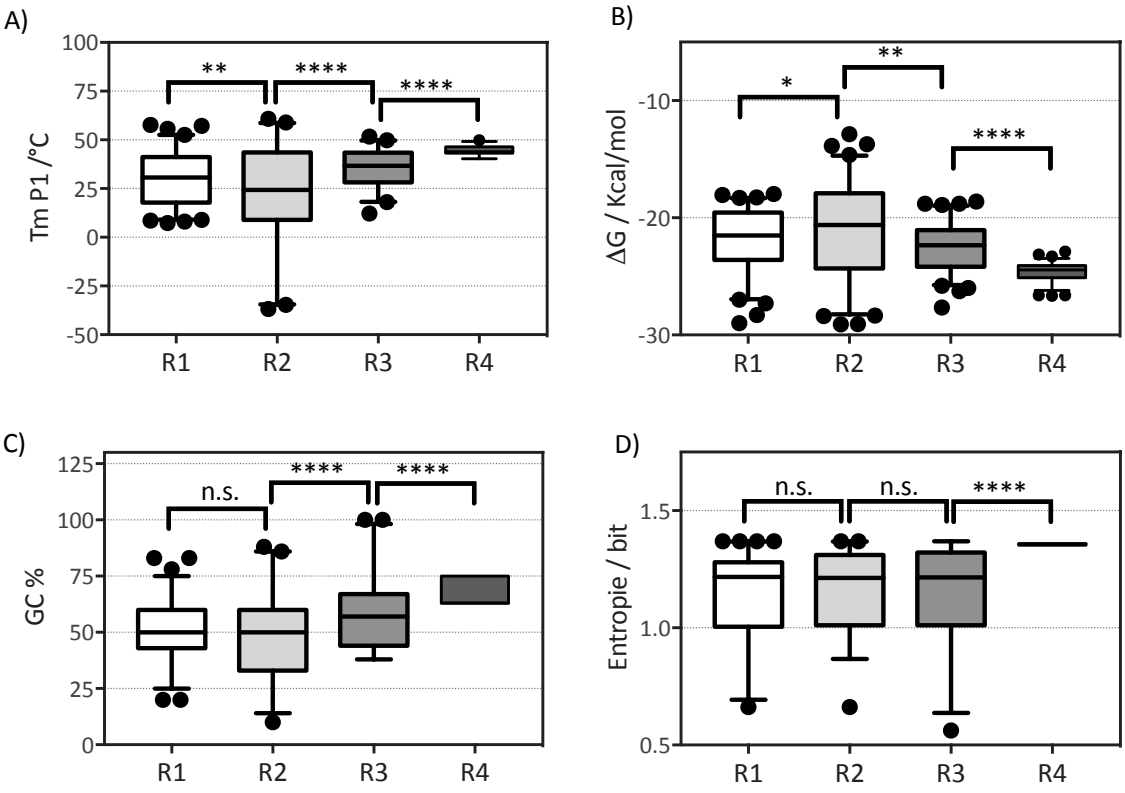


Abbildung 3.7.B Vergleich der biophysikalischen Parameter von Runde 1 bis Runde 4 Boxplot der biophysikalischen Parameter **A)** Tm / °C, **B)** ΔG / kcal/mol, **C)** GC% und **D)** Entropie / bit von Runde 1 - 4. Die univariate Statistik mittels t-Test berechnet (gepaarter, zweigeteilter t-Test; * = p value < 0,05; ** = p value < 0,01; *** = p value < 0,001; **** = p value < 0,0001). Whisker stellen P5 und P95 dar.

3.8 Ein Riboswitch mit 40-fachem Schaltfaktor

Mit der Ergänzung des *machine learning*-Programms um einen CNN und den unter Kapitel 3.1.4 genannten weiteren Maßnahmen konnte nicht nur der mittlere Schaltfaktor der getesteten Konstrukte signifikant verbessert werden (Abbildung 3.7.A), es konnte auch ein außergewöhnlich guter Riboswitch, R4-G8, gefunden werden, der ein mehr als doppelt so gutes Schaltverhalten und eine höhere Basalexpression aufweist als das Ausgangskonstrukt LG3 (Abbildung 3.8.A).

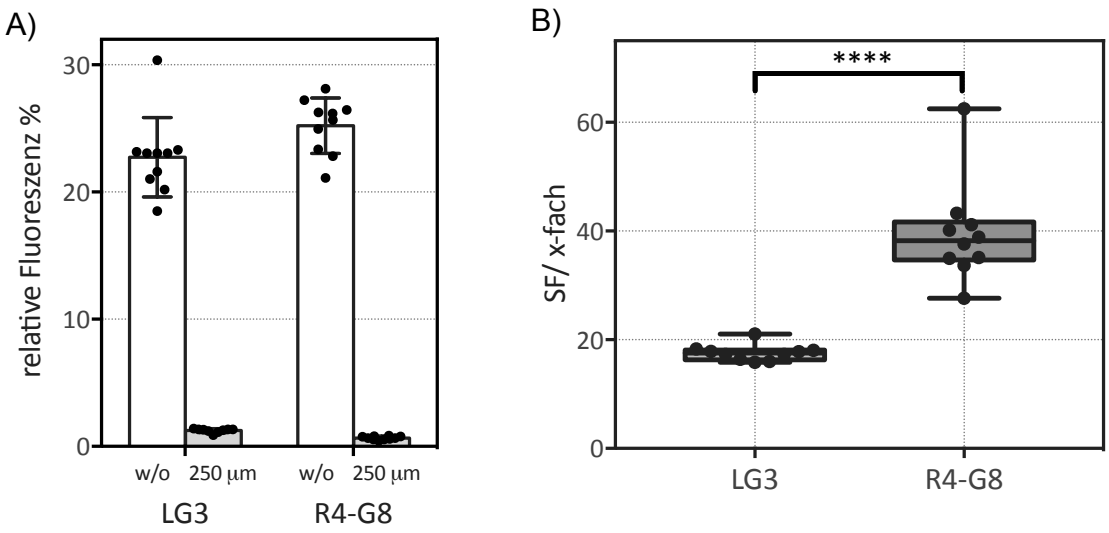


Abbildung 3.8.A Vergleich der Basalexpression und des Schaltverhaltens der Konstrukte LG3 und R4-G8 A) Vergleich der relativen Fluoreszenz von LG3 und R4-G8 ohne (w/o) und mit 250 µM Tetrazyklin. **B)** Boxplot: Vergleich des Schaltfaktors (SF) der beiden Konstrukte aus jeweils 10 Messungen. Die univariate Statistik mittels t-Test berechnet (gepaarter, zweigeteilter t-Test; **** = p value < 0,0001). Whisker stellen P10 und P90 dar.

Die relative Fluoreszenz im An-Zustand des neuen Konstruktes ist etwas höher, dagegen ist der Aus-Zustand (Liganden-gebundener Zustand) niedriger. Die Expression wird somit in diesem Konstrukt durch Ligandenbindung noch effektiver unterdrückt. Dieser Unterschied wirkt sich enorm auf den Schaltfaktor aus. Abbildung 3.8.B zeigt die Histogramme der mit dem Cytometer aufgezeichnete phänotypische Populationsanalyse der Hefe-Zellen. welche mit den Plasmiden der Konstrukte LG3 und R4-G8 transformiert worden waren. An der Häufigkeit der GFP-Fluoreszenzniveaus erkennt man, dass sich die Zellen welche mit TC inkubiert wurden, sehr viel näher am Hintergrund (dieser entspricht den Zellen, welche mit dem Plasmid IBB transformiert wurden) befinden, als die Zellen mit dem TC-Dimer LG3. Das GFP Signal wird sehr viel stärker unterdrückt, was die Folge einer verminderten Genexpression von GFP ist.

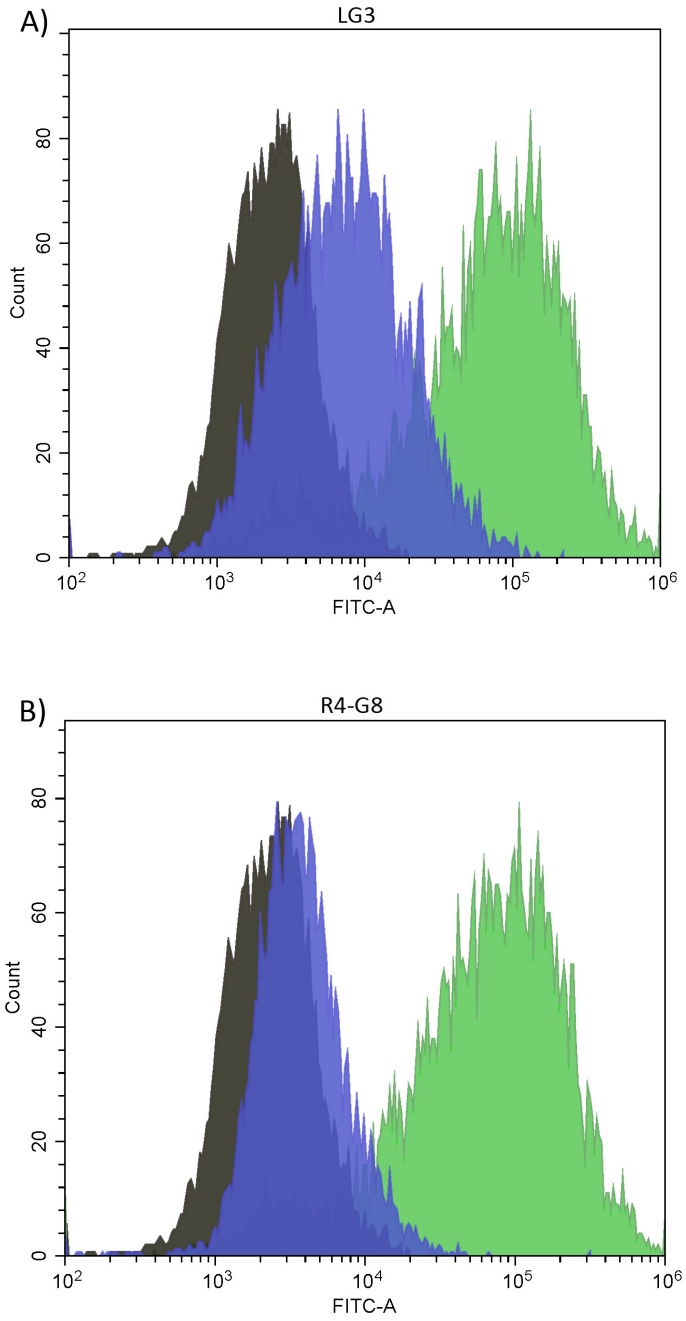


Abbildung 3.8.B Einzelzellanalyse des durch die im 5´UTR eines GFP-Gens inserierten Konstrukte LG3 und R4-G8. Das Histogramm zeigt die Häufigkeit jedes GFP-Fluoreszenzniveaus für Zellen die mit und ohne 250 µm TC inkubiert wurden und welche die Riboswitche LG3 und R4-G8 im 5´UTR eines GFP-Gens enthalten. Für die Messung wurden 5000 Ereignisse pro Population aufgezeichnet. Die x-Achse ist biexponentiell aufgetragen. Eine Population der Zellen wurden ohne TC inkubiert (grün) und eine Population mit 250 µm TC inkubiert (blau). In schwarz dargestellt ist eine Population an Zellen, welche mit dem Plasmid IBB transformiert wurde, welchem das AUG vor dem GFP-Gen fehlt. Diese Population an Zellen wurde als Hintergrundwert angenommen.

Der TC-Dimer Riboswitch R4-G8 verfügt über einen 3´ P1-Stamm von 8 Nukleotiden mit der Sequenz: AGGUGACC (Abbildung 3.8.C). Der GC-Gehalt des Stamms liegt bei 63% und der Tm ist etwas niedriger als der Mittelwert der anderen Stämme. Die Basenpaarungen am Anfang und am Ende des Stammes sind stark, wobei die Sequenz mit einem Adenin beginnt und zwei Cytosinen endet. In Runde 4 gibt es nur einen weiteren Schalter mit exakt denselben biophysikalischen Parametern (R4-C5); dieser Schalter reguliert jedoch nur 17-fach und hat die Sequenz: ACUGGGAC.

Ein Grund für den hohen Schaltfaktor des Stammes ist sein niedriger Aus-Zustand in Bezug zur Basalexpression in Anwesenheit von 250 µM TC. Nur zwei weitere Stämme aus Runde drei zeigen einen noch niedrigeren Aus-Zustand (R3-D2 und R3-E4), weisen aber auch einen niedrigen An-Zustand auf, weshalb ihr Schaltfaktor nur bei um die 20-fach liegt.

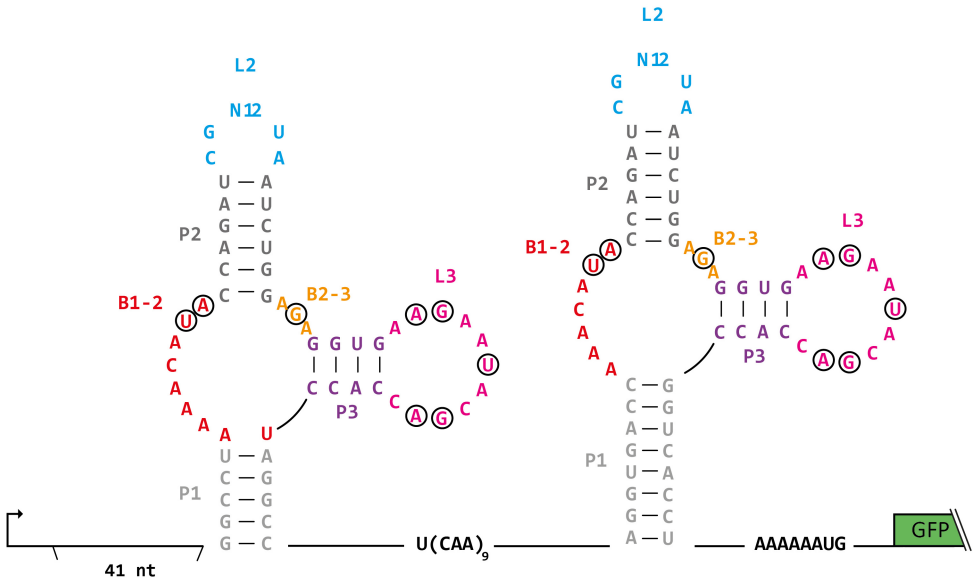


Abbildung 3.8.C 2D-Sekundärstruktur des Dimers R4-G8. Verändert wurde im Vergleich zum Ausgangskonstrukt LG3 nur der Stamm P1 des 3´Aptamers. Dieser umfasst hier 8 Nukleotide und hat die Sequenz AGGUGACC.

3.9 Parameter und Sequenzanalyse der Stämme in Bezug auf ihren Schaltfaktor

Für eine detaillierte Analyse der gewählten Parameter wurden alle Konstrukte mit einer Stammlänge von acht nt in drei Gruppen eingeteilt (siehe Tabelle 3.9) und hinsichtlich der biophysikalischen Parameter, der Basalexpression und der Expression im Aus-Zustand verglichen. Diese Analyse sollte auch die Frage beantworten, warum das Konstrukt R4-G8 besser schaltet als die anderen getesteten Konstrukte. Um eine bessere Vergleichbarkeit zu gewährleisten, ist die folgende Analyse nur auf die Stämme mit einer Länge von acht nt begrenzt.

Tabelle 3.9 Einteilung der Schaltfaktor (SF) in drei Bereiche

	Anzahl	SF - Bereich in x-fach
SF hoch ⇑	52	39,5 - 13,7
SF mittel ⇔	53	13,7 - 11,8
SF niedrig ⇓	53	11,0 - 3,8

In Abbildung 3.9.A wird gezeigt, wie sich die jeweiligen Parameter des P1-Stammes und die Expressionen im An und Aus-Zustand in diesen Gruppen verhalten.

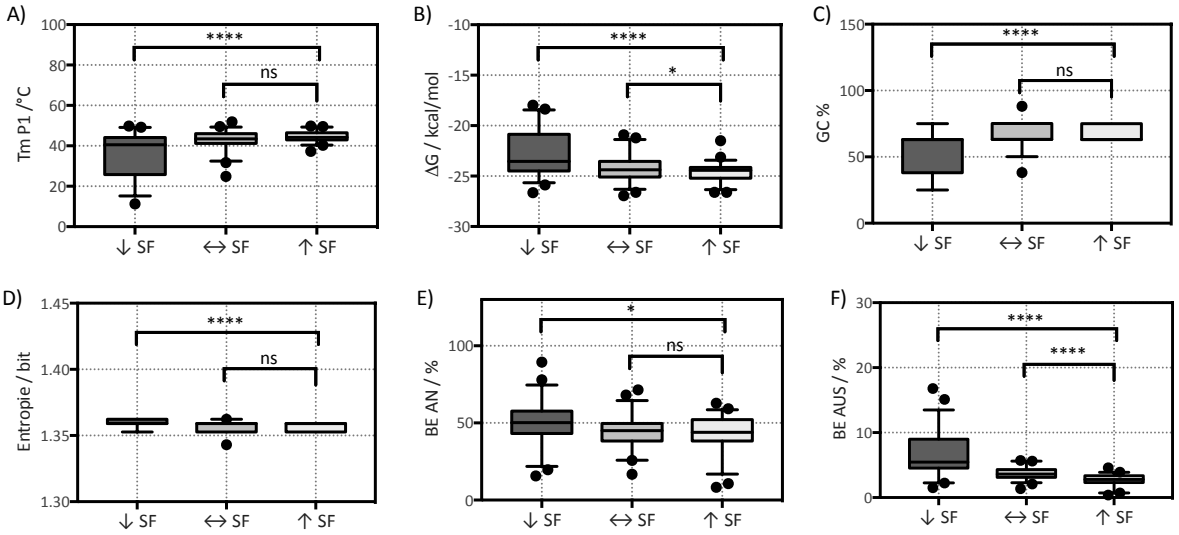


Abbildung 3.9.A Vergleich der biophysikalischen Parameter in Bezug zu niedrig, mittel und hoch schaltenden Riboswitchen. A) Boxplot der biophysikalischen Parameter Schmelztemperatur Tm in °C, **B)** ΔG in kcal/mol , **C)** GC-Gehalt in % und **D)** der Shannon-Entropie in bit, sowie **E)** der Basalexpression (BE) im An-Zustand und **F)** Aus-Zustand. Alle Gruppen (Schaltfaktor (SF) niedrig n = 52, Schaltfaktor mittel n = 54, Schaltfaktor hoch n= 54) wurde die univariate Statistik mittels t-Test berechnet (gepaarter, zweigeteilter t-Test; **** = p value < 0,0001; * = p value > 0,0332). Whisker stellen P10 und P90 dar.

Vergleicht man Stämme mit niedrigem Schaltfaktor in Bezug zu denjenigen mit hohem Schaltfaktor derselben Länge, ist der Unterschied in allen Parametern deutlich signifikant. Sogar die Veränderung der Expression des An-Zustands ist signifikant. Die Veränderung der Parameter der Stämme mit mittleren Schaltfaktor in Bezug auf die Stämme mit hohem Schaltfaktor ist beim ΔG und der Basalexpression signifikant. Besonders auffällig ist die Veränderung der Expression des Aus-Zustandes. Sie ist in ihrer Veränderung sowohl in Bezug auf Schalter mit niedrigem Schaltfaktor als auch in Bezug auf Schalter mit hohem Schaltfaktor signifikant.

Auch für die Sequenzanalyse wurden die Stämme in diese drei Gruppen aufgeteilt. In Bezug auf R4-G8 sollte verglichen werden, ob dieses spezielle Sequenzmotiv und seine Basenabfolge AGGUGACC in höher schaltenden Konstrukten häufiger vorkommt. Einen eindeutigen Trend lässt die Heatmap (Abbildung 3.9.B A)) hier nicht erkennen. Teilweise kommen die Basen an den entsprechenden

Positionen in niedrig schaltenden Konstrukten sogar häufiger vor. Wird der Stamm von R4-G8 jedoch in Sequenzabschnitte unterteilt, lässt sich erkennen, dass die Sequenzteile des Stammes R4-G8 an denselben Positionen bei mehr Stämmen der Gruppe mit einem hohen oder mittleren Schaltfaktor vorkommen, als bei Stämmen mit einem niedrigen Schaltfaktor (Abbildung 3.9.B B)). Der Trend ist deutlich und setzt sich über alle Sequenzabschnitte von R4-G8 fort.

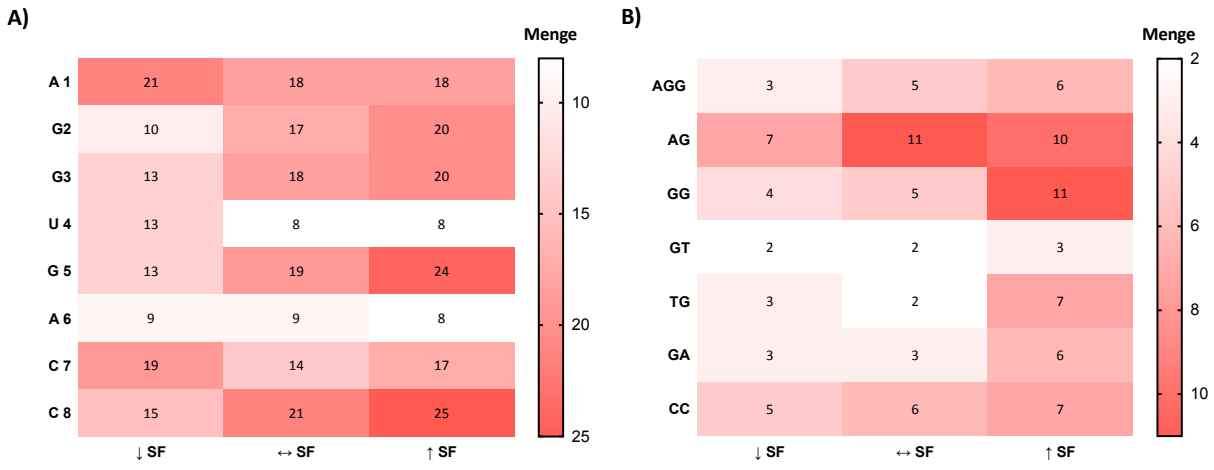


Abbildung 3.9.B: Heatmap zur Häufigkeit der Basen und Sequenzabschnitte im Vergleich A) Heatmap der Häufigkeit der Basen im P1-Stamm in niedrig, mittel und hoch schaltenden Konstrukten. **B)** Heatmap der Sequenzabschnitte im P1-Stamm in niedrig, mittel und hoch schaltenden Konstrukten.

3.10 Derivate von R4-G8

Wie die Daten der Sequenzanalyse zeigten, kommen bestimmte Sequenzabschnitte häufiger im Bereich mit einem höherem Schaltfaktor vor. Daher wurden im nächsten Schritt Mutanten vom Stamm R4-G8 hergestellt, um zu untersuchen, welche Rolle die einzelnen Basen im Stamm spielen. Für jede Base im Stamm gibt es vier Möglichkeiten A, U, C oder G. Für die Derivate wurde jede bestehende Base im Stamm einzeln, sättigend mutagenisiert, während die anderen Basen konstant gehalten wurden. Die Daten der 24 aus der Mutagenese resultierenden Varianten sind in Tabelle 3.10.A zu sehen. Obwohl jeweils nur eine Base in den Stämmen geändert wurde, unterscheiden sich sowohl die biophysikalischen Parameter als auch das Schaltverhalten zum Teil stark vom Ausgangskonstrukt.

Tabelle 3.10.A Derivate von R4-G8 mit dazugehörigen biophysikalischen Parametern und der Expressionen im An- (BE) und Aus-Zustand (Aus) sowie dem Schaltfaktor (SF).

Veränderte Base	Name	Stamm P1 [5´-3´] *	BE [%]	Aus [%]	SF [X-fach]	Tm [°C]	ΔG [kcal/mol]	GC-Gehalt [%]	Entropie [bit]
0	**R4-G8**	**A**GGUGACC	25,2	0,7	**39,5**	43,9	-25,00	63	1,3209
1	**Derivat 1**	**U**GGUGACC	2,2	1,3	**1,5**	44,1	-24,79	63	1,3209
	Derivat 2	**C**GGUGACC	31,5	1,8	**18,6**	46,8	-25,19	75	1,2555
	Derivat 3	G**G**GUGACC	16,5	1,6	**11,2**	50,1	-25,83	75	1,213
2	**Derivat 4**	A**A**GUGACC	38,2	5	**8,1**	33,8	-22,98	50	1,3209
	Derivat 5	A**U**GUGACC	3	2,8	**1,2**	34,7	-23,07	50	1,3863
	Derivat 6	A**C**GUGACC	43,5	4,4	**9,9**	41	-24,22	63	1,3209
3	**Derivat 7**	AG**A**UGACC	2,2	2,4	**0,9**	35	-23,5	50	1,3209
	Derivat 8	AG**U**UGACC	44,9	5,8	**8,1**	33,8	-22,98	50	1,3863
	Derivat 9	AG**C**UGACC	33,8	3	**12,4**	43,5	-25,39	63	1,3209
4	**Derivat 10**	AGG**A**GACC	36,1	4,3	**8,6**	44,1	-25,43	63	1,0822
	Derivat 11	AGG**C**GACC	32,2	2	**16,4**	49,8	-26,36	75	1,0822
	Derivat 12	AGG**G**GACC	8,6	2,1	**4,2**	52,9	-26,75	75	1,0397
5	**Derivat 13**	AGGU**A**ACC	56,7	6	**9,6**	34,8	-22,71	50	1,3209
	Derivat 14	AGGU**U**ACC	49,5	4,2	**12,6**	34,8	-22,71	50	1,3863
	Derivat 15	AGGU**C**ACC	37,1	3,3	**11,4**	43,9	-25	63	1,3209
6	**Derivat 16**	AGGUG**U**CC	30,5	2,9	**10,8**	43,9	-25	63	1,3209
	Derivat 17	AGGUG**C**CC	16,1	2,1	**7,9**	52,3	-26,71	75	1,2555
	Derivat 18	AGGUG**G**CC	62,9	11,3	**5,7**	52,3	-26,71	75	1,213
7	**Derivat 19**	AGGUGA**A**C	51	4,6	**11,3**	33,8	-22,98	50	1,2555
	Derivat 20	AGGUGA**U**C	51,8	3,7	**12,5**	35	-23,5	50	1,3209
	Derivat 21	AGGUGA**G**C	41,2	3,5	**11,9**	43,5	-25,39	63	1,213
8	**Derivat 22**	AGGUGAC**A**	43,6	4,8	**9,4**	37	-23,15	50	1,2555
	Derivat 23	AGGUGAC**U**	40,6	3,9	**10,8**	36,8	-22,27	50	1,3209
	Derivat 24	AGGUGAC**G**	53,6	10,6	**5,3**	40,2	-22,69	63	1,213

* die mutagenisierte Base ist fett markiert, Basen, bei denen die Mutation zu einem AUG Startcodon führt, sind unterstrichen.

In Abbildung 3.10.A ist das Schaltfaktor der einzelnen Derivate im Vergleich zu R4-G8 aufgezeigt. Abbildung 3.10.B zeigt die Basalexpression und den mittleren Schaltfaktor der Mutationsanalyse im Vergleich zum mittleren Schaltfaktor der 4. Runde.

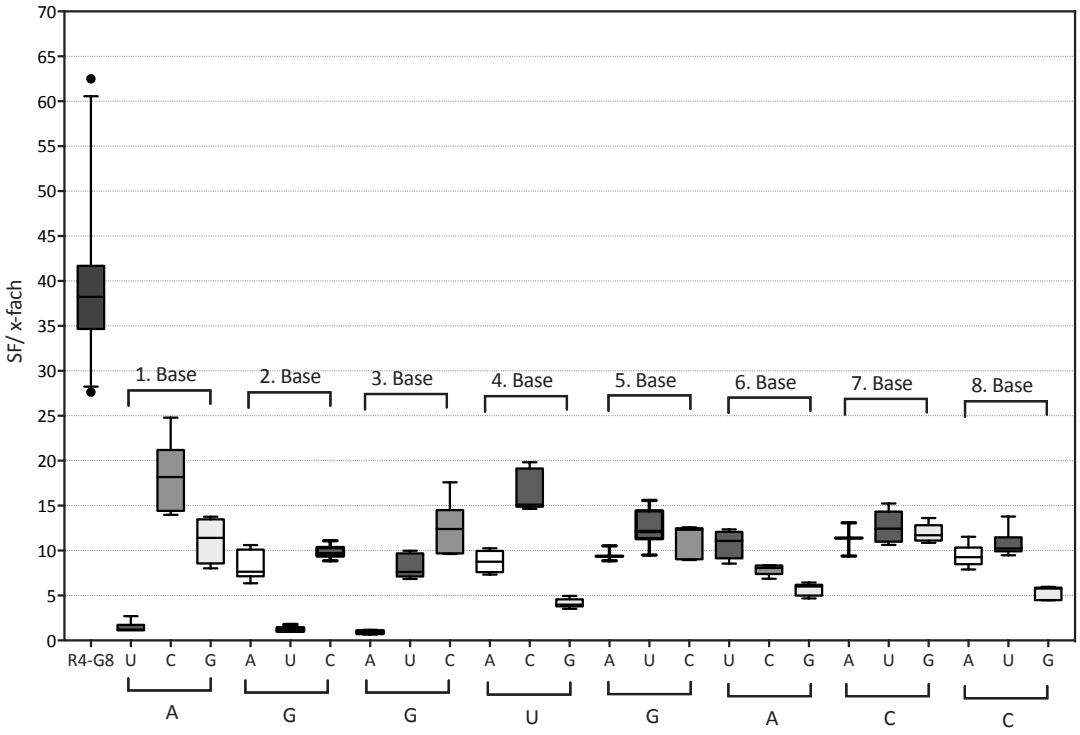


Abbildung 3.10.A Boxplot des Schaltfaktors (SF) der 24 Derivate mit singulären Basenaustausch von R4-G8. Die Base, die mutiert wurde, steht unter dem jeweiligen Boxplot, die eckige Klammer zeigt die Base des ursprünglichen Konstruktes R4-G8 an. Whisker stellen P10 und P90 dar.

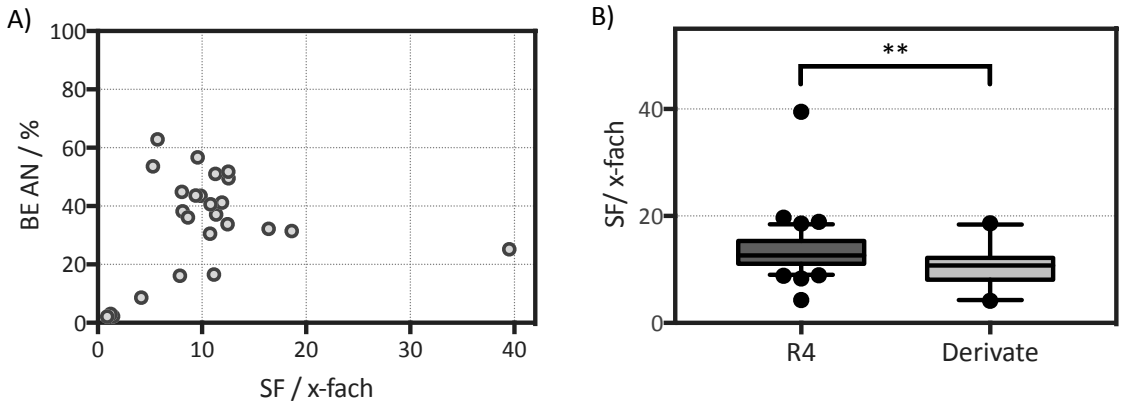


Abbildung 3.10.B Analyse der Derivate A) Basalexpression in % in Bezug auf den Schaltfaktor der Mutationsanalyse (jeder Punkt zeigt den Mittelwert aus 3 unabhängigen Messungen). **B)** Boxplot des Schaltfaktors der 24 Derivate im Vergleich zu Runde 4. Die univariate Statistik wurde mittels t-Test berechnet (gepaarter, zweigeteilter t-Test; ** = p value < 0,01). Whisker stellen P5 und P95 dar.

Einige Konstrukte zeigten keine Schaltaktivität mehr, was auf ein Startcodon (AUG) in der Sequenz zurückzuführen ist (Tabelle 3.10.A, unterstrichene Derivate). Wie bereits in Kapitel 3.3.1 angemerkt, beeinträchtigt ein AUG im P1-Stamm, unabhängig davon ob es *in frame* ist oder nicht, die Expression. Keines der Konstrukte erreicht einen annähernd ähnlichen Schaltfaktor wie das Ausgangskonstrukt.

Auch der mittlere Schaltfaktor der Mutationsanalyse (10,3-fach) bleibt unter dem der Runde 4 (13,3-fach) (die drei Konstrukte mit dem AUG in der Sequenz wurden bei der Berechnung nicht mit einbezogen). Mit der veränderten Sequenz haben sich zum Teil auch die biophysikalischen Parameter des Stammes geändert, was eine Erklärung für die zum Teil sehr unterschiedlichen Schaltfaktoren darstellt. Wie bereits in Kapitel 3.9 aufgezeigt, ist das Schaltverhalten sehr stark mit bestimmten biophysikalischen Parametern verknüpft.

Drei Konstrukte fallen durch ihren sehr geringen Schaltfaktor besonders stark auf: Die Konstrukte Derivat 12, Derivat 18 und Derivat 24. Bei all diesen Derivaten wurde eine Base gegen ein Guanin ausgetauscht. Obwohl bei allen der Schaltfaktor sehr stark negativ beeinflusst wird, kann bei zweien sogar eine viel höhere Basalexpression beobachtet werden im Vergleich zum Ausgangskonstrukt R4-G8. Derivat 12 erreicht einen Schaltfaktor von 4,2-fach und eine Basalexpression von 8,6%. Hier wurde die vierte Base verändert: Ein Uracil wurde gegen ein Guanin ausgetauscht, so dass die Sequenz des Stammes nun AGGGGACC lautet. Derivat 18 hat mit 62,9% eine sehr hohe Basalexpression, erreicht aber nur einen Schaltfaktor von 5,7-fach. Hier wurde die sechste Base, ein Adenin, gegen ein Guanin ausgetauscht, die Sequenz dieses Derivats ist AGGUGGCC. Bei Derivat 24 wurde die achte Base ausgetauscht, so dass die neue Sequenz nun AGGUGACG lautet. Die Basalexpression dieses Konstruktes erreicht 53,6% und der Schaltfaktor 5,3-fach. Wird bei der siebten Base das Cytosin gegen ein Guanin ausgetauscht, hat dies nicht so einen starken Einfluss auf den Schaltfaktor (Derivat 21). Es lässt sich festhalten, dass jede Veränderung von R4-G8 einen starken Einfluss auf das Schaltverhalten hat, aber dass an einigen Positionen dieser Effekt noch stärker wird, insbesondere wenn in ein Guanin ausgetauscht wird.

Der Vergleich des GC-Gehalts und des ΔG-Wertes der Derivate mit diesen Parametern aus Runde 4 zeigt eine breitere Streuung der Parameter bei den Derivaten auf (Abbildung 3.10.C). Es gibt einige Derivate mit einem GC-Gehalt von 50%, welcher nicht als das Optimum gilt (Vergleich GC-Gehalt Runde 1-4).

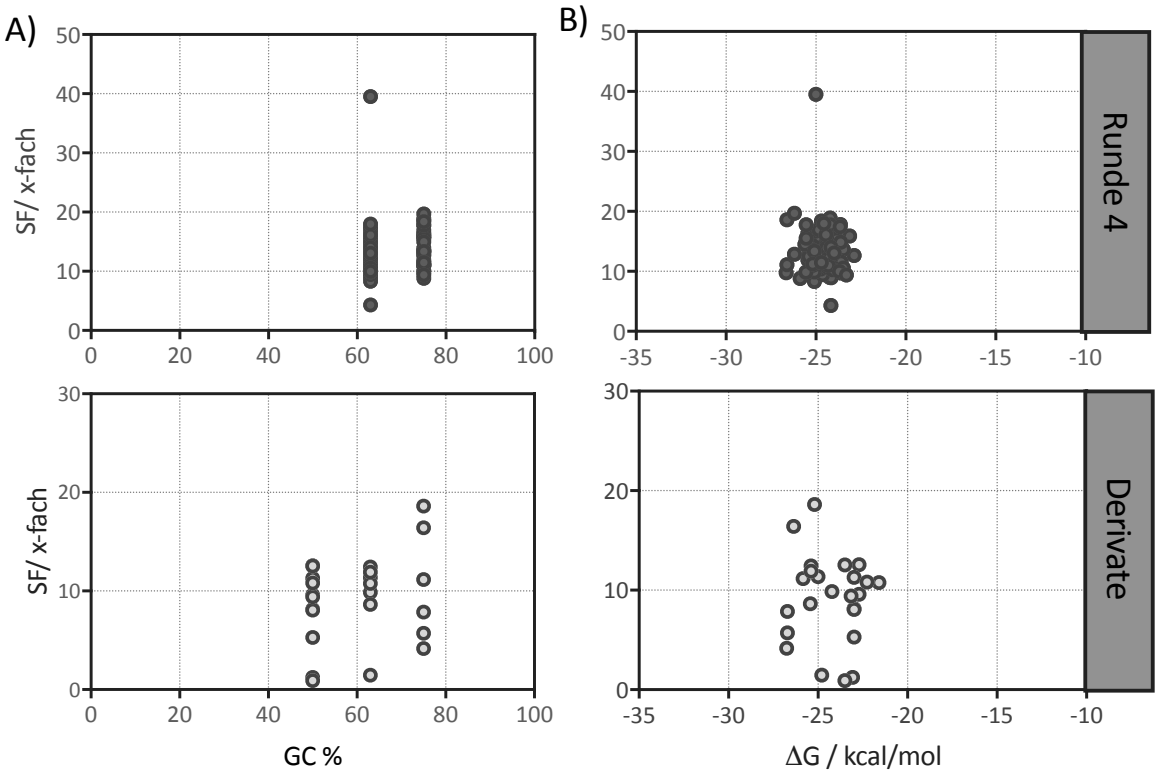


Abbildung 3.10.C GC-Gehalt und ΔG der Mutationsanalyse und Runde 4 im Vergleich. Biophysikalischen Parameter GC% A) und ΔG B) der Mutationsanalyse und Runde 4 in Bezug auf den Schaltfaktor (SF) (jeder Punkt zeigt den Mittelwert aus 3 unabhängigen Messungen).

Die besten Schaltfaktoren werden normalerweise mit einem GC-Gehalt von 63% oder 75% erreicht. Auch die ΔG-Werte der Mutationsanalyse sind etwas weiter gestreut als die der 4. Runde. Abgesehen von den Riboswitchen, welche ein AUG im P1-Stamm enthalten, schaltet auch in der Mutationsanalyse kein Dimer mit einem ΔG um 25 kcal/mol schlechter als 9-fach. Erneut stellt sich dies als ein sehr wichtiger Faktor für ein gutes Schaltverhalten dar. Auch ein GC-Gehalt von 63% scheint vorteilhaft zu sein. Zwar werden bei den Derivaten bessere Schaltfaktoren mit einem GC-Gehalt von 74% erreicht, jedoch sind bei diesem hohen GC-Gehalt auch zwei Dimere dabei, welche ein weniger gutes Schaltverhalten zeigen.

Auch die Shannon-Entropie und der Tm des P1-Stammes weisen in der Mutationsanalyse eine größere Streuung auf als in Runde 4 (Abbildung 3.10.D). Ein Riboswitch aus der Derivaten-Runde mit einer niedrigeren Entropie weißt einen hohen Schaltfaktor auf.

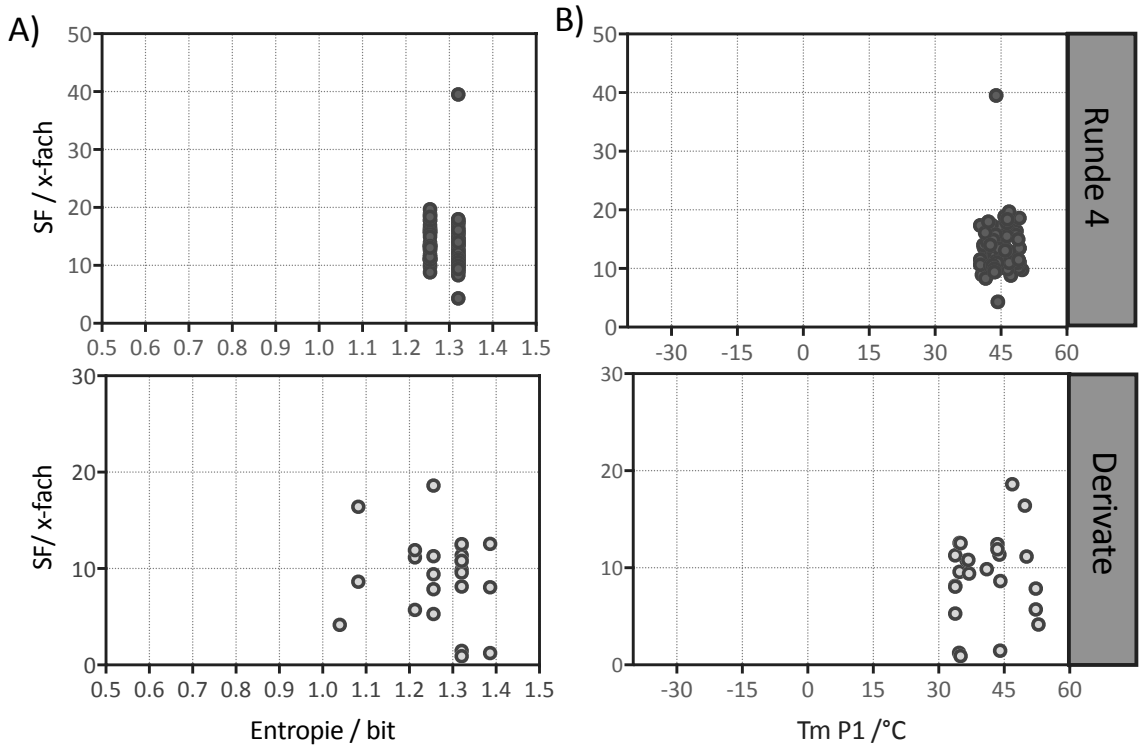


Abbildung 3.10.D: **Shannon-Entropie und Tm der Mutationsanalyse und Runde 4 im Vergleich.** Vergleich der biophysikalischen Parameter Entropie (A) und Tm (B) der Mutationsanalyse und Runde 4 in Bezug auf den Schaltfaktor (SF) (jeder Punkt zeigt den Mittelwert aus 3 unabhängigen Messungen).

Das auch Stämme mit einer niedrigeren Entropie ein gutes Schaltverhalten zeigen, konnte schon in den vorherigen Runden beobachtet werden. Bei dem Tm des P1-Stammes setzt sich wieder ein Stamm mit einem Tm von ca. 45°C durch. Wiederholt werden mit diesem Tm die besten Schalter gefunden, der Trend ist eindeutig und zieht sich durch alle Runden. Zur weiteren Analyse der Derivate werden deshalb in nächsten Schritt nur die P1-Stämme betrachtet, welche in der Mutationsanalyse einen Tm von ca. 45°C aufweisen und kein AUG in der Sequenz haben (Abbildung 3.10.E). Dabei handelt es sich die in Tabelle 3.10.B dargestellten Stämme.

Tabelle 3.10.B Derivate mit einem Tm nahe 45°C; Basalexpression (BE), Expression im Aus-Zustand (Aus) und Schaltfaktor (SF)

Name	Stamm P1 [5´-3´]	BE [%]	Aus [%]	SF [X-fach]	Tm P1 [°C]	ΔG [kcal/mol]	GC-Gehalt [%]	Entropie [bit]
R4_G8	AGGUGACC	25,21	0,66	39,50	**43,92**	**-25**	62,5	1,32089
Derivat 2	**C**GGUGACC	31,46	1,81	18,62	**46,84**	**-25,19**	75	1,35272
Derivat 10	AGG**A**GACC	36,13	4,33	8,64	**44,13**	**-25,43**	63	1,08220
Derivat 15	AGGU**C**ACC	37,15	3,30	11,36	**43,92**	**-25**	63	1,35910
Derivat 16	AGGUG**U**CC	30,5	2,9	10,8	**43,92**	**-25**	63	1,32089

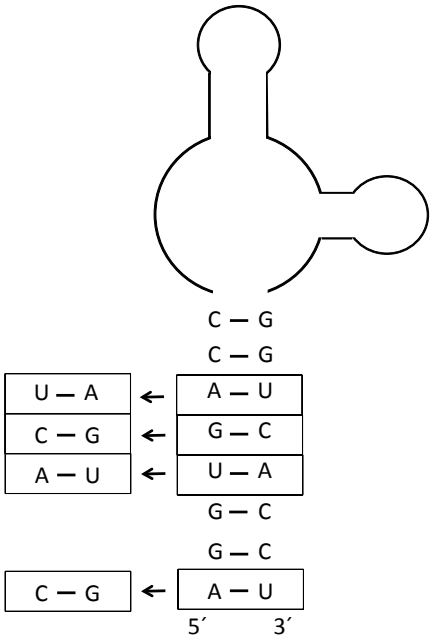


Abbildung 3.10.E Schematische Darstellung der 2D-Struktur mit ausgetauschten Basen der Mutationsanalyse von R4-G8, die zu einem Tm nahe 45°C führen. Wird das erste Basenpaar A-U gegen ein C-G (Derivat 2, Schaltfaktor 18,6-fach) Basenpaar ausgetauscht, das vierte Basenpaar U-A gegen ein A-U (Derivat 10, Schaltfaktor 8,6-fach), das fünfte G-C Basenpaar gegen ein C-G Basenpaar (Derivat 15, Schaltfaktor 11,3-fach) und das sechste A-U Basenpaar gegen ein U-A (Derivat 16, Schaltfaktor 10,8-fach), bleibt der Tm des P1-Stamms in einem ähnlichen Bereich wie der P1-Stamm von R4-G8.

Derivat 2 zeigt hier den höchsten Tm des Stammes P1 und weißt generell den besten Schaltfaktor der Mutationsanalyse auf. Bei diesem Derivat wurde das erste Adenin des Stammes gegen ein Cytosin ausgetauscht. Im Vergleich mit R4-G8 beeinflusst das sowohl den An- als auch den Aus-Zustand. Basalexpression im Liganden-ungebundenen Zustand ist etwas höher und liegt hier bei 31 %. Die Expression des Liganden-gebundenen Zustands des Riboswitches ist mehr als doppelt so hoch und liegt bei etwa 1,8%. Stämme in diesem Tm Bereich weisen in der Mehrheit (ausgenommen Stamm R3-D2) eine Basalexpression zwischen 30 und 55% auf und haben mindestens einen Schaltfaktor von zehn oder höher. Möglicherweise beeinflusst der etwas höhere Tm die Basalexpression in eine höhere Richtung. Ausgenommen R3-D2, weißt Derivat 2 in diesem Tm-Bereich den niedrigsten Aus-Zustand auf.

Es lässt sich also feststellen, dass keines der Derivate von R4-G8 einen vergleichbar guten Schaltfaktor erreicht. Auch die Derivate, die ähnliche biophysikalische Parameter aufweisen, erreichen nicht annähernd diesen guten Schaltfaktor. Eine ähnliche Sequenz mit identischen biophysikalischen Parametern ist keine Garantie für einen hohen Schaltfaktor.

3.10.1 Derivate von R4-G8 mit einer höheren Levenshtein-Distanz

Weitere 18 Derivate von R4-G8 wurden kloniert und gemessen (Tabelle 3.10.1). Diese Derivate unterscheiden sich in den meisten Fällen nur in zwei Nukleotiden vom ursprünglichen Konstrukt. Da die mittleren Sequenzmotive in besser schaltendenden Konstrukten häufiger gefunden wurden als die randständigen Motive (Vergleich Heatmap 3.9.B), wurden hauptsächlich diese Nukleotide durch weitere Mutationen verändert. Es wurde versucht, durch die Veränderung von zwei oder mehr Basen eine starke Veränderung im Expressionsverhalten oder dem Schaltfaktor herbeizuführen und so einen Aufschluss über die Wichtigkeit dieser Basen in der Sequenz zu erhalten. Unter den 18 Derivaten wurde eines entworfen, welches eine spiegelverkehrte Sequenz aufweist (Derivat 26).
Bei allen Derivaten konnte kein Dimer gefunden werden, welches mit dem Schaltfaktor von R4-G8 verglichen werden kann. Obwohl die Konstrukte Derivat 36 und Derivat 37 exakt die gleichen biophysikalischen Parameter aufweisen wie R4-G8, erreichen sie keinen hohen Schaltfaktor.

Tabelle 3.10.1 Derivate und ihre biophysikalischen Parameter mit einer Levenshtein-Distanz (LD) größer 1; Basalexpression (BE), Expression im Aus-Zustand (Aus) und Schaltfaktor (SF).

Name	Stamm P1 [5´-3]	BE [%]	Aus [%]	SF [X-fach]	Tm [°C]	ΔG [kcal/mol]	GC-Gehalt [%]	Entropie [bit]	LD	Nr. veränderter Basen
R4_G8	AGGUGACC	25,21	0,66	**39,5**	43,92	-25	62,5	1,32089	0	0
Derivat 25	**CC**GUGAA**G**	53	6,15	**9,01**	36,38	-21,6	63	1,3209	3	1. 2. 8.
Derivat 26	**CCAGUGGA**	45,87	5,11	**9,09**	43,28	-23,76	63	1,3209	8	1. -8.
Derivat 27	AGG**CA**ACC	34,26	2,82	**12,14**	43,01	-24,69	63	1,0822	2	4. 5.
Derivat 28	AGG**CUGA**C	33,02	2,68	**12,35**	43,47	-25,39	63	1,3209	4	4. 5. 6. 7
Derivat 29	AGGU**UG**CC	41,42	3,06	**13,97**	43,01	-24,69	63	1,3209	2	5. 6.
Derivat 30	AGGUGA**GG**	32,23	3,13	**10,99**	43,13	-23,47	63	0,9003	2	7. 8.
Derivat 31	A**C**GU**C**ACC	49,72	4,32	**12,28**	41,02	-24,22	63	1,213	2	2. 5.
Derivat 32	AGG**A**GA**G**C	29,7	3,02	**9,99**	43,68	-25,82	63	0,9743	2	4. 7.
Derivat 33	AGG**AGU**CC	32,45	2,93	**11,53**	44,13	-25,43	63	1,3209	2	4. 6.
Derivat 34	**UC**GUGACC	43,88	4,12	**11,64**	41,41	-24,44	63	1,3209	2	1. 2.
Derivat 35	AGG**CU**ACC	27,62	2,16	**13,23**	44,64	-25,12	63	1,3209	2	4. 5.
Derivat 36	AG**UGAC**CC	29,02	3,02	**9,73**	43,92	-25	63	1,3209	4	3. 4. 5. 6.
Derivat 37	AG**UG**GACC	36,28	2,96	**12,35**	43,92	-25	63	1,3209	2	3. 4.
Derivat 38	AGGUG**GA**C	36,28	3,25	**11,14**	43,92	-25	63	1,213	2	6. 7.
Derivat 39	AGGUG**GG**C	20,53	2,27	**9,01**	52,26	-26,71	75	1,0735	2	6. 7.
Derivat 40	AGG**GGU**CC	10,35	2,63	**3,97**	52,88	-26,75	75	1,213	2	4. 6.
Derivat 41	AGG**AC**ACC	29,99	3,02	**9,92**	43,92	-25	63	1,0822	2	4. 5.
Derivat 42	A**CC**UGACC	34,42	3,17	**10,93**	43,92	-25	63	1,213	2	2. 3.

Die meisten Schaltfaktoren dieser zweiten Mutationsanalyse befinden sich im mittleren Bereich. Den höchsten Schaltfaktor erreicht Derivat 29. Hier wurden die fünfte und die sechste Base verändert. Die Veränderung der fünften Base von einem Guanin zu einem Uracil hatte in der ersten

Mutationsanalyse einen Schaltfaktor von 12,6-fach erzielt. Die Mutation der sechsten Base von einem Adenin zu einem Guanin führte in der ersten Mutationsanalyse sogar einen sehr niedrigen Schaltfaktor von 5,7-fach. Zusammen konnten diese beiden Veränderungen den Schaltfaktor nun wieder etwas anheben und im Vergleich zu den anderen Konstrukten dieser zweiten Mutationsanalyse sogar verbessern. Die nachfolgenden Abbildungen (3.10.1 A und 3.10.1 B) zeigen die graphische Darstellung der Ergebnisse.

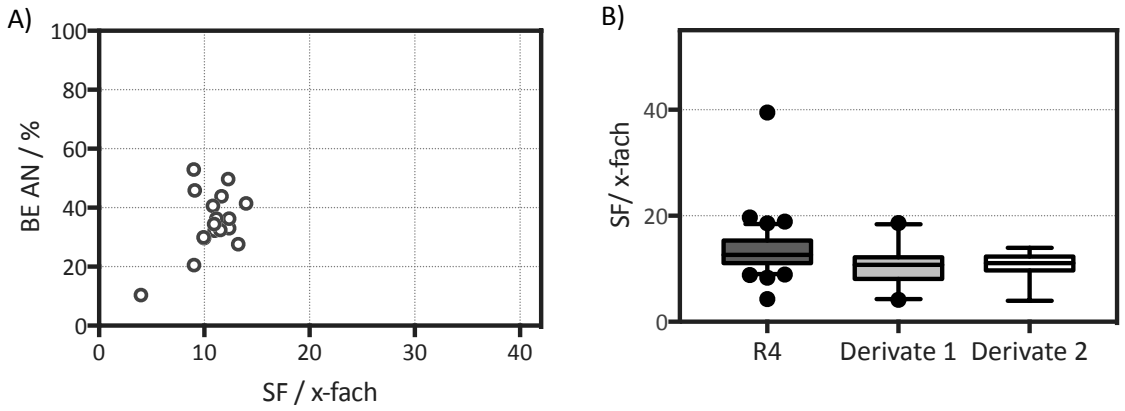


Abbildung 3.10.1 A Ergebnisse der Messungen der 2. Mutationsanalyse von R4-G8, Schaltfaktoren der zweiten (Derivate 2) im Vergleich zur ersten Mutationsanalyse (Derivate 1) und der Runde 4 A) Basalexpression (BE) in Bezug auf den Schaltfaktor (SF) von Derivate mit einer Levenshtein-Distanz größer eins (jeder Punkt zeigt den Mittelwert aus 3 unabhängigen Messungen). **B)** Boxplot des Schaltfaktors (SF) der 24 Derivate im Vergleich zu Runde 4. Whisker stellen P5 und P95 dar.

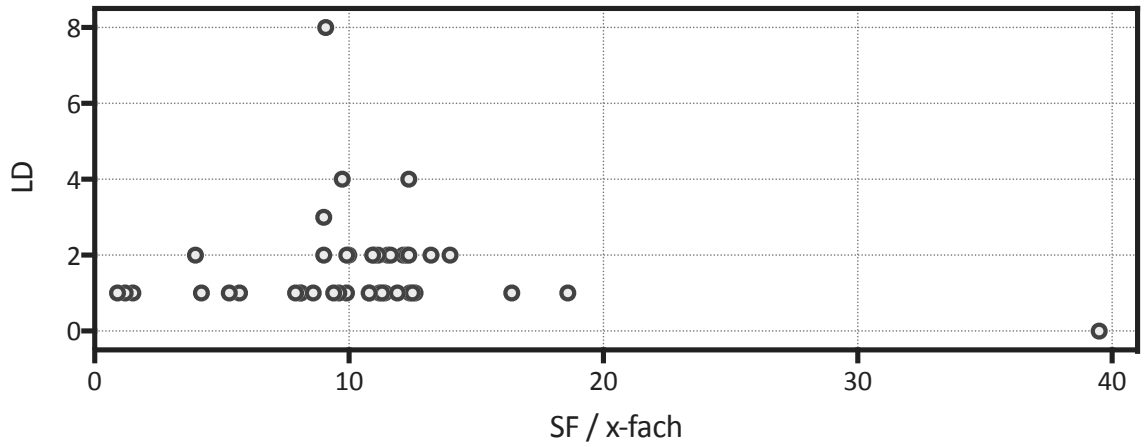


Abbildung 3.10.1 B: Levenshtein-Distanz (LD) der Derivate in Bezug auf den Schaltfaktor Schaltfaktor (SF) (jeder Punkt zeigt den Mittelwert aus 3 unabhängigen Messungen).

Die Daten zeigen, dass schon kleine Veränderungen der Sequenz ausreichen, um den Schaltfaktor von G8-R4 negativ zu beeinflussen. Zum einen mag das an der damit einhergehenden Veränderung der biophysikalischen Parameter liegen. Jedoch zeigt auch Derivat 37 mit den gleichen Parametern

und einer recht ähnlichen Sequenz ein schlechteres Schaltverhalten. Die Sequenzabschnitte GG (Position zwei und drei) und UG (Position vier und fünf) sind wesentlich häufiger in Schaltern mit einem hohen Schaltfaktor zu finden als in Konstrukten mit einem niedrigen Schaltfaktor (Vergleich Kapitel 3.9), diese beiden Sequenzabschnitte kommen im Konstrukt Derivat 37 nicht vor.

3.11 Der Einfluss der Stammendung auf die Basalexpression

Der Stamm des 3´Aptamers liegt unmittelbar vor der Kozak-Sequenz. Dass dieser Bereich die Translationseffizienz beeinflussen kann, wurde bereits gezeigt (J. Li et al. 2017). Insbesondere die Nukleotide im Bereich -11 bis -14 hatten in der Studie von Li *et al* einen besonders starken Einfluss auf die Translationseffizienz. In genau diesem Bereich befindet sich das 5´Ende des P1-Stammes. Die Stammendungen CC und GG kommen öfter in den Stämmen mit einem höheren Schaltfaktor vor. Zudem zeigt sich, dass eine Basalexpression von um die 30% oft sehr gute Schaltfaktor erzeugt (siehe Tabelle 8.5 Anhang). Daher wird in diesem Abschnitt untersucht, welchen Einfluss die beiden letzten Nukleotide auf die Basalexpression im Liganden-ungebundenen Zustand haben. Da der Tm des P1-Stammes einen starken Einfluss auf die Basalexpression hat (auf diesen Zusammenhang wird im folgenden Kapitel 3.12 eingegangen werden), wurden für diesen Vergleich alle gemessenen Stämme mit einer Schmelztemperatur von 40°C und 50°C, einer Stammlänge von acht nt und einer Stammendung wie in Abbildung 3.11.A dargestellt ausgewählt. Mit einbezogen wurden auch die Ergebnisse der beiden Mutationsanalysen.

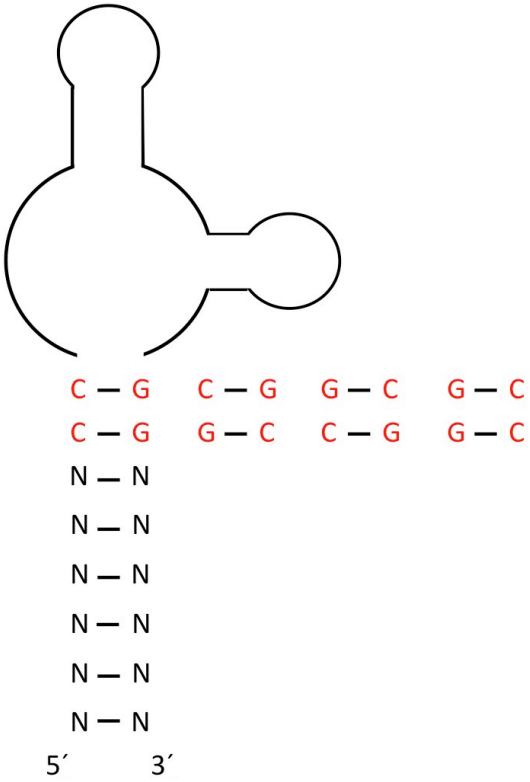


Abbildung 3.11.A Schematische 2D-Struktur des TC-Dimers. Alle Stämme mit einer Stammendung CC; CG; CG und GG mit acht nt Länge und einem Tm zwischen 40°C und 50°C wurden in die Analyse mit einbezogen.

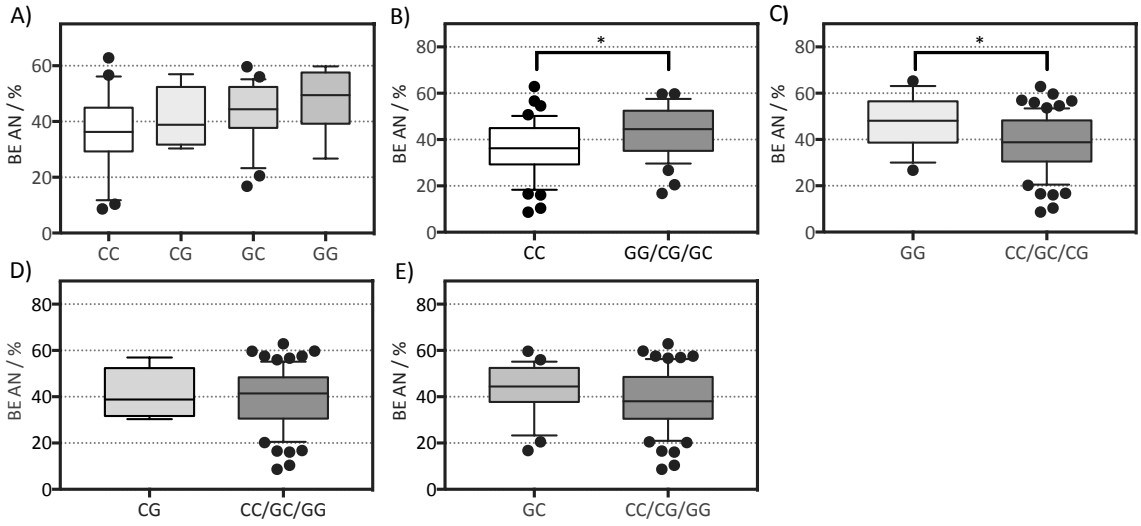


Abbildung 3.11.B A) Boxplot der Basalexpression (BE) der Stämme mit den Stammendungen CC/ CG/ GC und GG. Whisker stellen P5 und P95 dar. **B)** Boxplot der Basalexpression (BE) der Stämme mit den Stammendungen CC im Vergleich zu den Endungen GG/CG und GC. Die univariate Statistik wurde mittels t-Test berechnet (gepaarter, zweigeteilter t-Test; * = p value < 0,0332). Whisker stellen P5 und P95 dar. **C)** Boxplot der Basalexpression (BE) der Stämme mit den Stammendungen GG im Vergleich zu den Endungen CC/GC und CG. Die univariate Statistik wurde auf Grund der unterschiedlich großen Stichproben mittels Welch-Test berechnet (gepaarter, zweigeteilter t-Test; * = p value < 0,0332). Whisker stellen P5 und P95 dar. **D)** Boxplot der Basalexpression (BE) der Stämme mit den Stammendungen CG im Vergleich zu den Endungen CC/GC und GG. Whisker stellen P5 und P95 dar. **E)** Boxplot der Basalexpression (BE) der Stämme mit den Stammendungen GC im Vergleich zu den Endungen CC/GC und GG. Whisker stellen P5 und P95 dar.

Abbildung 3.11.B zeigt die Basalexpression der Konstrukte mit den Stammendungen CC, CG, GC und GG. Man kann erkennen, dass die Endung CC eher bei Konstrukten mit einer etwas niedrigeren Basalexpression zu finden ist und die Endung GG eher bei Konstrukten mit einer etwas höheren. Vergleicht man die einzelnen Endungen gegen die Gesamtheit der anderen drei Endungen kann man feststellen, dass die Endung CC zu einer signifikanten Verringerung der Basalexpression führt. Für den Vergleich der restlichen Endungen konnte mit dem gepaarten t-Test keine Signifikanz ermittelt werden. Da sich die Stichproben-Häufigkeit bei GG, CG und GC jedoch zum Teil stark unterschieden, wurde hier auch der Welch-T-Test durchgeführt. Mit diesem konnte für die Endung GG verglichen mit dem Rest eine signifikante Erhöhung der Basalexpression ermittelt werden.

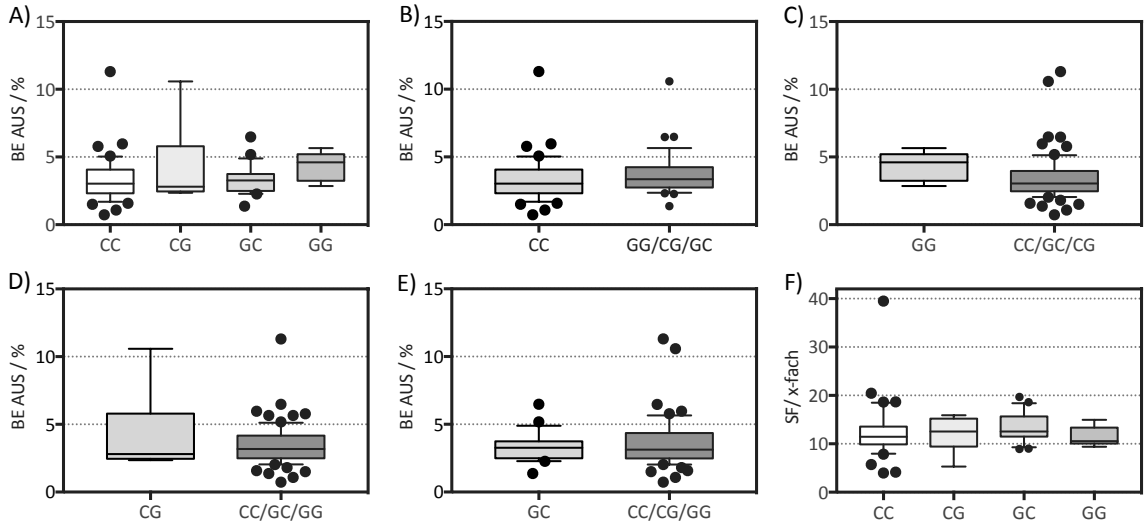


Abbildung 3.11.C: A) Boxplot der Expression (BE) im Aus-Zustand der Stämme mit den Stammendungen CC/ CG/ GC und GG. Whisker stellen P5 und P95 dar. **B)** Boxplot der Expression (BE) im Aus-Zustand der Stämme mit den Stammendungen CC im Vergleich zu den Endungen GG/CG und GC. Whisker stellen P5 und P95 dar. **C)** Boxplot der Expression (BE) im Aus-Zustand der Stämme mit den Stammendungen GG im Vergleich zu den Endungen CC/CG und GC. Whisker stellen P5 und P95 dar. **D)** Boxplot der Expression (BE) im Aus-Zustand der Stämme mit den Stammendungen CG im Vergleich zu den Endungen CC/GC und GG. Whisker stellen P5 und P95 dar. **E)** Boxplot der Expression (BE) im Aus-Zustand der Stämme mit den Stammendungen GC im Vergleich zu den Endungen CC/CG und GG. Whisker stellen P5 und P95 dar. **F)** Boxplot des Schaltfaktors (SF) der Stämme mit den Stammendungen CC/ CG/ GC und GG. Whisker stellen P5 und P95 dar.

Für den Aus-Zustand und den Schaltfaktor lässt sich jedoch keine durch die Endung bedingte signifikante Änderung feststellen. Jedoch fällt bei der reinen Betrachtung der Häufigkeit auf, dass die Endung C in besser schaltenden Konstrukten häufiger vorkommt, als die Endung G (Abbildung 3.11.D).

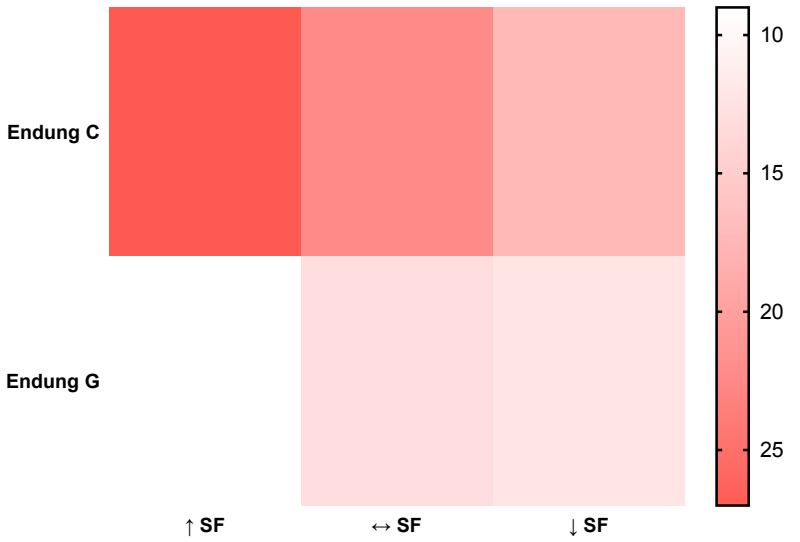


Abbildung 3.11.D Heatmap. Häufigkeit der 3´ Endungen des P1-Stammes in Bezug auf den Schaltfaktor.

Die Endung CC verändert die Basalexpression von GFP geringfügig hin zu einer niedrigeren Basalexpression, wohingegen die Endungen CC, GG, CG und GC keinen Einfluss auf die Expression im Liganden-gebundenen Zustand und den Schaltfaktor haben. Jedoch kommt ein Cytosin in gut schaltenden Konstrukten am Ende des P1-Stammes wesentlich häufiger vor als ein Guanin, wie bereits in Kapitel 3.9 aufgearbeitet wurde (Abbildung 3.9.B).

3.12 Biophysikalische Parameter in Bezug auf Expression und Schaltfaktoren

Die bisherigen Analysen der Parameter, Runden und Schaltfaktoren sowie Tabelle 8.5 im Anhang, welche alle gemessenen Konstrukte dieser Arbeit enthält, zeigen einen gewissen Trend der Dimere in Bezug auf ihre biophysikalischen Parameter. Um zu verdeutlichen, in welchem Maße die biophysikalischen Parameter die Basalexpression im An- und Aus-Zustand beeinflussen, wurde Grafik 3.12.A erstellt und die Korrelation nach Pearson berechnet. Aufgetragen sind die einzelnen biophysikalischen Parameter aller gemessenen Dimere in Bezug auf die Expression im Liganden-ungebundenen Zustand (An) und -gebundenen Zustand (Aus). Besonders stark korrelieren der Tm und der ΔG mit der Expression und es fällt auf, dass diese Korrelation im Aus-Zustand des Schalters noch stärker ist als im An-Zustand. Generell ist diese Korrelation beim GC-Gehalt schon etwas weniger deutlich, ist jedoch auch hier wieder im Aus-Zustand geringfügig stärker ausgeprägt als im An-Zustand. Nicht sehr ausgeprägt ist die Korrelation der Expression in Bezug auf die Entropie, was sowohl auf den An-Zustand als auch auf den Aus-Zustand zu trifft.

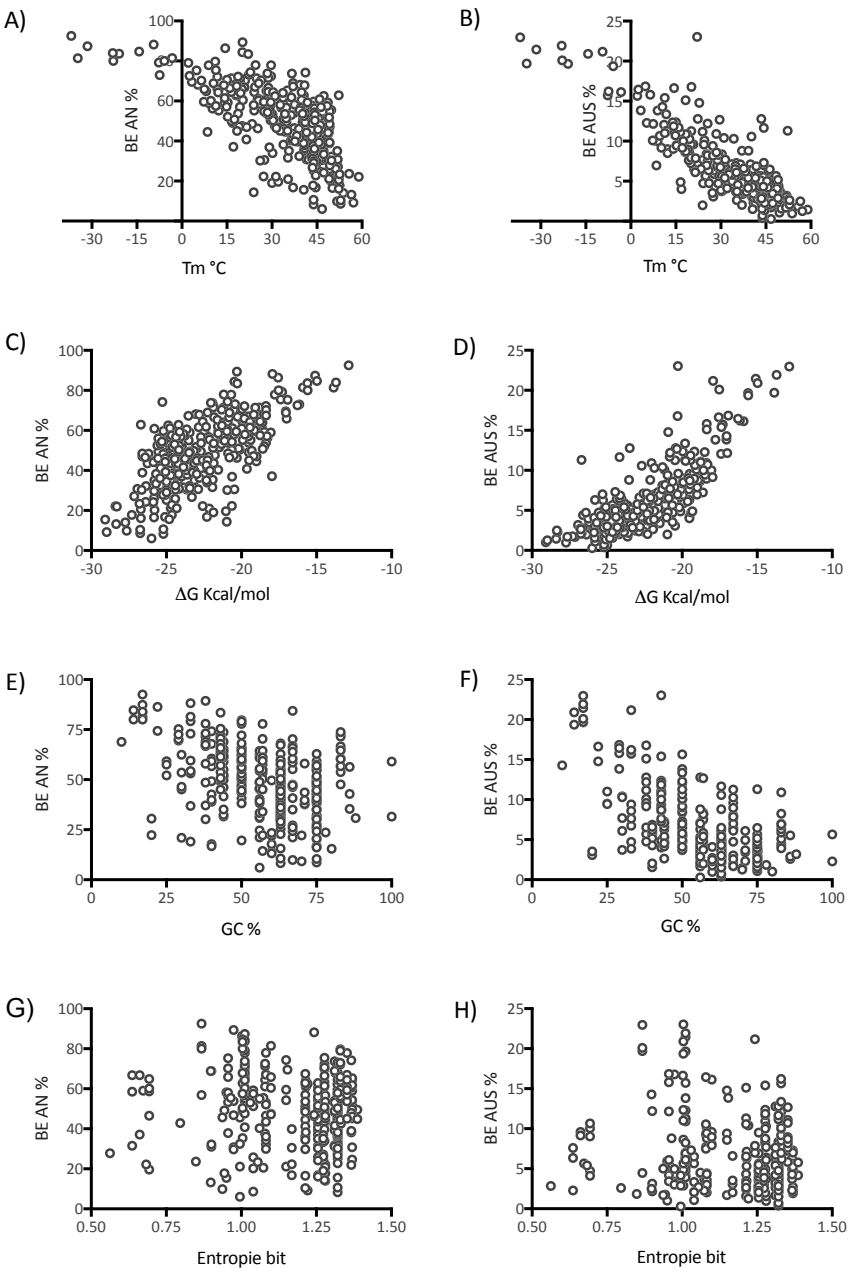


Abbildung 3.12.A: **Biophysikalische Parameter in Bezug auf die Expression im An- und Aus-Zustand der Dimere. A)** Basalexpression im An-Zustand in % in Bezug auf den Tm in °C des P1-Stamms. Korrelationsanalyse: Pearson r = -0,6844, R^2 = 0,4684, p value < 0,0001. **B)** Basalexpression im Aus-Zustand in % in Bezug auf den Tm in °C des P1-Stamms. Korrelationsanalyse: Pearson r = -0,8675, R^2 = 0,7525, p value < 0,0001. **C)** Basalexpression im An-Zustand in % in Bezug auf den ΔG in kcal/mol. Korrelationsanalyse: Pearson r = -0,6891, R^2 = 0,4749, p value < 0,0001. **D)** Basalexpression im Aus-Zustand in % in Bezug auf den ΔG in kcal/mol. Korrelationsanalyse: Pearson r = -0,8121, R^2 = 0,6594, p value < 0,0001. **E)** Basalexpression im An-Zustand in % in Bezug auf den GC-Gehalt in % des P1-Stamms. Korrelationsanalyse: Pearson r = -0,4214, R^2 = 0,1776, p value < 0,0001. **F)** Basalexpression im Aus-Zustand in % in Bezug auf den GC-Gehalt in % des P1-Stamms. Korrelationsanalyse: Pearson r = -0,6315, R^2 = 0,3988, p value < 0,0001. **G)** Basalexpression im An-Zustand in % in Bezug auf die Entropie des P1-Stamms in bit. Korrelationsanalyse: Pearson r = -0,119, R^2 = 0,01417, p value < 0,0185. **H)** Basalexpression im Aus-Zustand in % in Bezug auf die Entropie des P1-Stamms in bit. Korrelationsanalyse: Pearson r = -0,2446, R^2 = 0,05983, p value < 0,0001.

Da eine Korrelation zwischen der Expression sehr deutlich ist, ist auch eine Korrelation der biophysikalischen Parameter zum Schaltfaktor naheliegend. Um diese aufzuzeigen wurden vier verschiedene Heatmaps erstellt (Abbildung 3.12.B). Jede Heatmap zeigt die Häufigkeit, mit der ein Schaltfaktor bei einem gewissen Parameter vorkommt.

Auch hier lässt sich eine Korrelation zwischen Tm bzw. ΔG in Bezug auf den Schaltfaktor erkennen. Je höher der Tm und je niedriger der ΔG, desto höher wird die Wahrscheinlichkeit, einen hohen Schaltfaktor zu erhalten. Dieser Trend zeigt sich weniger deutlich bei der Entropie. Hier ist zwar auch eine hohe Entropie von vorteilhaft, jedoch nicht zwangsläufig notwendig. Auch der GC-Gehalt lässt einen recht deutlichen Trend erkennen. Die höchsten Schaltfaktoren werden mit einem GC-Gehalt zwischen 60% und 75% erreicht. Mit einem niedrigeren GC-Gehalt konnte kein Schaltfaktor über 14-fach erreicht werden.

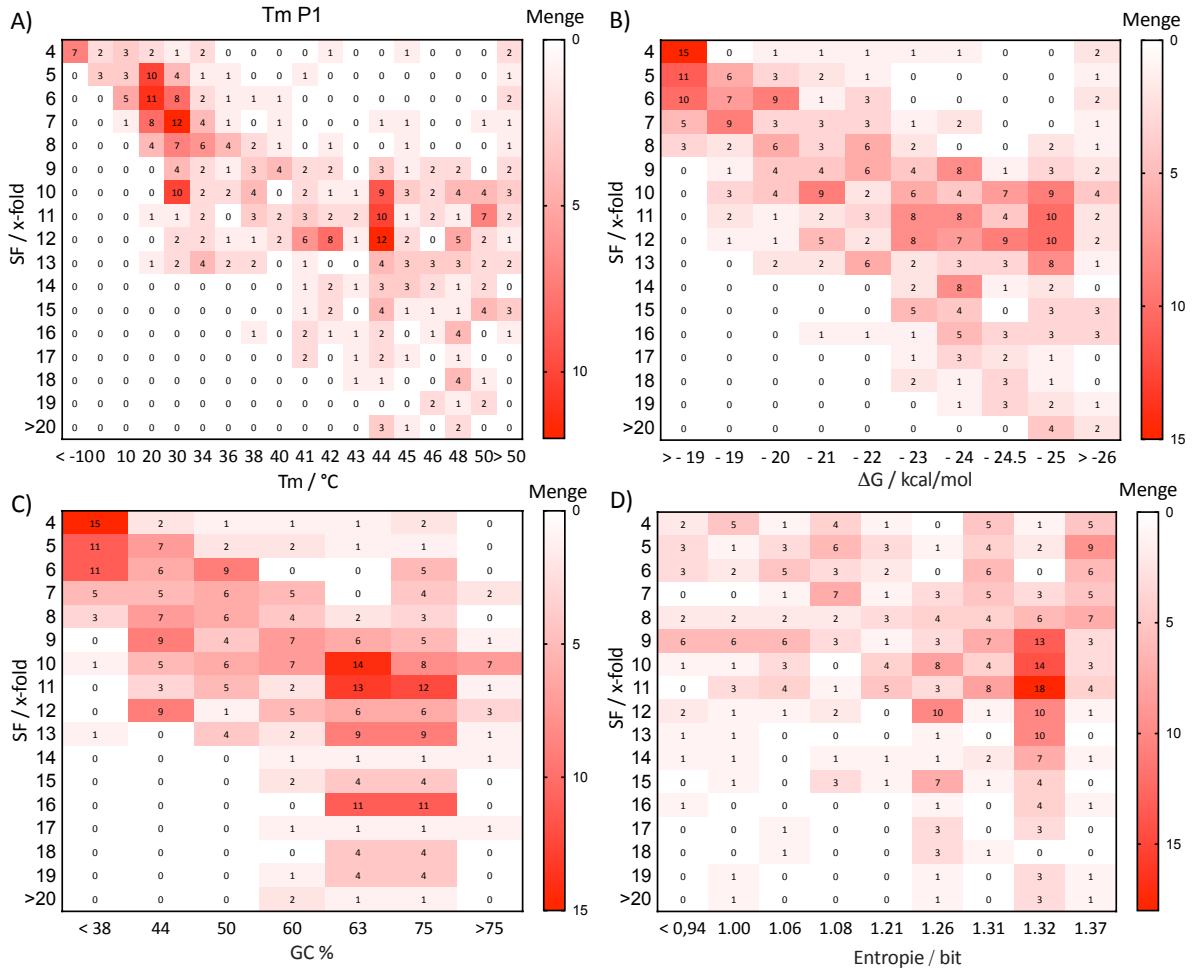


Abbildung 3.12.B Heatmaps der biophysikalischen Parameter (x-Achse) in Bezug auf die Häufigkeit bestimmter Schaltfaktor (SF)-Klassen (y-Achse). **A)** Der Tm des P1-Stammes in °C in Bezug auf die Häufigkeit der Schaltfaktoren (SF). **B)** Der ΔG des P1-Stammes in kcal/mol in Bezug auf die Häufigkeit der Schaltfaktoren (SF). **C)** Der GC-Gehalt des P1-Stammes in % in Bezug auf die Häufigkeit der Schaltfaktoren (SF). **D)** Die Entropie des P1-Stammes in bit in Bezug auf die Häufigkeit der Schaltfaktoren (SF).

Folgende Aussagen bezüglich der biophysikalischen Parameter des TC-Dimers können somit getroffen werden:

1. Befinden sich die biophysikalische Parameter Tm und ΔG in einem gewissen Bereich, erhöht sich die Wahrscheinlichkeit, einen sehr gut schaltenden Riboswitch zu finden.
    - Ein Tm des P1-Stamms zwischen 44°C und 50°C
    - Ein ΔG zwischen -24,5 kcal/mol und -26 kcal/mol
2. Biophysikalische Parameter innerhalb eines bestimmten Bereiches können ein Ausschlusskriterium für einen sehr hoch schaltenden Riboswitch sein und erhöhen die Wahrscheinlichkeit eines schlecht schaltenden Riboswitches:
    - Ein Tm des P1-Stamms kleiner 36°C
    - Ein ΔG größer -20 kcal/mol
    - Ein GC-Gehalt des P1-Stamms kleiner 60%
3. Die Entropie des P1-Stamms hat den geringsten Einfluss auf den Schaltfaktor
4. In Runde 3 und Runde 4 nähern sich die mittleren biophysikalischen Parameter der in diesen Runden gemessenen Riboswitche immer stärker den Bereichen an, die einen sehr gut schaltenden Riboswitch begünstigen.

3.13 Der Austausch des 5´Aptamers und der Effekt auf den Schaltfaktor

Im Weiteren sollte untersucht werden, ob ein Austausch des 5´Aptamers gegen ein anderes Aptamer den Schaltfaktor beeinflussen und verbessern kann. Um dies zu ermöglichen, wurden die ersten acht Basen des CAA-Spacer zwischen den Aptameren durch eine PacI-Schnittstelle ersetzt (Abbildung 3.13.A). Diese Schnittstelle ermöglichte den Austausch der Aptamere und beeinflusste die Faltung nicht.

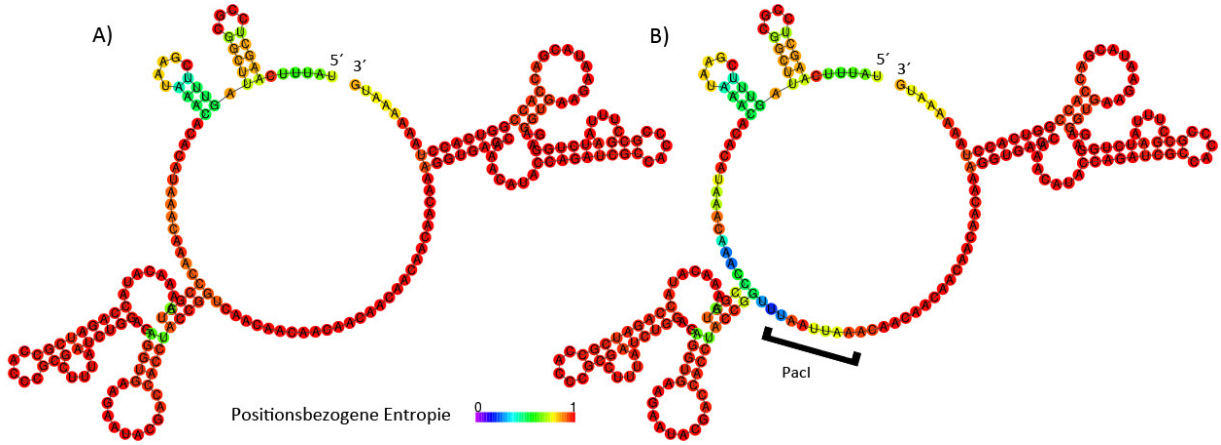


Abbildung 3.13.A: Faltungsvorhersage des 5´UTRs mit R4-G8 mittels RNA-fold mit und ohne PacI-Schnittstelle. A) Sekundärstruktur-Vorhersage mit RNA-Fold für das 5´UTR mit R4-G8 ohne PacI-Schnittstelle ab Transkriptionsstartpunkt bis zum Startcodon **B)** Sekundärstruktur-Vorhersage mit RNA-Fold für das 5´UTR mit R4-G8 mit PacI-Schnittstelle im CAA-Spacer ab Transkriptionsstartpunkt bis zum Startcodon.

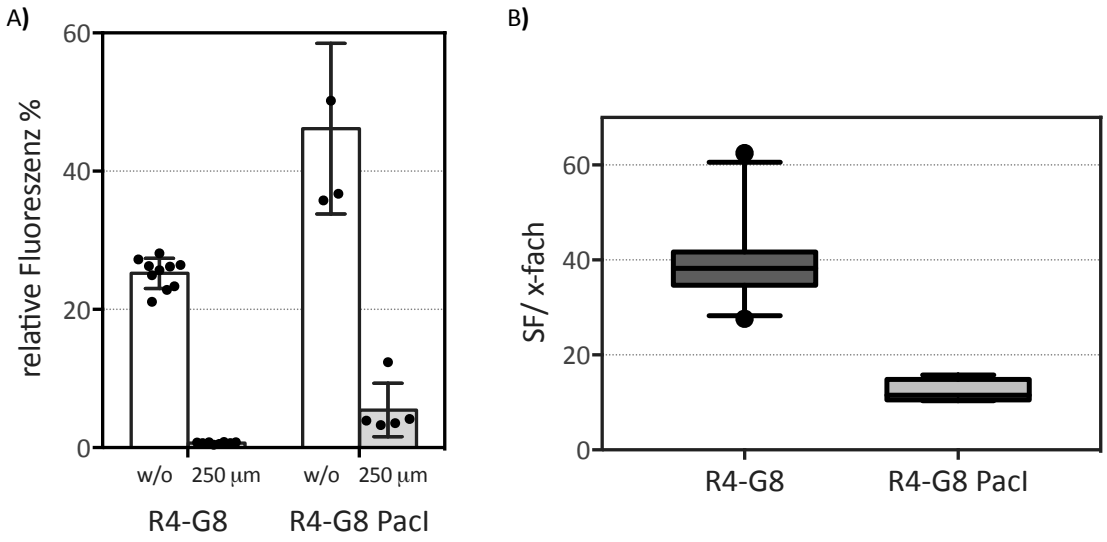


Abbildung 3.13.B Schaltleistung von R4-G8 und R4-G8 PacI im Vergleich A) Vergleich der relativen Fluoreszenz von R4-G8 und R4-G8 PacI ohne (w/o) und mit 250 µM Tetrazyklin. **B)** Boxplot: Vergleich des Schaltfaktors (SF) der beiden Konstrukte. Whisker stellen P1 und P90 dar.

Der durch die Schnittstelle geringfügig veränderte Kontext des Riboswitches zeigt einen wesentlich niedrigeren Schaltfaktor und eine höhere Basalexpression als das Ausgangskonstrukt R4-G8 (Abbildung 3.13.B).

Trotz dieser Unterschiede im Expressionsmuster wurde das Konstrukt mit der PacI-Schnittstelle für weitere Klonierungen genutzt. Im darauf folgenden Schritt wurden mehrere Konstrukte getestet, die aus zwei Aptameren bestanden: Ein Konstrukt, welches 2x das 5´Aptamer (R4-G8/R4-G8) enthielt, ein Konstrukt bei dem das 3´Aptamer gegen das 5´Aptamer ausgetauscht wurde (R4-G8/Tc1) und drei

Konstrukte, bei welchen das 3´Aptamer gegen ein andere bereits getestete Konstrukte ausgetauscht wurde (R4-E8/R4-G8, R2-B7/R4-G8, R4-F5/R4-G8). Abbildung 3.13.C zeigt die Faltungsvorhersage für diese Konstrukte.

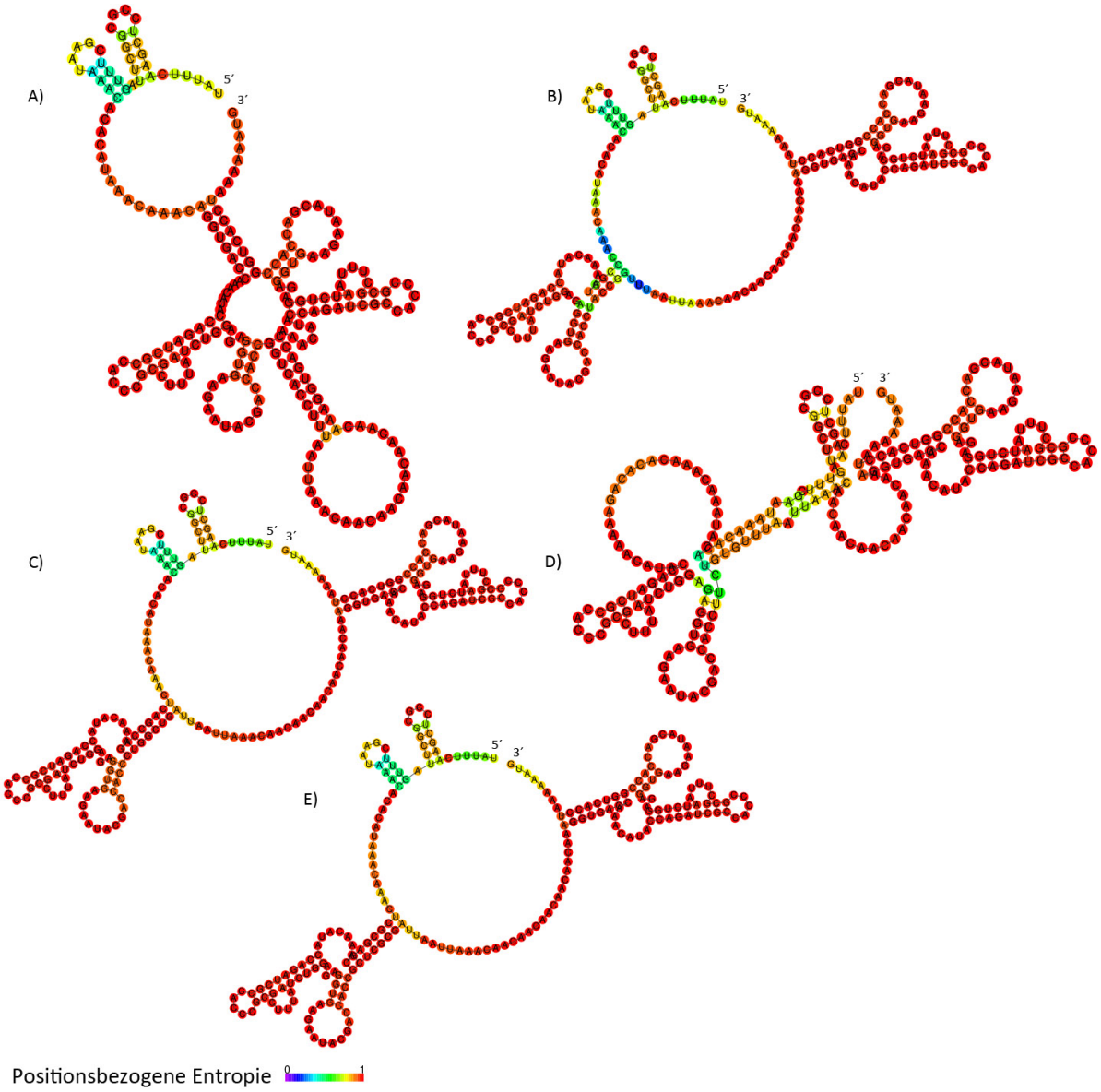


Abbildung 3.13.C: Sekundärstrukturvorhersage TC-Dimere mit ausgetauschtem 3´TC-Dimer des Servers RNA-Fold. 5´UTR beginnt am Transkriptionsstartpunkt und endet mit dem Startcodon. **A)** R4-G8/R4-G8, **B)** R4-G8/Tc1 **C)** R4-E8/R4-G8 **D)** R2-B7/R4-G8 und **E)** R4-F5/R4-G8.

Für die Konstrukte R4-B8/R4-G8 und R2-B7/R4-G8 wird von RNAfold eine Fehlfaltung vorhergesagt, welche die Aptamere ganz oder teilweise maskieren. Für die restlichen Konstrukte zeigt sich hingegen eine korrekte Faltung, beide Aptamere werden ausgebildet.

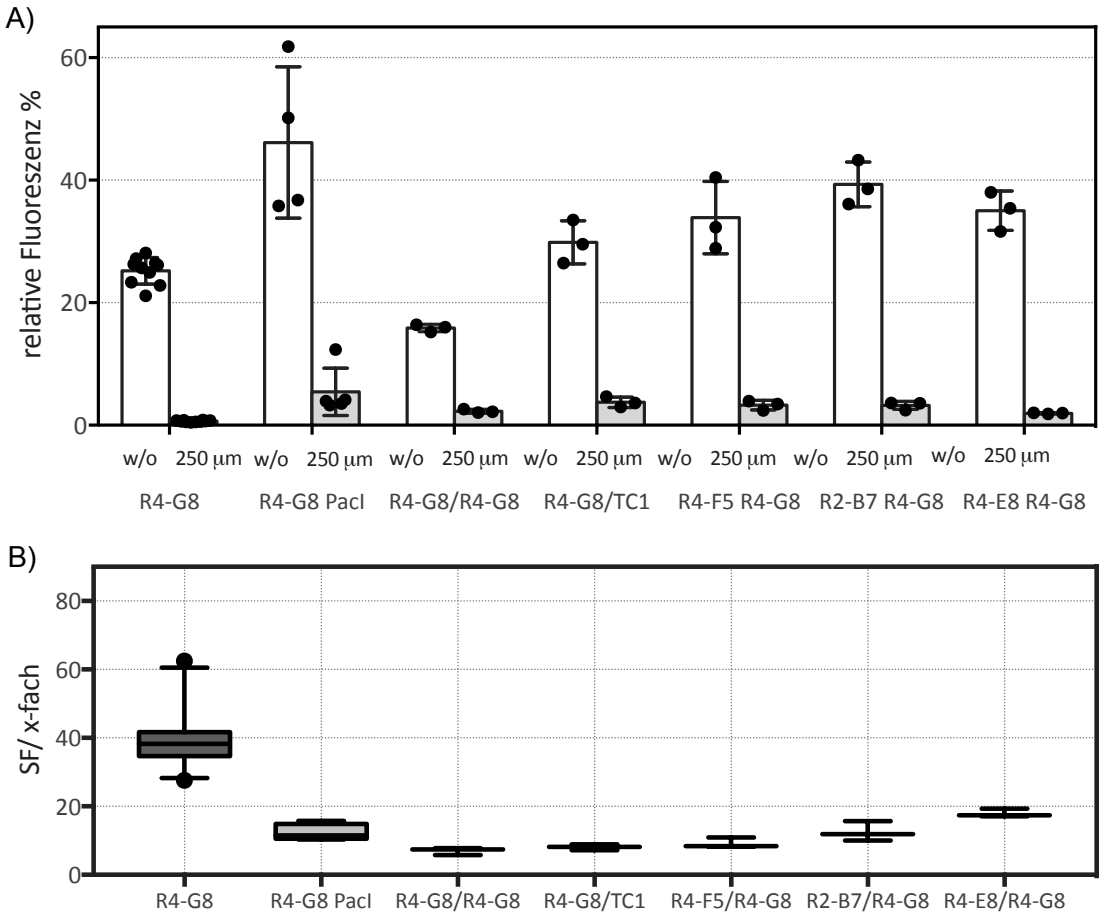
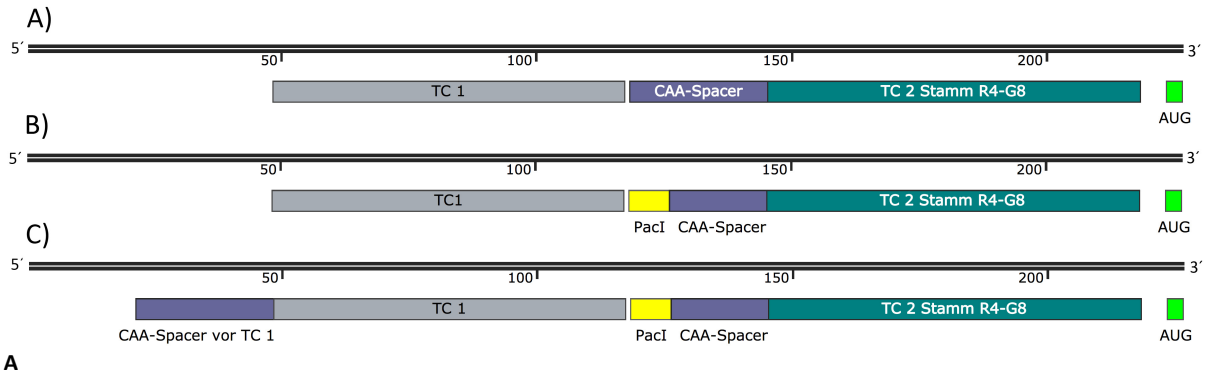


Abbildung 3.13.D Expression und Schaltfaktor der R4-G8-Doppelkonstruktemit PacI-Schnittstelle und ausgetauschtem 5´Aptamer A) Vergleich der relativen Fluoreszenz von R4-G8, R4-G8 PacI, R4-G8/R4-G8, R4-G8/TC1, R4-F5 R4 G8, R2-B7 R4-G8 und R4-E8 R4-G8 ohne (w/o) und mit 250 µm Tetrazyklin. **B)** Boxplot: Vergleich des Schaltfaktors der Konstrukte. Whisker stellen P1 und P90 dar.

Abbildung 3.13.D vergleicht die Messungen der jeweiligen Konstrukte. Kein Konstrukt erreicht einen zu R4-G8 vergleichbaren Schaltfaktor. Obwohl die Faltung der beiden Konstrukte R4-B8/R4-G8 und von Konstrukt R2-B7/R4-G8 nicht korrekt vorhergesagt wurde, kann auch hier ein „Schalten“ des Riboswitches beobachtet werden. Im Vergleich zu den anderen Konstrukten erreicht R4-E8/R4-G8 mit knapp 20-fach den besten Schaltfaktor und auch einen besseren Schaltfaktor als das Konstrukt R4-G8 PacI.

Dass die eingefügte Schnittstelle PacI Fehlfaltungen verursacht, wurde zwar von RNA-Fold nicht vorhergesagt, ist aber möglich. Daher wurde ein weiteres Konstrukt designt, bei welchem die Sequenz vor dem 5´Aptamer durch einen CAA-Spacer ersetzt wurde. Eine Fehlfaltung des Aptamers sollte durch einen zweiten, 15 nt langen, 5´CAA-Spacer möglichst unterdrückt werden (Abbildung 3.13.E). Tatsächlich führte das Einfügen dieses zweiten Spacers zu einer niedrigeren Basalexpression und zu einem höheren Schaltfaktor. Dieser konnte im Vergleich zum Ausgangskonstrukt von 12-fach auf 17-fach erhöht werden (Abbildung 3.13.F).



A

Abbildung 3.13.E 5´UTR mit TC-Dimer und PacI-Schnittstelle. **A)** Schema 5´UTR mit TC-Dimer R4-G8 und **B)** eingefügter PacI-Schnittstelle sowie **C)** zweitem CAA-Spacer vor dem ersten TC-Aptamer. Die Gesamtanzahl der Basen des 5´UTRs blieb in allen Fällen unverändert.

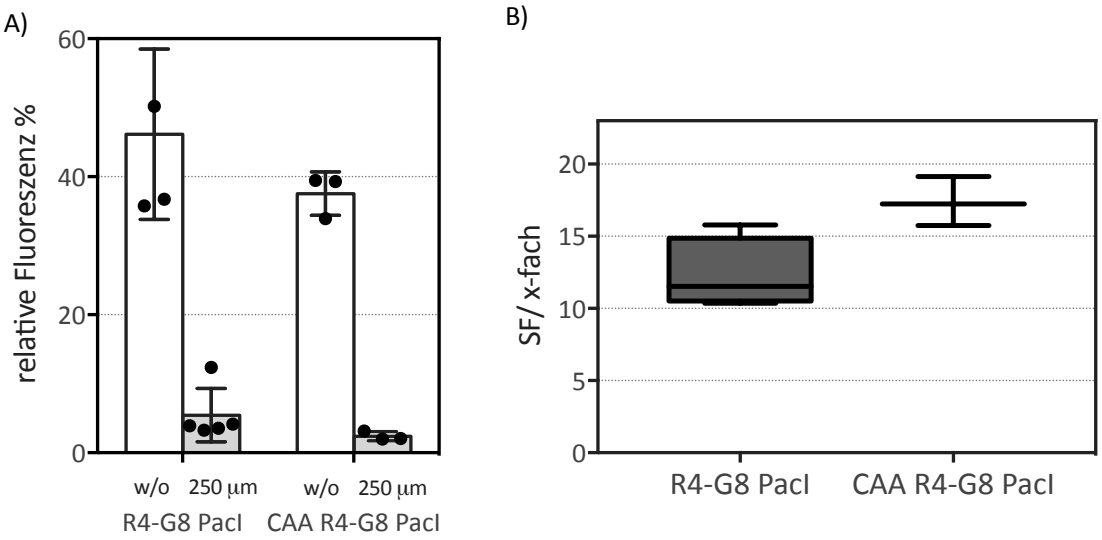


Abbildung 3.13.F Expression und Schaltfaktor von R4-G8 PacI und CAA R4-G8 PacI A) Vergleich der relativen Fluoreszenz von R4-G8 PacI und CAA R4-G8 PacI ohne (w/o) und mit 250 µM Tetrazyklin **B)** Boxplot: Vergleich des Schaltfaktors der beiden Konstrukte. Whisker stellen P1 und P90 dar.

3.14 Einzelkonstrukte – Einfluss auf Basalexpression und Schaltfaktor

Im weiteren Verlauf wurden die 3´Aptamere einiger Dimere als Einzelkonstrukte getestet. Es sollte die Frage beantwortet werden, ob Basalexpression und der Schaltfaktor mit den zugehörigen Dimeren in einem gewissen Verhältnis vergleichbar sind und welchen Einfluss die verschiedenen Stämme in diesen Monomeren haben. Im ersten Schritt wurden dazu drei verschiedene TC-Aptamere mit unterschiedlichen P1-Stämmen und unterschiedlich langen 5´UTRs getestet (126 bp und 135 bp)

(Abbildung 3.14.A). Es handelt sich dabei um die beiden Aptamere des Konstruktes LG3 (TC1M und TC2M) und um das 3´Aptamer des in dieser Arbeit neu gefundenen Dimers R4-G8 (R4-G8M).

Abbildung 3.14.A Sequenz der R4-G8 Monomere, Ausschnitt aus den Plasmidkarten. Oben das kürzere 5´UTR mit insgesamt 126 nt, unten das längere 5´UTR in welches ein zusätzlicher CAA-Spacer vor das Aptamer integriert wurde. Das 5´UTR wurde so auf 135 nt verlängert.

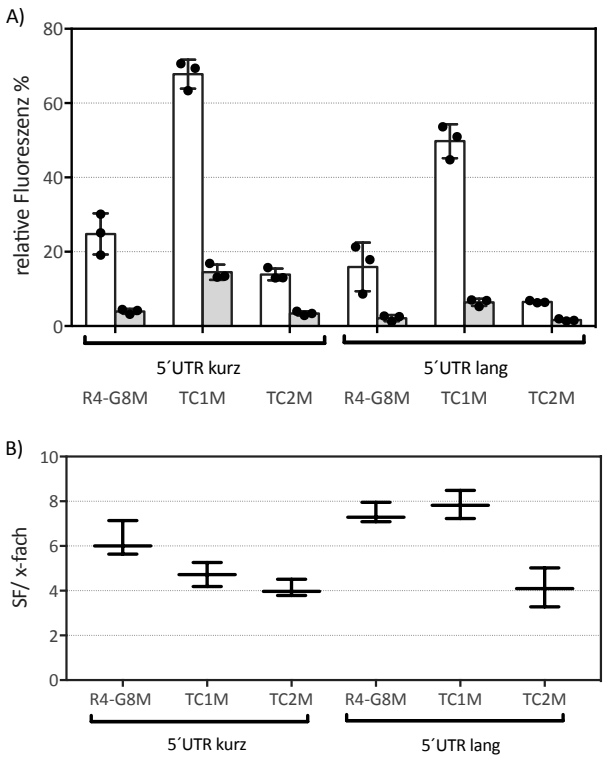


Abbildung 3.14.B Expression und Schaltfaktor der Monomere R4-G8Mo, TC1Mo und TC2Mo: A) Vergleich der relativen Fluoreszenz der Monomere R4-G8M, TC1M und TC2M mit kurzen und langen 5´UTRs ohne (w/o) und mit 250 µm Tetrazyklin. **B)** Boxplot: Vergleich des Schaltfaktors der sechs Konstrukte. Whisker stellen P10 und P90 dar.

Abbildung 3.14.B zeigt die relative Fluoreszenz der gemessenen Monomere im An- und Aus-Zustand und den Schaltfaktor. Generell zeigen die Konstrukte mit einem um 12 nt kürzen 5´UTR eine höhere GFP-Expression und einen etwas niedrigeren Schaltfaktor. Da mit den längeren 5´UTRs der Schaltfaktor höher war, wurde diese Länge für weitere Konstrukte genommen. Obwohl es sich bei den Riboswitchen nun nur noch um Monomere handelt, also in diesem Bereich des 5´UTRs nur noch eine Stammschleife vorhanden ist, kommt es bei den Aptameren R4-G8M und TC2 zu relativ niedrigen Expressionen. Nur der Riboswitch TC1 mit einem sehr kurzen Stamm von fünf nt zeigt eine sehr hohe Basalexpression.

Für die Klonierung weiterer Monomer-Konstrukte fiel die Entscheidung auf den längere 5´UTR, da hier ein besserer Schaltfaktor erzielt werden konnte. Die Sequenzen der P1-Stämme der Monomere finden sich in Tabelle 3.14. Anders als erwartet befindet sich die Basalexpression in einem den Dimeren ähnlichen Bereich, jedoch fällt auf, dass die getesteten Monomere alle einen wesentlich geringeren Schaltfaktor aufweisen als der Durchschnitt der Dimere. Die Schaltfaktoren liegen in einem Bereich zwischen 5- und 10-fach (Abbildung 3.14.C).

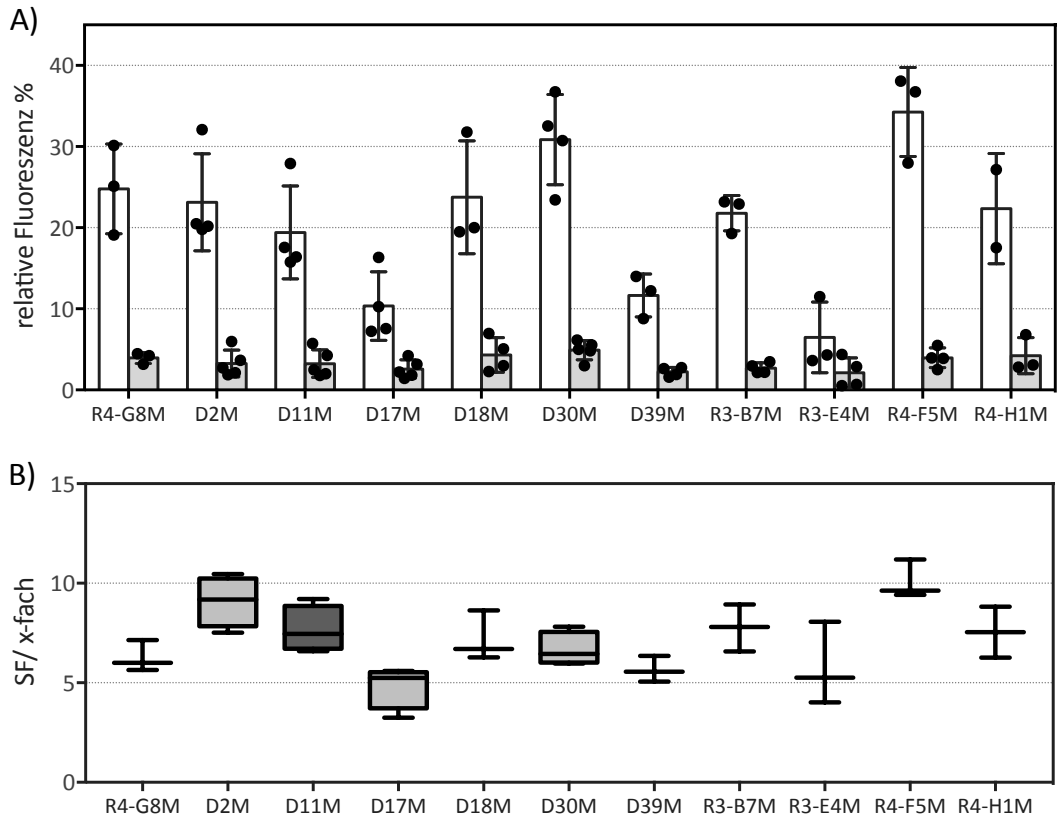


Abbildung 3.14.C Expression und Schaltfaktor (SF) der Einzelkonstrukte A) Vergleich der relativen Fluoreszenz der ohne (w/o) und mit 250 µM Tetrazyklin. **B)** Boxplot: Vergleich der Schaltfaktoren (SF) von 11 Konstrukten. Whisker stellen P10 und P90 dar.

Tabelle 3.14 Stamm-Namen und GFP-Expression in % der Dimere und Monomere, in Klammern ist die jeweilige Standardabweichung gezeigt.

		w/o Ligand (Expression in %)		250 µm TC (Expression in %)		SF [X-fach]	
Name	Stamm	Dimer	Monomer	Dimer	Monomer	Dimer	Monomer
R4-G8	AGGTGACC	25,2 (2,17)	15,9 (6,5)	0,7 (0,13)	2,2 (0,8)	39,5 (9,21)	7,4 (0,84)
Derivat 2-D2	CGGTGACC	31,5 (6,9)	23,1 (5,9)	1,8 (0,7)	2,6 (0,8)	18,6 (3,8)	9,1 (1,2)
Derivat 11-D11	AGGCGACC	32,2 (8,8)	19,4 (5,7)	2,0 (0,7)	2,6 (1,1)	16,4 (2,2)	7,7 (1,1)
Derivat 17-D17	AGGTGCCC	16,1 (4,8)	10,3 (4,2)	2,1 (0,6)	2,2 (0,8)	7,9 (0,6)	4,8 (1,1)
Derivat 18-D18	AGGTGGCC	62,9 (15,4)	23,7 (6,9)	11,3 (3,4)	3,4 (1,5)	5,7 (0,7)	7,2 (1,2)
Derivat 30-D30	AGGUGAGG	32,2 (4,8)	30,9 (5,5)	3,1 (1,1)	4,7 (1,3)	11,0 (2,7)	6,7 (0,8)
Derivat 39-D39	AGGTGGGC	20,5 (4,2)	11,7 (2,6)	2,3 (0,3)	2,1 (0,6)	9,0 (0,9)	5,7 (0,6)
R3_B7	ATCGGTGAC	45,5 (0,4)	21,8 (2,2)	1,9 (0,2)	2,9 (0,7)	23,8 (2,6)	7,8 (1,2)
R3_E4	AGGGCATC	8,3 (0,5)	6,5 (4,4)	0,4 (0,1)	1,4 (1,3)	21,5 (2,8)	5,8 (21)
R4_F5	TCGCGAGC	48,4 (3,7)	34,3 (5,5)	2,5 (0,4)	3,4 (0,8)	19,6(1,6)	10,1 (1,0)
R4_H1	TACCGAGC	41,6 (2,0)	22,3 (6,8)	2,3 (0,2)	2,9 (0,2)	18,0 (1,4)	7,5 (1,8)

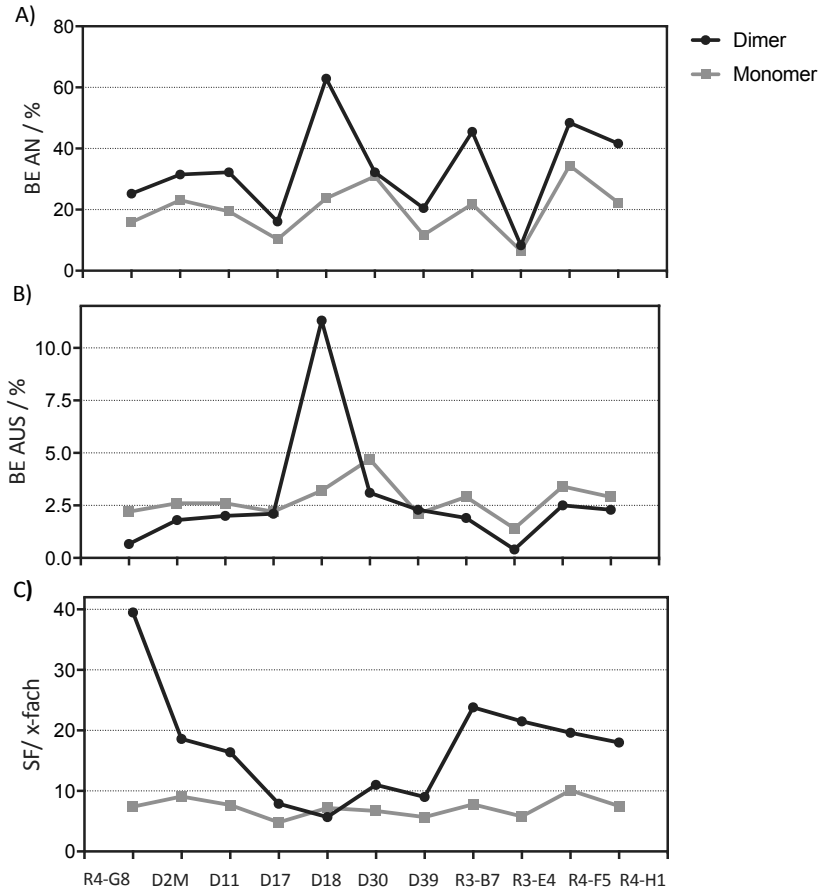


Abbildung 3.14.D Vergleich der Expressionsmuster (BE) und Schaltfaktoren (SF) von Monomeren mit den entsprechenden Dimeren. Die Dimerkonstrukten besitzen als 5‘ Aptamer Tc1. **A)** Expression im Liganden-ungebundenen (BE AN) Zustand der Monomere und Dimere. **B)** Expression im Liganden-gebundenen Zustand (BE AUS) der Monomere und Dimere. **C)** Schaltfaktoren (SF) der Monomere und Dimere.

Der Vergleich mit den Dimeren zeigt bei der Basalexpression ein ähnliches Muster (Tabelle 3.14 und Abbildung 3.1A.D). In Bezug zu anderen Riboswitchen scheint sich das Expressionsmuster von Monomer und Dimer kaum zu unterscheiden. Obwohl bei den Monomeren im 5´UTR nur eine Stammschleife vorkommt, wird die Expression hier mehr gehemmt. Auch das Muster der Expression im Liganden-gebundenen Zustand ist in Bezug auf andere Riboswitche ähnlich. Dies kann man jedoch kaum auf den Schaltfaktor übertragen. Da sich die Werte der Basalexpression und des Aus-Zustands zum Teil unterscheiden kann beim Schaltfaktor kein Bezug zwischen Dimeren und Monomeren erkannt werden.

3.15 Ein Tetrazyklin-Tobramycin-Hybrid-Riboswitch als NOR-Gate

Nachdem sich gezeigt hatte, dass ein Monomer mit dem gleichen P1-Stamm ein ähnliches Expressionsmuster im Liganden-ungebundenen und -gebundenen Zustand aufweist, sollte mit Hilfe einer kleinen Auswahl an P1-Stämmen ein funktionales NOR-Gate entwickelt werden. Bei diesem sollten die Aptamere nicht wie üblich in Reihe geschaltet, sondern hybridisiert werden. Ein Tobramycin -Aptamer, charakterisiert im Rahmen einer Bachelorarbeit von Eric Bräuchle, sollte mit einem TC-Riboswitch so fusioniert werden, dass beide ihre Funktionalität und ihre Fähigkeit zur spezifischen Ligandenbindung behalten. Aus der in dieser Arbeit klonierten und gemessenen Auswahl von 402 Konstrukten wurden drei ausgewählt. Die Monomere (3´Aptamere) dieser Riboswitche, welche sich nur in ihrem P1-Stamm unterscheiden, sollten mit dem Tobramycin-Aptamer V4_22 fusioniert werden. Da der P2-Stamm und der Loop L2 des TC-Aptamers nicht an der Bindung des Liganden beteiligt ist und deshalb ebenso veränderbar ist wie der P1-Stamm, wurde das komplette und ungekürzte Tobramycin-Aptamer mit seinem P1-Stamm auf dem Loop L2 platziert, zwischen Base 25 (Cytosin) und Base 26 (Adenin) (Abbildung 3.15.A). Für diese Klonierung wurden die in Tabelle 3.15 aufgeführten zwei P1-Stämme ausgewählt. Ein weiteres Konstrukt beinhaltet zwischen dem Loop L2 des TC-Aptamers und dem Stamm P1 des Tobramycin-Aptamers einen 14 nt langen Adenin-Linker (Abbildung 3.15.A A)). Dieser Linker sollte einen Loop ausbilden und somit den ΔG des gesamten Konstrukts erhöhen. Es bestand die Befürchtung, dass der niedrige ΔG des gesamten TC-Tobramycin-Hybrids zu einer recht niedrigen initialen Basalexpression führen würde.

Tabelle 3.15: Für die Klonierung des Hybridkonstrukts ausgewählte P1-Stämme; Basalexpression (BE), Expression im Aus-Zustand (Aus) und Schaltfaktor (BE).

Name	Stamm P1 [5´-3]	BE [%]	Aus [%]	SF [X-fach]	Tm [°C]	dG [kcal/mol]	GC-Gehalt [%]	Entropie [bit]
R4_F5	UCGCGAGC	48,4	2,5	19,7	46,9	-26,2	75	1,255
R4_G8	AGGUGACC	25,2	0,7	39,5	43,9	-25,0	63	1,321
R3_G6	AGUAAUCUGC	74,2	6,3	12,1	41,1	-25,3	40	1,366

Gewählt wurden zwei Konstrukte, die einen sehr guten Schaltfaktor aufweisen und ein Konstrukt mit einer hohen BE und einem mittleren Schaltfaktor. Das Konstrukt R4-F5 weist einen guten Schaltfaktor auf und eine Basalexpression von fast 50%. Vor Beginn der Klonierung wurde eine Faltungsvorhersage mit RNA-fold durchgeführt um zu überprüfen, ob sich die Hybride richtig falten würden (Abbildung 3.15.A).

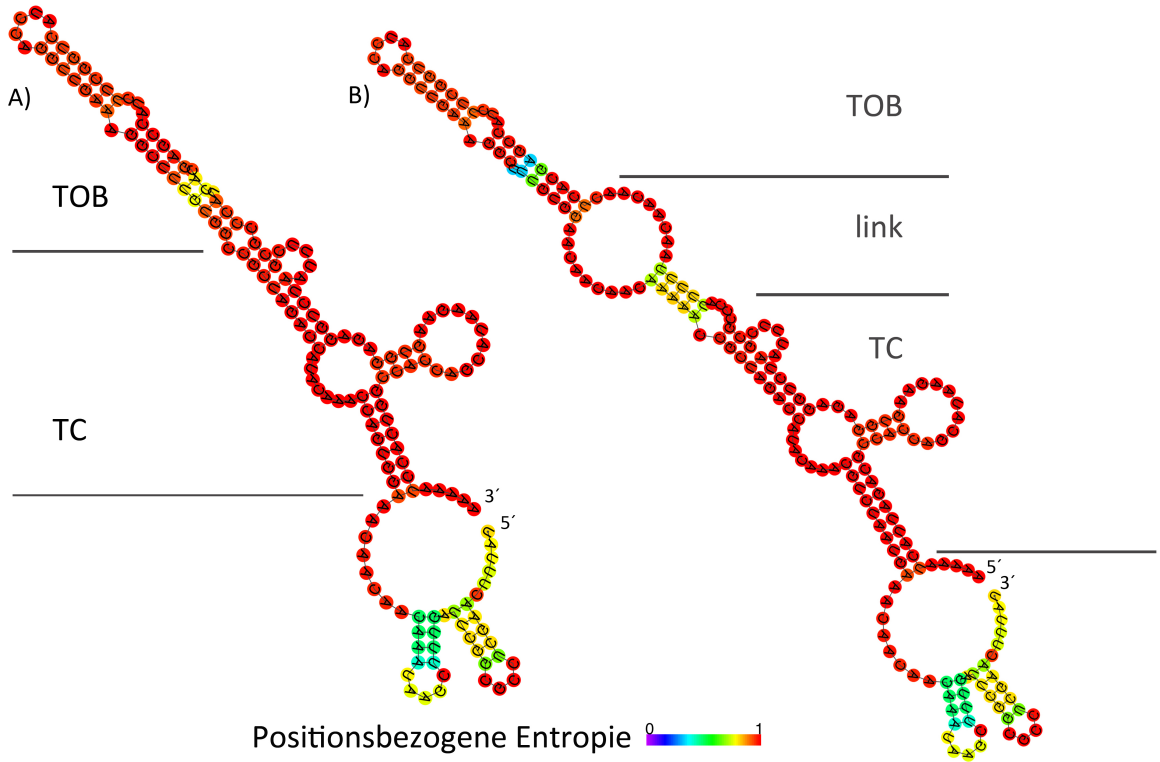


Abbildung 3.15.A RNA-Fold Faltungsvorhersage für das TC-Tobraramycin (TOB)-Hybrid. A) 5´UTR mit TC-TOB-Hybrid (TOB: Va22G8), das 5´UTR beginnt am Transkriptionsstartpunkt und endet am AUG **B)** 5´UTR mit TC-TOB-link-Hybrid (Va22G6 link), das 5´UTR beginnt am Transkriptionsstartpunkt und endet am AUG.

Abbildung 3.15.A A) zeigt die Faltung des 5´UTRs, in welchem das TC-Aptamer R4-G8 mit dem Tobramycin-Aptamer V422 fusioniert wurde (TC-TOB-Hybrid). Man kann erkennen, dass sich beide Aptamere in korrekter Weise falten. Abbildung 3.15.A B) zeigt das TC-Aptamer R3-G6 mit dem Linkerbereich, der einen Loop ausbildet. Obwohl Teile des Linkers mit dem TC-Aptamer interagieren, bleiben die für die Ligandenbindung relevanten Teile der Aptamere intakt.

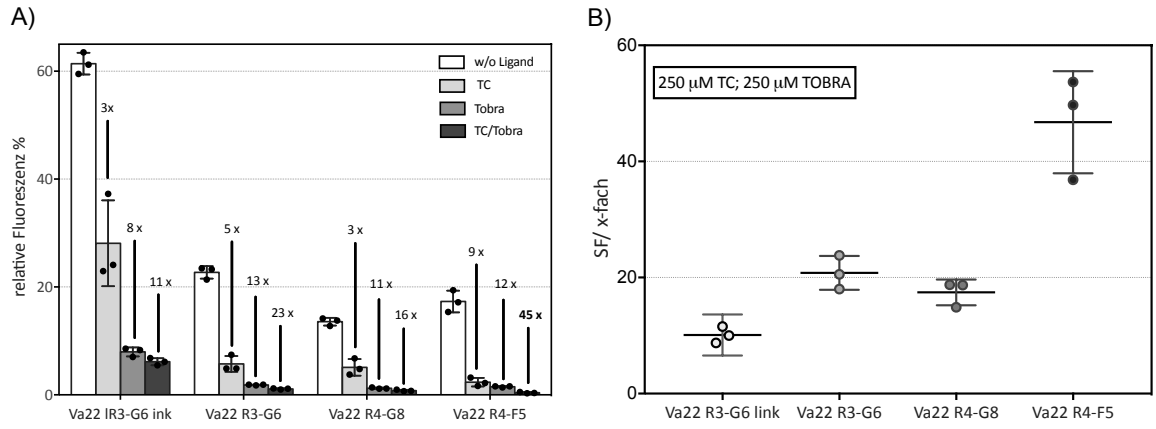


Abbildung 3.15.B Expression und Schaltfaktor der TC-TOB-Hybride Va22 R3-G6 link; Va22 R3-G6, Va22 R4-G8 und Va22 R4-F5 A) Vergleich der relativen Fluoreszenz der vier Hybrid-Konstrukte ohne (w/o) und mit 250 µM TC, 250 µM Tobramycin und 250 µM T + 250 µM Tobramycin (TC/Tobra), die jeweiligen Schaltfaktoren stehen über den einzelnen Balken. **B)** Scatterdotplot: Vergleich der Schaltfaktoren (SF) der 4 Konstrukte, inkubiert mit je 250 µM TC und Tobramycin. (Weiß: Va22 R3-G6; Hellgrau: Va22 R3-G6; Dunkelgrau: Va22 R4-G8; Schwarz: Va22R4-F5), Konfidenzintervall 95%.

Abbildung 3.15.B zeigt die Ergebnisse der GFP-Messungen. Bereits mit der Inkubation von 250 µM TC kommt es bei allen Konstrukten zu einer Reduktion der Fluoreszenz. Wenn die Hefen mit 250 µM Tobramycin angezüchtet werden, kommt es zu einer noch stärkeren Reduktion der Fluoreszenz. Am stärksten ist der Effekt in Gegenwart beider Liganden. Die verschiedenen Konstrukte unterscheiden sich zum Teil stark in ihrem Schaltverhalten. Die höchste Basalexpression erreicht das Konstrukt Va22 R3-G6 link, welches den 3´P1-Stamm des Konstruktes G6 aus der *machine learning*-Runde 3 enthält. Das Dimer weist eine sehr hohe Basalexpression mit fast 75% auf und wurde auch aus diesem Grund ausgewählt. Auch in Kombination mit dem Tobramycin-Aptamer kann noch eine recht hohe Basalexpression von über 60% erreicht werden. Weniger hohe Basalexpressionen, dafür aber bessere Schaltfaktoren weisen die anderen Konstrukte auf. Mit dem Austausch des Stammes R3-G6 durch die Stämme R4-G8 und R4-F5 wurde die Basalexpression um mindestens 2/3 reduziert. Der P1-Stamm von R3-G6 hat einen niedrigeren GC-Gehalt aber mehr Basen als die anderen Konstrukte. Die höhere Basalexpression hängt möglicherweise mit dem niedrigeren GC-Gehalt des Stammes zusammen. Jedoch hat der Stamm R4-F5 einen GC-Gehalt von 75% und dennoch eine höheren Basalexpression als R4-G8 mit einem GC-Gehalt von 63%. Erstaunlicherweise hat hier nicht das Konstrukt mit dem 3´P1-Stamm von R4-G8 den besten Schaltfaktor, sondern das Konstrukt mit dem 3´P1-Stamm von R4-F5. Dieses zeigt hier auch eine etwas höhere Basalexpression, was ein Grund für einen besseren Schaltfaktor sein kann. Das Experiment zeigt, dass zwei Aptamere auf einfache Weise mit einander fusioniert werden können, ohne die jeweilige Ligandenbindung zu stören. Der Schaltfaktor kann mit der Zugabe von beiden Liganden sogar weit über den Schaltfaktor der Einzelkonstrukte hinaus verbessert werden.

Abbildung 3.15.C zeigt die Histogramme der Einzelzellpopulationsmessungen der Hybrid-Konstrukte, welche mit dem Cytometer aufgezeichnet wurden. Man erkennt, dass zum einen bei dem Konstrukt Va22 R4-F5, verglichen mit den Konstrukt Va22 R4-G8, ein höheres GFP-Signal vorhanden ist und zum anderen, dass das Signal nach Inkubation mit TC bei dem Konstrukt Va22 R4-F5 stärker inhibiert wird. Die Signalstärke von GFP nach Inkubation mit Tobramycin bleibt bei beiden Konstrukten dagegen auf einem ähnlichen Niveau und wird durch die Veränderung des P1-Stamms nicht beeinflusst.

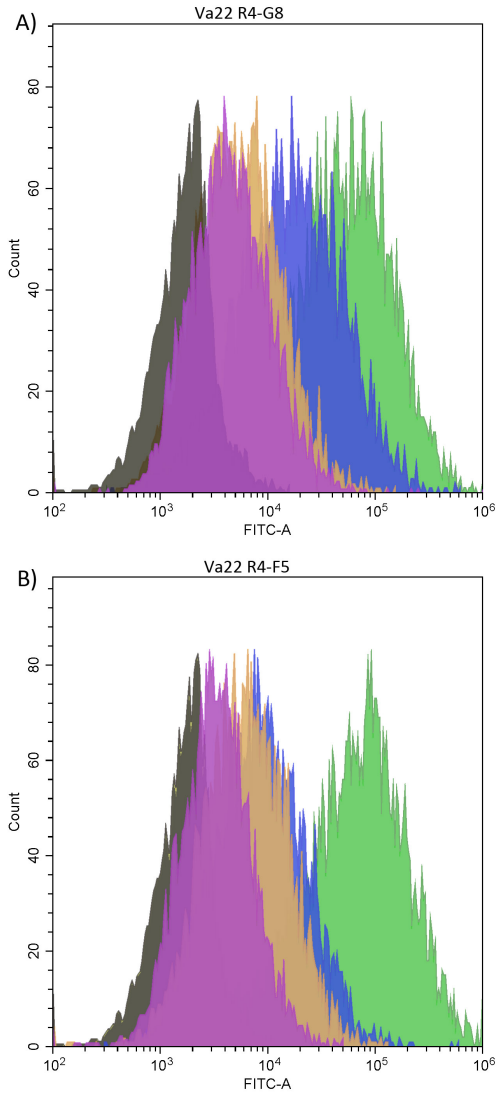


Abbildung 3.15.C: Einzelzellanalyse des durch die im 5´UTR eines GFP-Gens inserierten Konstrukte Va22 R4-G8 und Va22 R4-F5. Das Histogramm zeigt die Häufigkeit jedes GFP-Fluoreszenzniveaus für Zellen, die mit 250 µM TC, 250 µM Tobramycin oder mit je 250 µM TC und 250 µM Tobramycin inkubiert wurden und welche die Riboswitche Va22 R4-G8 und Va22 R4-F5 im 5´UTR eines GFP-Gens enthalten. Für die Messung wurden 5000 Ereignisse pro Population automatisch aufgezeichnet. Die x-Achse ist biexponentiell aufgetragen. Individuelle Populationen der Zellen wurden ohne TC (grün), mit 250 µM TC (blau), mit 250 µM Tobramycin inkubiert (orange) bzw. mit 250 µM TC und mit 250 µM Tobramycin zusammen inkubiert (lila). In schwarz dargestellt ist eine Population an Zellen, welche mit dem Plasmid IBB transformiert wurde, welchem das AUG vor dem GFP-Gen fehlt. Diese Population an Zellen wurde als Hintergrundwert angenommen.

Mit der Kombination der beiden Aptamere wurde ein funktionales NOR-Gate erzeugt, dessen Wirkweise schematisch in Abbildung 3.15.D zu sehen ist. Abbildung 3.15.E zeigt als Heatmap, wie sich der Schaltfaktor bei den unterschiedlichen Konstrukten und der Liganden-Zugabe verändert.

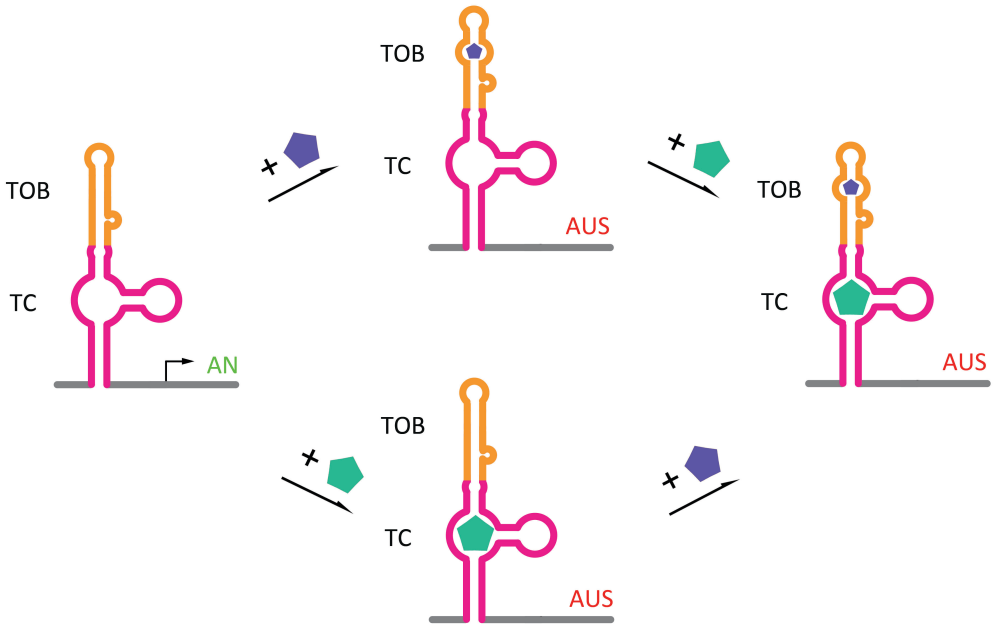


Abbildung 3.15.D Schema des TC-TOB-NOR-Gates. Das Tobramycin-Aptamer (TOB) ist mit dem Tetrazyklin-Aptamer (TC) fusioniert. Ohne Liganden kann die Translation stattfinden, die Genexpression ist angeschaltet. Bei Zugabe von Tobramycin oder TC wird die Genexpression ausgeschaltet, auch wenn beide Liganden anwesend sind, ist die Genexpression aus.

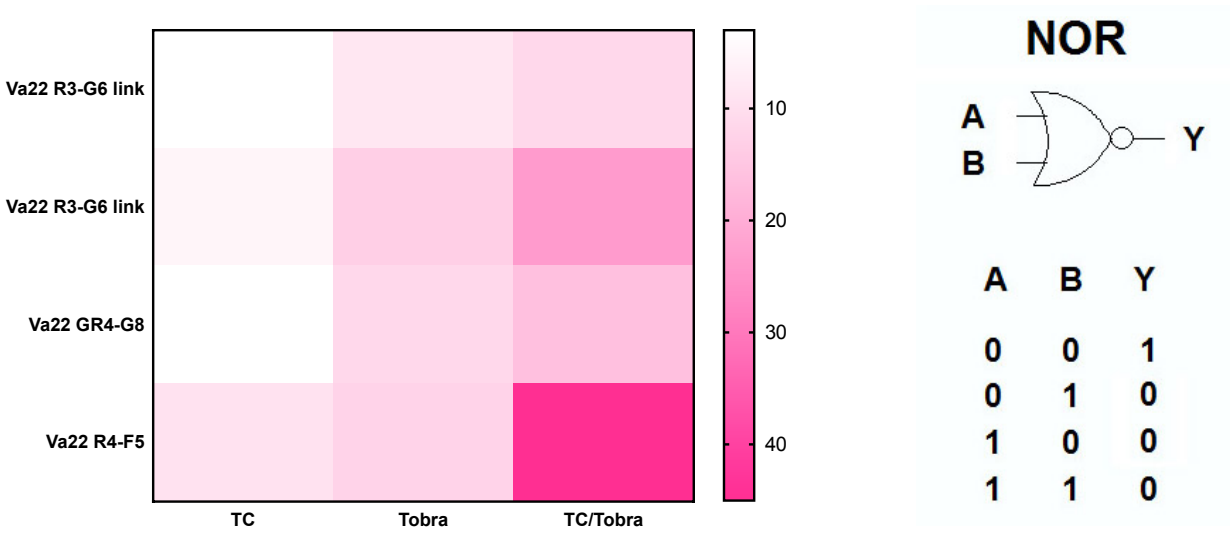


Abbildung 3.15.E Heatmap der Schaltfaktoren und Schema eines NOR-Gates. A) Heatmap der Schaltfaktoren in Bezug auf die verschiedenen Tobramycin-TC Hybride und der Zugabe der Liganden. **B)** Schema eines NOR-Gates. Sobald ein Signal im Eingang vorhanden ist oder beide vorhanden sind, ist kein Ausgangssignal mehr vorhanden (0).

4 Diskussion

4.1 Vorteile von *machine learning* als Methode zum Design und der Optimierung von Riboswitchen

Meist können Aptamere nach der Selektion nicht sofort als Riboswitche eingesetzt werden, sondern müssen erst einer Optimierung unterzogen werden. Deshalb folgt oft ein genetisches *in vivo* Screening, bei welchem ein durch SELEX angereicherter Aptamer-Pool im Zielorganismus auf die Schaltbarkeit eines Reportergens selektiert wird (Berens et al. 2015). *In vivo* durchgeführte Hochdurchsatz-Screenings, welche die gleichzeitige Messung vieler verschiedener Konstrukte durch die Kopplung von Zellsortierung (FACS) und NGS ermöglichen, haben sich bereits als vorteilhaft erwiesen und beschleunigen den Screening-Prozess. Zudem ist die Funktionalität eines RNA-Schalters stark kontextabhängig, da es leicht zu Um- bzw. Fehlfaltungen kommen kann. So ist das Ergebnis eines Screening-Prozesses schwer plan- oder abschätzbar und kann durch einen größeren Ansatz an Effizienz gewinnen. Nachdem durch die oben beschriebene Kombination von *in vitro* Selektion und *in vivo* Screening ein *in vivo* schaltendes Konstrukt gefunden wurde, muss in den meisten Fällen die Schaltungseffizienz weiter verbessert werden. Üblicherweise sind die Riboswitche anfangs noch schwach in ihrer Effektivität und müssen über Permutation, gerichtete und ungerichtete Mutation und rationale Design-Ansätze verbessert werden (Groher 2018, Weigand 2008, Süß 2003). Das rationale Design ist hier ein üblicher Ansatz, jedoch ist dieser oft langwierig, von vielen Iterationen geprägt und stark von den Fähigkeiten und Erfahrungen des Versuchsplaners abhängig. Kleinere Ansätze müssen neu designt, überprüft und verbessert werden (McKeague et al. 2016). Das *machine learning* und *deep learning*, bei welchen strukturell über die gemessenen Konstrukte Programme zur Leistungsverbesserung trainiert werden können, bieten hier einen interessanten Ansatz, um den Prozess der Optimierung zu automatisieren und zu beschleunigen. Zusätzlich werden viele Daten gewonnen und eine Vielzahl an Konstrukten erstellt, die zu unterschiedlichen Zwecken eingesetzt werden können. Dies könnte sich als ein großer Vorteil erweisen. Da die Designregeln eines neuen Aptamers bislang nur unzureichend verstanden sind, war es bisher nicht möglich, ein Aptamer rechnergestützt zu entwerfen, das eine Regulation der Genexpression ermöglicht (Berens et al. 2015; Schneider & Suess 2015).

4.2 Versuchsdesign - Grenzen und Optimierungsvorschläge

Beim Gegenstand dieser Studie, dem TC-Aptamer als Riboswitch-Dimer, handelt es sich um einen bereits sehr guten und optimierten Riboswitch mit einem vergleichsweise hohen Schaltfaktor und einer Basalexpression nahe 20 %. Es stellt eine Herausforderung dar, einen bereits so guten Schalter, der sich nach einem iterativen Optimierungsprozess bereits nahe seinem Optimum befindet, weiter zu verbessern.

Aus Gründen der Praktikabilität und der zur Verfügung stehenden finanziellen Mittel wurde eine Klonierungsstrategie gewählt, bei welcher pro Runde maximal 96 verschiedene Konstrukte kloniert wurden. Daraus resultierte die Entscheidung, eine Verbesserung der Riboswitche zunächst nur mit einem *machine learning* Programm zu erzielen, da dieses weniger Datenpunkte braucht als ein *deep learning* Ansatz.

Um ein erstes Datenset für das *machine learning* zur Verfügung zu haben, wurde das erste Set von 96 Konstrukten so generiert, dass der P1-Stamm des 3´Aptamers variabel gehalten wurde. Das erste Datenset war im Hinblick auf die erzeugte Variabilität der Daten deutlich unterrepräsentiert. Zwar konnte in der darauffolgenden Runde ein Schalter gefunden werden, welcher einen besseren Schaltfaktor aufwies, als das Ausgangskonstrukt LG3, jedoch konnte der mittlere Schaltfaktor der 2. Runde gegenüber der 1. Runde keine Verbesserung erzielen. In der 3. Runde konnte schließlich erstmals eine Verbesserung des mittleren Schaltfaktors im Vergleich zu den ersten beiden Runden erzielt werden. Mit der Ergänzung des *machine learning* Programms um ein *deep learning* Programm in der 4. Runde, konnte eine weitere signifikante Verbesserung erzielt und zudem ein sehr gut schaltender Riboswitch gefunden werden. Obwohl sich die Leistung des Ausgangskonstrukts bereits auf einem sehr hohen Schaltniveau befand, konnte dieses durch nur vier Runden *machine- und deep learning* verdoppelt werden. Diese Leistung ist umso erstaunlicher, da das Programm bis zur 4. Runde mit nur 243 Sequenzen trainiert worden war. Dies stellt eine sehr geringe Datenmenge für ein *machine learning* Programm dar, denn die Vorhersagegenauigkeit hängt auch von der Größe des Trainingssatzes ab (Faber et al. 2016; Schmidt et al. 2017). Zhang *et. al.* zeigten 2018, dass der skalierte Fehler verschiedener *machine learning* Modelle bei kleinen Trainingssätzen von 100-200 Proben bei 10% oder mehr liegt (Zhang et al. 2016). Dabei ist der mittlere absolute skalierte Fehler ein Maß für die Genauigkeit von Prognosen (Hyndman 2006). Ein weiterer Hinweis, dass mehr Daten zu besseren Ergebnissen führen, ist die Verbesserung des AUCs von Runde 2 auf Runde 4. So lag die Vorhersagekraft nach der 1. Runde noch bei 0,7. Von der 2. bis zur 3. Runde konnte dieser Wert auf 0,89 erhöht werden. Der beste AUC konnte schließlich durch den in der vierten Runde zusätzlich angewandten CNN, den veränderten Hyperparamter sowie die Verkleinerung des Lösungsraums durch die Beschränkung auf Stämme mit 8 nt, erzeugt werden, hier lag die Vorhersagekraft bei 0,92 (A.-C. Groher et al. 2018).

Um die Leistung des *machine learning* Ansatzes zu verbessern, wäre es daher empfehlenswert, wesentlich mehr Konstrukte zu analysieren. Für die Generierung von Daten für die Anwendung maschinellen Lernens sollten im besten Fall mehrere tausend Konstrukte teilweise randomisiert, kloniert und gemessen werden. Eine Analyse dieser Daten und eine zweite, auf *machine learning* bzw. *deep learning* basierten Runde, würde ein noch tieferes Verständnis der notwendigen biophysikalischen Parameter und Sequenzen ermöglichen.

Um das *machine learning* Programm noch gezielter trainieren zu können, sollte auch das Design des Ausgangskonstruktes überdacht werden. Die Optimierungen in dieser Arbeit wurden an einem Dimer durchgeführt. Ein Dimer, welches zudem aus zwei zum größten Teil identischen Aptameren besteht, unterliegt gewissen Beschränkungen. So kann es durch die gleichen Sequenzabschnitte der beiden Aptamere zu ungewollten Fehlfaltungen kommen. In Kapitel 3.13 konnte gezeigt werden, dass solche Fehlfaltungen nicht vollständig durch den CAA-Spacer verhindert werden konnten. Zwischen den beiden komplett identischen Aptameren kam es laut Vorhersage zu Fehlfaltungen. Begründet wird dies wahrscheinlich dadurch, dass die Sequenzen komplett komplementär sind. In der vorhergesagten Faltung sind die Aptamer-Motive in diesen Sekundärstrukturen nicht mehr erkennbar (Abbildung 3.13.C A) und D)).

4.3 Analyse der biophysikalischen Parameter

Eine rein auf biophysikalischen Parametern basierende Optimierung von Riboswitchen ist bei bisherigen Ansätzen unüblich. In der Regel werden auch bei bisher verwendeten Strategien Sequenzen analysiert und Sequenzabschnitte, welche sich als gut erwiesen, teilweise beibehalten. Da ein Lernen auf Basis der Sequenz zu Beginn auf Grund der geringen Datenmenge als nicht sinnvoll erachtet wurde, war für die Verbesserung des Riboswitches nur ein *machine learning* Ansatz möglich, der auf biophysikalischen Parametern lernte. Rückblickend ermöglicht dieser Ansatz jedoch eine Analyse der biophysikalischen Parameter der Stämme, welches durch einen Ansatz, der nur auf Basis der Sequenz lernt, nicht möglich gewesen wäre.

Bereits in den ersten drei Runden konnte beobachtet werden, dass die Veränderung der biophysikalischen Parameter in enger Weise mit der Veränderung der Basalexpression und des Schaltfaktors zusammenhängt. Gerade die Verschlechterung des mittleren Schaltfaktors in Runde 2 und die damit korrelierende Veränderung der Basalexpression scheint dies zu zeigen. Während sich von Runde 2 auf 3 und von Runde 3 auf 4 der mittlere Schaltfaktor erhöht, sich gleichermaßen die mittlere Basalexpression reduziert und die biophysikalischen Parameter sich mehr in einem gewissen Raum verdichten, kommt es in Runde 2 zu einer Verschlechterung des Schaltfaktors, zu einer

höheren mittleren Basalexpression und zu wesentlich breiter verstreuten Werten der biophysikalischen Parameter.

Die Auswertung aus Kapitel 3.12, welche die biophysikalischen Parameter im Hinblick auf die Expressionen und den Schaltfaktor darstellen, machen eine noch genauere Analyse möglich. Es zeigt sich bei allen Parametern ein deutlicher Trend und ein Zusammenhang zwischen den Expressionen im An- und Auszustand, dem Schaltfaktor und den biophysikalischen Parametern. Zudem ist erwähnenswert, dass der Zusammenhang bestimmter biophysikalischer Parameter mit der Expression im Aus-Zustand stärker ist als mit der Expression im An-Zustand. Generell ist die Korrelation der Expression und des Schaltfaktors beim ΔG und Tm ausgeprägt. Bei der Betrachtung der Heatmap (Abbildung 3.12.B), welche den Zusammenhang des Schaltfaktors mit den biophysikalischen Parametern darstellt, kann man erkennen, dass bestimmte Parameter im P1-Stamm gewissermaßen ein „Ausschlusskriterium“ für gute Schalter sind. So wird bei einem Tm unter 10°C kein Schalter mit einem Schaltfaktor höher als 8-fach gefunden. Wohingegen bei einem Tm von 46-48°C kein Schalter gefunden wurde, der einen SF geringer als 9-fach aufweist. Der Tm-Bereich zwischen 44°C und 48°C hat sich als der beste Bereich herausgestellt. P1-Stämme mit diesem Tm haben eine hohe Wahrscheinlichkeit für einen hohen Schaltfaktor. Beim ΔG zeichnet sich ein ähnliches Bild ab. Hier werden die höchsten Schaltfaktoren mit Konstrukten erreicht, die einen ΔG kleiner bzw. gleich -25 kcal/mol aufweisen. Besonders deutlich ist hier auch der Trend der aufzeigt, dass ein höherer ΔG mit einem schlechteren Schaltfaktor verbunden ist. Und auch hier zeigt sich, dass bei einem ΔG zwischen -24,5 kcal/mol und -25,0 kcal/mol keine Schaltfaktoren unter 8-fach erreicht werden. Auch beim GC-Gehalt lässt sich der Trend erkennen, allerdings ist er im Gegensatz zum Tm und ΔG nicht ganz so deutlich ausgeprägt. Am wenigsten ausgeprägt ist der Trend bei der Entropie. Eine generelle Aussage ist hier kaum möglich.

Der starke Einfluss des Tm und des ΔG-Wertes auf den Schaltfaktor und den Aus-Zustand lässt sich auch damit begründen, dass beides sehr stark die Stabilität der RNA-Struktur beeinflusst. Eine Struktur, die eine hohe basale Expression zulässt, ist auch im Liganden-gebundenen Zustand instabiler als eine Struktur, welche die Expression schon im ungebundenen Zustand mehr einschränkt. Die wesentlichen biophysikalischen Parameter einer RNA-Struktur, ΔG, Tm, GC-Gehalt und Entropie, bedingen einander und hängen miteinander zusammen. Sie alle beeinflussen die Stabilität der Struktur und somit die Basalexpression und auch die Expression im Liganden-gebundenen Zustand. Die Auswertung der Schaltfaktoren (Abbildung 3.12.B) zeigt jedoch auch, dass der Tm und damit die RNA-Struktur des Stamms nicht zu stabil sein darf, um einen guten Schaltfaktor zu generieren. Eine zu stabile Struktur mit einem Tm über 50 °C erreicht hier maximal einen Schaltfaktor von 16-fach. Es kann aber nicht ausgeschlossen werden, dass es auch P1-Stämme mit einem Tm über 50°C geben könnte, die einen höheren Schaltfaktor erreichen könnten.

4.4 Sequenz-Analyse der P1-Stämme

In Kombination mit dem *machine learning* Programm führte das *deep learning* Programm (CNN) zu einer weiteren Verbesserung des mittleren Schaltfaktors in Runde 4. Dass neben den biophysikalischen Parametern auch die Sequenz einen signifikanten Einfluss auf die Regulation hat, zeigte eine Testrunde, welche zwischen Runde 3 und 4 durchgeführt wurde (Kapitel 3.5). Die durch rationales Design konstruierten Dimere ähneln den bis dahin acht beste Konstrukte in ihrer Sequenz, denn es wurde versucht, diese den guten Konstrukten ähnlich zu gestalten. Die durch rationales Design konstruierten Riboswitche enthalten mehrheitlich (in sieben von zehn Fällen) ein Adenin als erste Base (5´) und der Stamm endet in den meisten Fällen mit einem Cytosin-Guanin-Basenpaar. Ein Adenin-Uracil- oder Uracil-Adenin-Basenpaar am Ende wurde vermieden. Die Sequenzen, welche mit dem *machine learning* Programm vorhergesagt wurden, sind zum größten Teil kürzer und verfügen nur über sieben Basenpaare. Die Analyse der Stammlängen in Kapitel 3.2, 3.3 und 3.4 zeigt, dass auch die Stammlänge einen Einfluss auf den Schaltfaktor und die Basalexpression hat und Stammlängen von 8 bis 9 nt häufiger gute Schaltfaktoren generieren als kürzere P1-Stämme. Zudem enden die mit dem *machine learning* Programm vorhergesagten Sequenzen in allen Fällen mit einem Adenin oder Uracil. Der Vergleich der mittleren Schaltfaktoren zeigt somit, dass auch das Einbeziehen der Sequenz in die Designstrategie den Schaltfaktor verbessern konnte. In der vierten Runde konnte somit mit dem CNN auch ein Lernen auf Grund der Sequenz ermöglicht werden. In dieser Runde konnte nicht nur der mittlere Schaltfaktor der Runde weiter verbessert werden, es konnte auch ein sehr guter Riboswitch (R4-G8) gefunden werden. Durch die anschließend durchgeführte Sequenzanalyse wurde ersichtlich, dass nicht die einzelnen Basen an bestimmten Positionen für einen guten Schaltfaktor wichtig sind, sondern viel mehr Sequenzabschnitte von 2 oder 3 Basen eine große Bedeutung zukommt.

Der Einfluss der Sequenzabschnitte auf den Schaltfaktor kann zumindest teilweise durch die „*Nearest-Neighbor*-Theorie“ begründet werden (Andronescu et al. 2013; Zuber et al. 2018; Mathews 2006). Nach diesen Regeln wird die Stabilität eines Basenpaares von der Sequenz benachbarter gepaarter oder ungepaarter Basenpaare bestimmt. Dies impliziert, dass die Stabilität und damit die Wirkung einer Base in einem Stamm auch von den umgebenden Basen abhängig ist. So kann zum Beispiel ein Adenin-Uracil-Basenpaar mit einem vorangegangenem Cytosin-Guanin-Basenpaar eine andere lokale Stabilität aufweisen als ein Adenin-Uracil-Basenpaar, welches an ein Guanin-Cytosin-Basenpaar grenzt. Gewisse Sequenzmotive des sehr guten Riboswitches R4-G8 lassen sich mit einer größeren Häufigkeit in gut schaltenden Riboswitchen finden (Bsp. AGG/GG/TG/GA und CA) (Abbildung 3.9.B). Das kann ein Hinweis darauf sein, dass diese Motive tatsächlich einen guten Schaltfaktor begünstigen. Denn das „Lernen“ des Programms führt zwangsweise zu einer Anhäufung

von Sequenzmotiven, die sich als günstig erwiesen haben. Ein größerer *machine learning* Ansatz würde auch hier eindeutigere Interpretationen ermöglichen.

4.5 Der Einfluss des Stammendes auf die Basalexpression

Kapitel 3.11 beschäftigt sich mit der Analyse von vier verschiedenen Stammendungen des P1-Stammes. Analysiert wurden die Endungen CC, GG, GC und CG (5´-3´). Diese Endungen zeichnen sich alle durch die gleiche Anzahl an Wasserstoffbrücken aus (drei) und verfügen über eine gleiche Stabilität. Durch die Analyse konnte aufgezeigt werden, dass Stammendungen mit zwei Cytosinen eine signifikant niedrigere Basalexpression bedingen als eine Stammendung mit zwei Guaninen bzw. einem Guanin und einem Cytosin. Die Begründung liegt hier möglicherweise nicht an der Stammstabilität und der Sekundärstruktur, sondern an der Nähe des 3´Endes der Aptamersequenz zur Kozak-Sequenz, wodurch Translationseffizienz möglicherweise beeinflusst werden kann (J. Li et al. 2017). Dass die Sequenz die Basalexpression auch unabhängig von Sekundärstrukturen beeinflussen kann, könnte durch die folgende Theorie begründen werden: Wenn das Ribosom mit dem Scanning-Prozess beginnt und die erste Stammschleife aufgetrennt hat, ist es wahrscheinlich, dass diese sich in hoch exprimierten Genen nicht wieder ausbildet, weil sich nunmehr viele Ribosomen auf der mRNA befinden (Mao 2014). Wenn man also die Basalexpression betrachtet, betrachtet man nach dieser Theorie unter anderem eine Wahrscheinlichkeit, mit welcher das erste Ribosom die Sekundärstruktur aufwinden konnte. Allerdings schlägt eine andere Theorie indes vor, dass sich die Stammschleifen nach Entwindung durch die Helikase-Funktion des Ribosoms möglichst schnell wieder zurückbilden, um eine Kollision der Ribosomen auf der RNA zu vermeiden (Mauger, 2019). Jedoch gelten diese Annahmen für den kodierten Bereich der mRNA, wenn das Ribosom aus seinen zwei Untereinheiten zusammengesetzt ist. Wie es sich bei dem 5‘UTR der mRNA verhält, wenn die mRNA von der kleinen Untereinheit des Ribosoms gescannt wird, wurde noch nicht untersucht. Es steht aber fest, dass Sekundärstrukturen in diesem Bereich der mRNA die Translationseffizienz negativ beeinflussen (Vega Laso et al. 1993).

Wie bereits in der Einleitung erwähnt, konnte 2017 von Li *et. al.* gezeigt werden, dass die Nukleotide direkt vor dem Startcodon einen positiven oder negativen Einfluss auf die Basalexpression haben. Besonders wichtig scheinen hier die Nukleotide im Bereich -11 bis -14 zu sein (die Base -1 ist die erste Base vor dem AUG im 5´UTR der mRNA), diese Nukleotide haben einen starken Einfluss auf die Translationseffizienz. In Abbildung 3.8.C erkennt man, dass sich die beiden analysierten Nukleotide genau in diesem Bereich vor der Kozak-Sequenz befinden: An Position -12 und -13. Auf dieser Seite des Stammes befinden sich auf Grund der Basenpaarung zwei Guanine und es konnte bereits gezeigt

werden, dass Guanine im Bereich der Kozak-Sequenz einen hemmenden Einfluss auf die Translationseffizienz haben (J. Li et al. 2017). Da auch im Liganden-gebundenen Zustand eine (geringe) Genexpression stattfindet, könnte diese Basenendung im Einzelfall eine leicht reduzierte Translationseffizienz zur Folge haben und so den berechneten Schaltfaktor verbessern. Auch die Beobachtung, dass ein Cytosin am Stammende deutlich mit einem guten Schaltfaktor assoziiert ist (Abbildung 3.9.B), kann diese Vermutung untermauern.

4.6 Analyse des Dimers R4-G8

Die Frage, warum R4-G8 sich in seinem Schaltverhalten so stark von den anderen Konstrukten unterscheidet, kann nicht abschließend beantwortet werden. Vergleichbare biophysikalische Parameter wie von R4-G8 sind auch bei weiteren Konstrukten vorhanden:

Tabelle 4.6.A Biophysikalische Parameter ausgewählter Stämme im Vergleich zu R4-G8. Schaltfaktor (SF), Basalexpression (BE) und Expression im Aus-Zustand (Aus).

Name	Stamm P1 [5'-3']	BE [%]	Aus [%]	SF [x-fach]	Tm [°C]	ΔG [kcal/mol]	GC-Gehalt [%]	Entropie [bit]
Derivat 15	AGGU**C**ACC	37,1	3,3	11,4	43,9	-25	63	1,3209
Derivat 16	AGGUG**U**CC	30,5	2,9	10,8	43,9	-25	63	1,3209
Derivat 36	AG**UGAC**CC	29,0	3,0	9,7	43,9	-25	63	1,3209
Derivat 37	AG**UG**GACC	36,3	3,0	12,4	43,9	-25	63	1,3209
R4_C5	A**CUG**G**GA**C	39,7	2,4	16,7	43,9	-25	63	1,3209
R4_G8	AGGUGACC	25,21	0,66	39,5	43,9	-25	63	1,3209

Da die ausgewählten Konstrukte sehr ähnliche biophysikalische Parameter im Vergleich zu R4-G8 aufweisen, kann der Schaltfaktor nicht allein aus diesen Parametern abgeleitet werden und weitere Eigenschaften müssen in Betracht gezogen werden. Eine der Möglichkeiten wurde in der 4. Runde durch das Einbeziehen der Sequenz aufgezeigt. Dennoch ist nicht auszuschließen, dass weitere Parameter, welche in dieser Studie nicht getestet wurden, auch einen Einfluss haben.

Es fällt auf, dass die Derivate von R4-G8 einen schlechteren Schaltfaktor aufweisen als das Konstrukt R4-C5, welches zwar mit R4-G8 eine gewisse Ähnlichkeit aufweist, sich aber in 4 Basen von diesem Konstrukt unterscheidet. Auch die anderen Derivate von R4-G8 zeigen ein ähnliches Bild. Obwohl sie dem Konstrukt R4-G8 sehr ähnlich sind, erreichen sie nicht dessen guten Schaltfaktor. Es wird deutlich, dass jede Base im Stamm von R4-G8 unabdingbar für diesen sehr hohen Schaltfaktor ist, weder ähnliche biophysikalische Parameter noch eine ähnliche Sequenz können diesen erreichen. Dennoch erreicht R4-C5 trotz eines Unterschieds von 4 Basen einen Schaltfaktor von fast 17-fach und das beste Konstrukt der Mutationsanalyse (Derivat 2), welches einen Schaltfaktor von 18,6-fach hat, weißt andere biophysikalische Parameter auf als R4-G8. Man kann davon ausgehen, dass sich der

Stamm R4-G8 mit seiner Sequenz und seinen biophysikalischen Parametern in einem Optimum befindet. Wenn man sich von diesem Optimum entfernt, führt das zu einem schlechteren Schaltfaktor. Jedoch kann mit einer anderen Sequenz und anderen biophysikalischen Parametern womöglich ein neues Optimum erreicht werden und ein ähnlich guter Schaltfaktor zu Stande kommen.

Es lassen sich für das Konstrukt R4-G8 folgende Punkte festhalten:

1. Mit seinen biophysikalischen Parametern befindet sich der P1-Stamm in dem Bereich in dem sich auch andere Riboswitch-Varianten befinden, welche einen hohen Schaltfaktor aufweisen. (Vergleich Heatmap Abbildung 3.12.B):
    - Tm: 43°C bis 50°C
    - ΔG: -24,5 kcal/mol bis -25 kcal/mol
    - GC-Gehalt: 60% bis 75%
2. Die Sequenzmotive von R4-G8 befinden sich deutlich häufiger in Stämmen gut schaltender Konstrukte, wie in der nachfolgenden Tabelle gezeigt ist.

Tabelle 4.6.B: P1-Stamm-Sequenzen, die einen gut schaltenden Riboswitch erzeugt haben

Name	Sequenz	Schaltfaktor
R3-E4	AGGGCAUC	**21,49**
R3-D10	AGGCAUCC	**20,47**
R1-D6	GGUGUGCC	**18,68**
R3-E8	AUUGGGCC	**18,35**
R4-B11	AGGCGUUC	**17,34**
R4-E5	CGGUCGAC	**16,93**
R4-A3	CGGCGCUA	**16,26**
R4-A8	CGGAUCCG	**15,88**
R4-B8	AGGCUCCU	**15,58**
R4-A10	AGUAGGCC	**14,05**
R4-G9	UACGGACC	**14,04**
R4-B9	CGGCUCAG	**13,93**

3. Biophysikalische Parameter und Sequenz befinden sich wahrscheinlich in einem Optimum, wird dieses verändert, stört das den Schaltfaktor.

4.7 Das Einfügen einer Schnittstelle beeinflusst stark die Basalexpression und den Schaltfaktor

Mit dem Einfügen der PacI-Schnittstelle innerhalb des CAA-Spacer verändern sich die Eigenschaften des Konstruktes stark, die Basalexpression wie auch der Expression im Liganden gebundenen Zustand steigen an und der Schaltfaktor sinkt dadurch. Durch das Einfügen der Schnittstelle wurden aus dem CAA-Spacer drei Cytosine entfernt und vier Uracile eingefügt (CAACAACA -> UUAAUUAA).

Die Faltungsvorhersage mit RNA-fold sagt nur eine geringfügige Änderung in der Sekundärstruktur des 5´UTRs voraus, jedoch ändert sich die positionsbezogene Entropie im 5´UTR. Daher ist es nicht unwahrscheinlich, dass sich durch das Einfügen der Schnittstelle doch alternative Sekundärstrukturen in der Zelle ausbilden und die Basalexpression beeinflussen. Ein weiterer CAA-Spacer, welcher vor dem ersten UC eingefügt wurde, reduzierte die Basalexpression wieder und verbesserte den Schaltfaktor. Durch den CAA-Spacer änderte sich die Anzahl der Cytosine in diesem Bereich nicht und ein Uracil wurde durch ein Adenin ersetzt, jedoch bilden sich zwei zusätzliche Basenpaare am Stamm P1 des 5´Aptamers aus und verlängern diesen von fünf auf sieben Basenpaare (Abbildung 4.7). Auch die positionelle Entropie verändert sich wieder und wird schwächer. Was wir also beobachten können, ist, dass das Schaltverhalten eines Riboswitches in großem Maße kontextabhängig ist. Bereits geringfüge Veränderungen genügen, um den Schalter in seiner Basalexpression und seinem Schaltverhalten zu ändern.

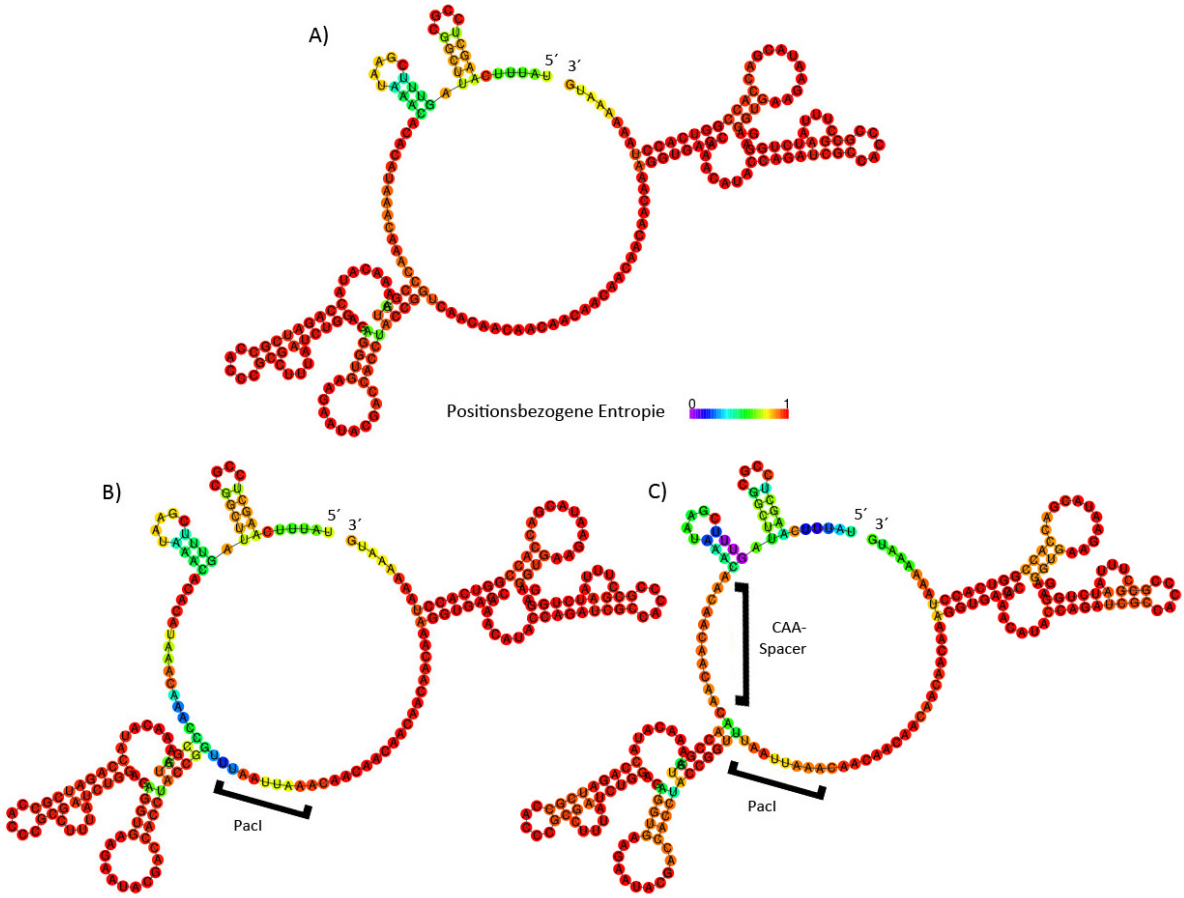


Abbildung 4.7 Faltungsvorhersage des 5´UTRs mit R4-G8 mittels RNA-fold mit und ohne PacI-Schnittstelle und zusätzlichem CAA-Spacer. A) Sekundärstruktur-Vorhersage mit RNA-Fold für das 5´UTR mit R4-G8, ohne PacI-Schnittstelle ab Transkriptionsstartpunkt bis AUG **B)** Sekundärstruktur-Vorhersage mit RNA-Fold für das 5´UTR mit R4-G8, mit PacI-Schnittstelle im CAA-Spacer ab Transkriptionsstartpunkt bis AUG **C)** Sekundärstruktur-Vorhersage mit RNA-Fold für das 5´UTR mit R4-G8, mit PacI-Schnittstelle im CAA-Spacer und zusätzlichem CAA-Spacer vor dem 5´Aptamer, vom Transkriptionsstartpunkt bis zum AUG.

Zudem muss bedacht werden, dass die Basalexpression auch die Wahrscheinlichkeit beinhaltet, mit welcher sich eine RNA in der Zelle zu einer gewissen Sekundärstruktur faltet, denn es wird angenommen, dass die RNA auch *in vivo* in einer Faltungsheterogenität vorliegen kann (Marek et al. 2011), wie auch in den folgenden Kapiteln noch diskutiert werden wird. Liegt also ein Teil der mRNA in der Zelle in einer anderen Sekundärstruktur vor, kann sich dies auf die Expression im Aus-, sowie im An-Zustand auswirken, was wiederrum auf den Schaltfaktor beeinflusst.

4.8 Die Einzelkonstrukte zeigen ein den Dimeren ähnliches Basalexpressionsmuster

Überraschenderweise zeigten die Monomere eine Basalexpression, welche mit den Dimeren vergleichbar ist. Da die Monomere nur noch eine Stammschleife enthielten, wäre eine höhere Basalexpression zu erwarten gewesen. Die gewählten 5´UTRs waren mit 135 nt bzw. 126 nt kürzer als das 5´UTR des Dimers, welches bei R4-G8 und anderen Konstrukten mit einem acht Basenpaar langen P1-Stamm 223 nt lang ist. Die Vergleichsmessung der Monomere mit dem kürzeren 5´UTR verglichen mit dem neun nt längeren 5´UTR zeigte eine Veränderung der Basalexpression und des Schaltfaktors. Bei den drei getesteten Riboswitchen konnte mit dem kürzeren 5´UTR eine höhere Basalexpression und ein niedrigerer Schaltfaktor festgestellt werden. Für dieses Phänomen gibt es zwei mögliche Erklärungen, zum einen könnten die unterschiedlich hohen Basalexpressionen in der tatsächlichen Länge des 5´UTRs begründet sein oder der zur Verlängerung des 5´UTRs eingefügte CAA-Spacer hatte einen Effekt auf die Basalexpression. Durch den Spacer wurden drei zusätzliche Cytosine in das 5´UTR eingesetzt. Da das Dimer mit etwa 223 bp einen längeren 5´UTR hat, aber zum Teil sogar höhere Basalexpressionen aufweist, kann die Länge nicht der Grund für die niedrigere Basalexpression sein.

Das Faltungsprogramm RNA-fold sagt die wahrscheinlichste RNA-Sekundärstruktur vorher und verzeichnet die einzelnen Basen mit entsprechenden Markierungen für die positionsbezogenen Entropien. Das Faltungsprogramm mfold baut zwar genau wie RNA-fold für seine Berechnungen auf den thermodynamischen Parametern des Zuker-Algorithmus auf, es werden aber auch alternative RNA-Sekundärstrukturen mit den entsprechenden ΔG-Werten vorhergesagt (Zuker & Stiegler 1981). Ein niedriger ΔG-Wert begünstigt die Bildung einer RNA-Sekundärstruktur. Ähnlich wie bei Proteinen verläuft die Faltung der RNA hierarchisch (Pan et al. 1997) und ist in bestimmten Teilen auch kontextabhängig: Salzkonzentration, Temperatur und (Nukleinsäure-bindende) Proteine können vor allem *in vivo* einen Einfluss auf die Faltung der RNA-Struktur haben (Marek et al. 2011). Für die beiden TC1-Monomere mit den unterschiedlich langen 5´UTRs werden von mfold die in Abbildung 4.8.A dargestellten alternative Sekundärstrukturen vorhergesagt.

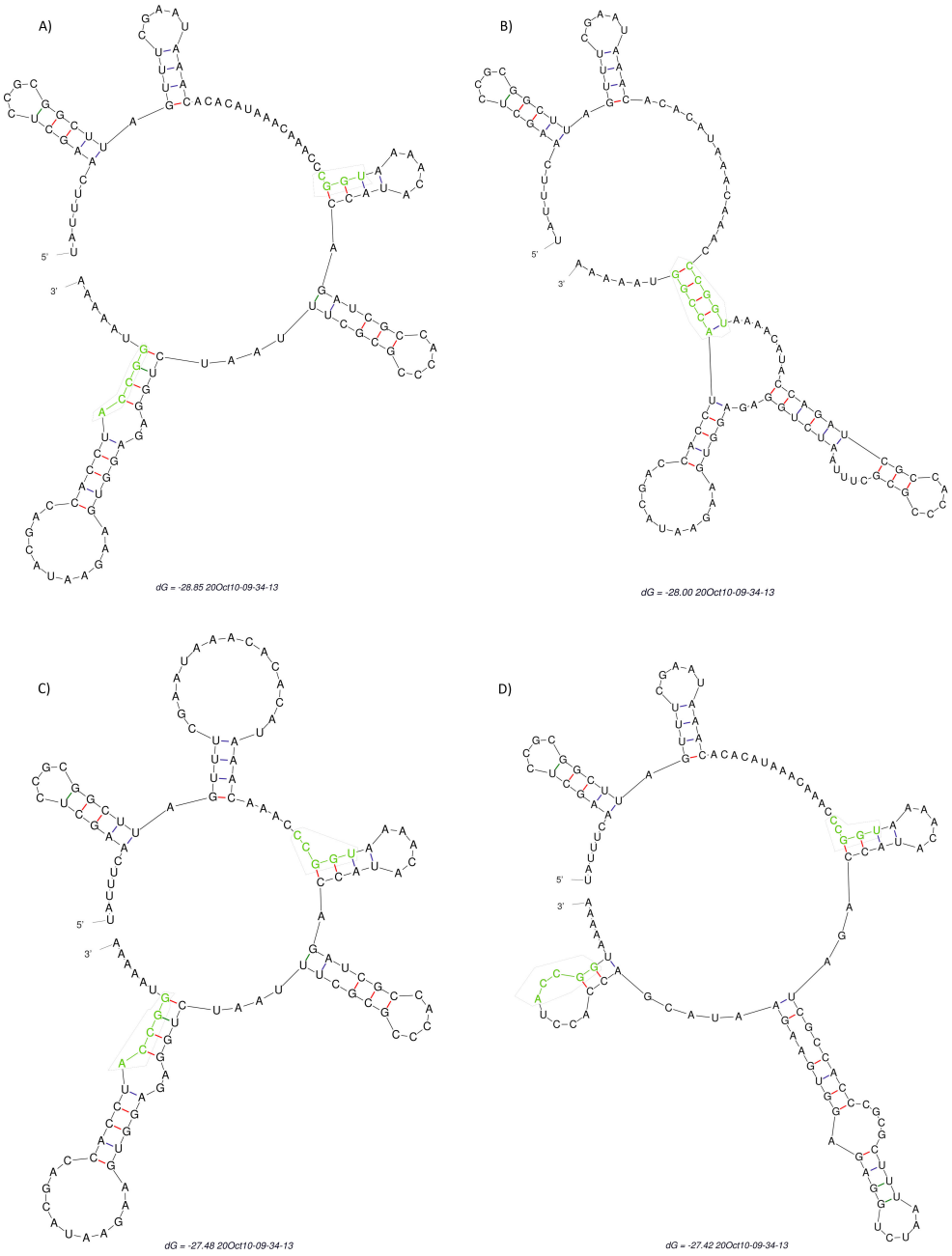


Abbildung 4.8.A Sekundärstrukturvorhersage für TC1 mit kurzem 5´UTR. Sekundärstrukturen mit mfold vorhergesagt, angeordnet nach absteigendem ΔG. Die Basen des P1-Stammes sind grün markiert **A)** ΔG = - 28,85 kcal/mol, die Aptamerstruktur wird nicht korrekt vorhergesagt; **B)** ΔG = -28,00 kcal/mol, die Aptamerstruktur wird korrekt vorhergesagt; **C)** ΔG = -27,48 kcal/mol die Aptamerstruktur wird nicht korrekt vorhergesagt; **D)** ΔG = - 27,42 kcal/mol die Aptamerstruktur wird nicht korrekt vorhergesagt.

Laut mfold-Vorhersage wird für das Monomer TC1 mit dem kürzeren 5´UTR die Aptamerstruktur nicht korrekt vorhergesagt. Hier wird mit dem niedrigsten ΔG-Wert eine andere Sekundärstruktur vorausgesagt. Möglicherweise liegen hier die höhere Basalexpression und der niedrigere Schaltfaktor begründet. Wenn nur ein Teil der gebildeten mRNAs in der Zelle eine stabile Stammschleife

(Aptamer) ausbildet, führt dies zu einer höhere Basalexpression, das Ribosom kann ungehindert seinen Scanning-Prozess fortführen. Für das TC1-Monomer mit dem etwas längeren 5´UTR wird die Aptamerstruktur dagegen in beiden Fällen korrekt vorhergesagt, hier ist die Basalexpression niedriger und der Schaltfaktor höher (Abbildung 4.8.B). Es ist folglich möglich, dass die Aptamerstruktur bei einigen Konstrukten nicht korrekt ausgebildet wird. Dieser Faktor könnte sowohl die Basalexpression als auch das Schaltverhalten des Aptamers beeinflussen.

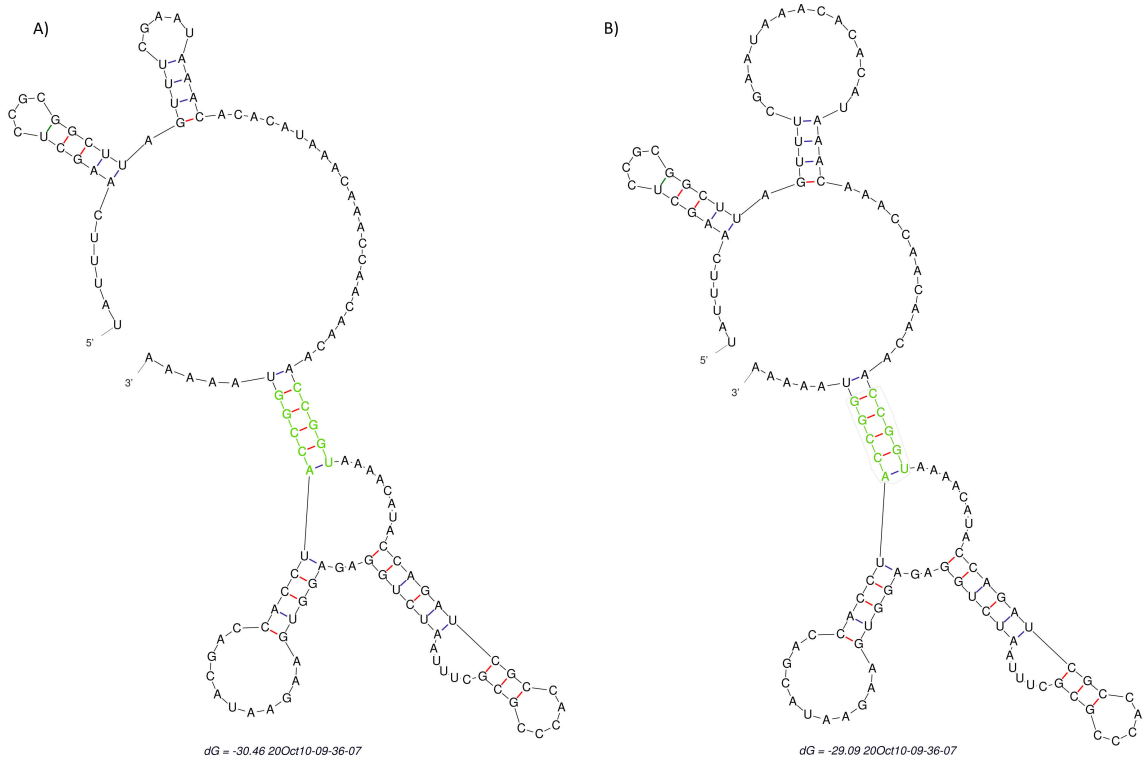


Abbildung 4.8.B TC1 mit längerem 5´UTR. Sekundärstrukturen mit mfold vorhergesagt **A)** ΔG = -30,46 kcal/mol, die Aptamerstruktur wird korrekt vorhergesagt; **B)** ΔG = -29,09 kcal/mol die Aptamerstruktur wird korrekt vorhergesagt.

Die Basalexpression der Monomere ist mit der Basalexpression der entsprechenden Dimere vergleichbar (Abbildung 3.14.C). Die Basalexpression der Monomere liegt zwar auf einem niedrigeren Niveau, verhält sich aber im Vergleich mit den anderen Konstrukten ähnlich. Nur das Konstrukt Derivat 18 hat in diesem Kontext eine vergleichsweise hohe Basalexpression und einen hohen Schaltfaktor. Wenn man bedenkt, dass in den Dimeren das erste Aptamer konstant gehalten und nur das zweite Aptamer verändert wurde, ist diese Beobachtung zu erwarten gewesen. Das erste Aptamer erfüllte seine Aufgabe stets konstant und kann in diesem Kontext als stabile Einheit betrachtet werden.

4.9 Zwei verschiedene Aptamere können miteinander zu einem NOR-Gate fusioniert werden

Für die Generierung logischer Gatter in Zellen wurden bislang zwei verschiedene Aptamere als Tandem eingesetzt (Schneider et al. 2017). Eine Hybridisierung von zwei Aptameren zu einer Sekundärstruktur, die dabei ihre volle Funktion behalten, ist eher ungewöhnlich und in der Literatur bislang noch nicht beschrieben. Bemerkenswert ist zudem, dass sich dieses komplexe Konstrukt nach der Faltungsvorhersage über RNA-fold und mfold richtig faltet. Beide Aptamere werden korrekt gefaltet und können so ihre Bindungseigenschaften behalten. Da erwartet wurde, dass eine so große Stammschleife die Basalexpression stark negativ beeinflussen würde, wurde in Konstrukt „Va22 R3-G6 link“ eine Linker-Region eingefügt. Diese Region besteht aus insgesamt 14 Basen pro Seite, 5 Basenpaare bilden einen kleinen Stamm aus Adenin-Uracil-Basenpaaren und je 9 Basen einen Loop (Abbildung 3.15.A). Dieser Linker hat einen enormen Einfluss auf die Basalexpression, welche sich von 20% auf 60%-erhöht hat. Für diese Erhöhung der Basalexpression gibt es mehrere Erklärungsmöglichkeiten. Zum einen verfügen stark exprimierte Gene in Hefe, wie bereits erwähnt, über eine Adenin-reichen 5´UTR, zum anderen wurde durch das Einfügen des Linkers der Tm des Konstruktes verringert. Stammschleifen mit einem geringeren Tm können von der kleinen Untereinheit des Ribosoms besser überwunden werden als Stammschleifen mit einem höheren Tm (Vega Laso et al. 1993). Für eine umfassende Erklärung der beobachteten Effekte wären jedoch weitere Untersuchungen notwendig, die im Rahmen dieser Arbeit nicht mehr durchgeführt werden konnten.

4.10 Ausblick

Die Analyse der Konstrukte zeigte, dass sowohl biophysikalische Parameter als auch einzelne Sequenzmotive eine wichtige Rolle für das Schaltverhalten des TC-Dimers spielen. Es konnte jedoch nicht abschließend geklärt werden, welche Gegebenheiten den Riboswitch R4-G8 so außergewöhnlich machen. Neben den untersuchten Parametern gibt es noch weitere Faktoren, welche eine Rolle spielen könnten. So kann man nicht davon ausgehen, dass sich alle Konstrukte in gleicherweise in der Zelle falten. Eine Faltungsheterogenität der RNAs wird durch sich wiederholende Sequenzabschnitte und Mutationen gefördert (Marek et al. 2011). Sich wiederholende Sequenzabschnitte können verhindert werden, indem für das Ausgangskonstrukt eines solchen Ansatzes ein Monomer gewählt wird. Auch ist ein RNA-Kontext zu bevorzugen, welcher relativ strukturarm ist und wenig Möglichkeiten bietet, mit der Aptamerstruktur zu interagieren.

Ein größerer Ansatz würde zum einen ein besseres Trainieren der verwendeten *machine learning* Algorithmen und zum anderen ein noch besseres Verständnis der Strukturen ermöglichen. Neben biophysikalischen Parametern und Sequenzen sollte auch der RNA-Kontext mehr beachtet und mögliche Faltungsalternativen berücksichtigt werden, sowie deren Einfluss auf Basalexpression und Schaltverhalten.

Zusammenfassend kann man jedoch festhalten, dass *machine learning* und *deep learning* gute Alternativen zur herkömmlichen Art der Riboswitch-Konstruktion darstellen. Die Programme bieten eine Vielzahl an Lernmöglichkeiten und ermöglichen es, die Konstrukte besser zu verstehen und gezielt weiterzuentwickeln.

5 Material

Die Materialien und Instrumente, die bei dieser Arbeit verwendet wurden, sind in den folgenden Tabellen aufgelistet. Der folgende Abschnitt enthält die Listen für Chemikalien und Reagenzien (Tabelle 5.1.A), Instrumente (Tabelle 5.1.B), Kits und kommerziell erhältliche Systeme (Tabelle 5.1.C), Enzyme und Proteine (Tabelle 5.1.D), Proteinstandards und DNA-Leitern (Tabelle 5.1.E), Zelllinien (Tabelle 5.1.F), Puffer und Lösungen (Tabelle 5.1.G), Oligonukleotide (Tabelle 5.2.A) und Plasmide (5.2.B).

5.1 Übersicht über die verwendeten Labormaterialien

Alle Puffer und Lösungen wurden mit deionisiertem oder Milli-Q Wasser hergestellt. Falls erforderlich, wurden die Puffer und Lösungen durch Autoklavieren bei 121°C und 2 bar für 20 min sterilisiert, außer SCD-Ura-Medium, welches für 10 min sterilisiert wurde. Hitzelabile Substanzen wurden durch Filtration keimfrei gemacht (Sterilfilter, 0,22 µm Porengröße). Oligonukleotide wurden bei Sigma-Aldrich, München, bestellt (entsalzt oder RP1 gereinigt). Die geklonten Sequenzen wurden von Seqlab, Göttingen, mittels Sanger-Sequenzierung und Kapillarelektrophorese analysiert.

Tabelle 5.1.A Chemikalien und Reagenzien

Chemikalien und Reagenzien	Bezugsquelle
Adenin	Roth, Karlsruhe
Agar	Oxoid, Heidelberg
Agarose peqGold Universal	Peqlab, Erlangen
Ammoniumsulfat	Roth, Karlsruhe
Ampicillin	Roth, Karlsruhe
Bromophenolblau	Roth, Karlsruhe
Butanol	Roth, Karlsruhe
Deoxynucleotidtriphosphate (dNTPs)	Peqlab, Erlangen
Dimethylsulfoxid (DMSO)	Peqlab, Erlangen
Dulbecco's Phosphate buffered Saline (PBS)	Life Technologies, USA
Ethanol, p.a.	Merck, Darmstadt
Ethanol, vergällt	VWR, Darmstadt
Ethidiumbromid	Roth, Karlsruhe
Ethylendiamintetraessigsäure (EDTA)	Roth, Karlsruhe
Glucose, wasserfrei	Roth, Karlsruhe
Glycerol, p.a.	Roth, Karlsruhe
Isopropanol, p.a.	VWR, Darmstadt
Leucin	Roth, Karlsruhe
MEM Aminosäuren, 50X	Sigma Aldrich
Magnesiumchlorid	Roth, Karlsruhe

Natriumchlorid (NaCl)	Roth, Karlsruhe
Salzsäure (HCl)	Roth, Karlsruhe
Sodium dodecyl sulfate (SDS), pellets	Roth, Karlsruhe
Triton-X-100	Roth, Karlsruhe
Uracil	Roth, Karlsruhe
Xylene cyanole	Roth, Karlsruhe
Yeast extract	Oxoid, Heidelberg
Yeast nitrogen base (w/o ammonium sulfate)	Difco
Yeast synthetic drop-out (-Ura/Leu/Trp)	Sigma Aldrich

Tabelle 5.1.B Geräte und Hersteller

Gerät	Hersteller
Biogfuge (Fresco17, Pico17, PrimoR)	Heraeus Christ, Osterode
Brutschrank	Heraeus Christ, Osterode
CytoFlex S (Flow Cytometer)	Beckman-Coulter, Krefeld
Feinwaage	Acculab, USA
Fluorolog-3 Spectrofluorometer	Horiba, Darmstadt
Geldokumentation mit UV-Schirm (254 nm /und 312 nm)	INTAS, Göttingen
Heizblock	VWR, Darmstadt
Inkubationsschüttler Multitron	Infors AG, Bottmingen
Magnetrührer IKA RET basic	IKA, Staufen
Milli-Q Wasserentsalzung mit RNase Filter	Millipore, Frankreich
NanoDrop ND-1000 Spektrophotometer	PeqLab, Erlangen
pH-Meter 766 Calimatic	Knick, Berlin
Thyermocycler Peqstar Universal 96	peqLab, Erlangen
Thermomixer comfort	Eppendorf AG, Hamburg
Zentrifugen	Heraeus Christ, Osterode

Tabelle 5.1.C Kits und kommerzielle Systeme

Kits und kommerzielle Systeme	Hersteller
ClonWizard® SV Gel and PCR Clean-Up Systeme	Promega, Walldorf
Frozen-EZ Yeast Transformation II	Zymo Research, USA
PureYield™ Plasmid Miniprep System	Promega, Walldorf
QIAfilter Plasmid Maxi Kit	QIAGEN, Hilden
QIAprep Spin Miniprep Kit	QIAGEN, Hilden
QIAquick Gel Extraction Kit	QIAGEN, Hilden

Tabelle 5.1.D Enzyme und Proteine

Enzyme / Proteine	Hersteller
Polymerasen	
Q5 High-Fidelity DNA-Polymerase [2 U/µl]	New England Biolabs, USA
Taq DNA-Polymerase [5 U/µl]	New England Biolabs, USA
Restriktionsendonukleasen	
AgeI-HF [20 U/µl]	New England Biolabs, USA

NheI-HF [20 U/µl]	New England Biolabs, USA
PacI [10 U/µl]	New England Biolabs, USA
SacII [20 U/µl]	New England Biolabs, USA

Diverse Enzyme und Proteine

Antarctic Phosphatase [5 U/µl]	New England Biolabs, USA
T4 DNA Ligase [400 U/µl]	New England Biolabs, USA
T4 Polynukleotidkinase [10 U/µl]	New England Biolabs, USA

Tabelle 5.6 Größenstandards

Größenstandards	Hersteller
peqGold Ultra Low Range DNA-Leiter II	PeqLab, Erlangen
pegGold 1 kB DNA-Leiter	PeqLab, Erlangen

Tabelle 5.1.E Zelllinien

Name	Genotyp / Anmerkung	Quelle
Prokaryotische Zelllinien		
E. coli DH5α	fhuA2 *lac*(del)U169 *pho*A *gln*V44 Φ80' *lac*Z(del)M15 *gyr*A96 *rec*A1 *rel*A1 *end*A1 *thi*-1 *hsd*R17	Sambrook, Molecular Cloning, 2001
E. coli Top 10	*F- mcrA* Δ(*mrr-hsdRMS-mcrBC*) Φ*80lacZΔM15* Δ *lacX74 recA1 araD139* Δ(*araleu*)7697 *galU galK rpsL (StrR) endA1 nupG*	New England Biolabs
Eukaryotische Zelllinie		
S. cerevisiae RS453α	matα ade2-1 trp1-1 can1-100 leu2-3 his3-1 ura3-52	Stadler u. Sauer (Sauer & Stadler 1993)

Tabelle 5.1.F Puffer und Lösungen

Puffer	Bestandteil	Konzentration
6x DNA loading dye	Tris-HCl pH 7.6 EDTA EDTA Essigsäure Glycerol Bromophenolblau Xylencyanol	40 mM 1 mM 20 mM 50% (v/v) Spatelspitze Spatelspitze
Ampicillin, Stocklösung	Ampicillin in 70% (v/v) EtOH	100 mg/ml
LB-Medium/Platten	Trypton Hefeextrakt NaCl Agar Ampicillin	1% (w/v) 0.5% (w/v) 1% (w/v) 2% (w/v) 100 µg/ml

PBS-Puffer	NaCl	140 mM
	Di-Natrium-Hydrogen-Phosphat-Dihydrat Na2HPO4 . 2H2O	10 mM
	KCl	2,7 mM
	Kalium-Di-Hydrogen-Phosphat KH2PO4	1,8 mM
SCD-Ura-Medium/Platten	50x MEM Amino acids	1x
	Adenin	12 µg/ml
	Ammoniumsulfat	0.55% (w/v)
	Yeast Nitrogen Base	0.2% (w/v)
	Glucose	2% (w/v)
	Agar für Platten	1.8% (w/v)
SOC-medium	Hefeextrakt	0.5%
	Trypton	0.2%
	NaCl	10 mM
	KCl	2.5 mM
	MgCl2	10 mM
	MgSO4	10 mM
	Glucose	20 mM
YPD	Hefeextrakt	1% (w/v)
	Pepton	2% (w/v)
	Glucose	2% (w/v)
	Agar	1.8% (w/v)

5.2 Übersicht über die in dieser Arbeit verwendeten Oligonukleotidsequnezen (Primer)

Für die Klonierung der TC-Dimere wurde stets das gleiche Primer-Design-Prinzip verwendet. Ausgetauscht wurde jeweils nur der Teil des Primers, welcher den veränderten Bereich des P1-Stammes enthielt (N). Die jeweils eingefügten Stämme können in Tabelle 8.5 eingesehen werden. Die weiteren in dieser Arbeit verwendeten Primer sind in Tabelle 5.2.A zu finden.

Fwd. Primer:
5´CTCTTCACCGGTCAACAACAACAACAACAACAACAACAA(N_{6-10})AAACATACCAGATCGCCACC3´

Rev. Primer:
5´CTCTTCGCTAGCCATTTTTT(N_{6-10})GGTGGTCGTATTCTTCACCTC3´

Template-Prime (129 rev)r:
5´AAACATACCAGATCGCCACCCGCGCTTTAATCTGGAGAGGTGAAGAATACGACCACC3´

Tabelle 5.2.A Verwendete Primer

Name	Sequnez 5´-> 3´
R4-G8 um ZS rev	TCCTTTGCTAGCCATTTTTTTTGTTGTTGTTGTTGTTTAATTAAAGGTGACCGGTGGTCGTATTCTTCACCTC
R4G8 um PacI TC1 fwd	GTCACCTTTAATTAAACAACAACAACAACAACAACCGGTAAAACATACCAGATCGCCACCC
R4-G8 do PacI fwd	TCACCTTTAATTAAACAACAACAACAACAACAAAGGTGACCAAACATACCAGATCGCCACC
TC1 PacI/ R4-G8 K rev	TGTTGTTTAATTAAACCGGTAGGTGGTCGTATTCTTCACCTC
R4-G8 PacI für K fwd	TACCGGTTTAATTAAACAACAACAACAACAACAAAGGTGACCAAACATACCAGATCGCCACC
1. TC R2-G7 fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACACACAGAAAAACATACCAGATCGCCACC
1. TC R2-G7 rev	TGTTGTTTAATTAAACACAGAAGGTGGTCGTATTCTTCACCTC
1. TC R4-E8 fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACTCAGCCAGAAACATACCAGATCGCCACC
1. TC R4-E8 rev	TGTTGTTTAATTAATCAGCCAGGGTGGTCGTATTCTTCACCTC
TC1 R4-F5 fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACTCGCGAGCAAACATACCAGATCGCCACC
TC1 R4-F5 rev	GTTGTTGTTTAATTAATCGCGAGCGGTGGTCGTATTCTTCACCTC
R4-G8 Mono fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACAGGTGACCAAACATACCAGATCGCCACC
R4-G8 MCAA k fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACCAACAACAAAGGTGACCAAACATACCAGATCGCCACC
R4-G8 Mono rev	TCCTTTGCTAGCCATTTTTTAGGTGACCGGTGGTCGTATTCTTCACCTC
LG3 TC1 MonoCAA k fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACCAACAACAACCGGTAAAACATACCAGATCGCCACC
LG3 TC1 Mono fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACCCGGTAAAACATACCAGATCGCCACC
LG3 Mono TC1 rev	TCCTTTGCTAGCCATTTTTTACCGGTAGGTGGTCGTATTCTTCAC
TC1 Mono Temp. rev	AGGTGGTCGTATTCTTCACCTCTCCAGATTAAAGCGCGGGTGGCGATCTGGTATGTTT
LG3 TC2 Mono CAA k	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACCAACAACAACTCGAGGCCTAAAACATACCAGATCGCCACC
LG3 TC2 Mono fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACCTCGAGGCCTAAAACATACCAGATCGCCACC
LG3 TC2 Mono rev	TCCTTTGCTAGCCATTTTTTCTCGAGGCCTAGGTGGTCGTATTCTTCACCTC
D26 Mono fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACCAACAACAACGGTGACCAAACATACCAGATCGCCACC
D11 Mono fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACCAACAACAAAGGCGACCAAACATACCAGATCGCCACC
D12 Mono fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACCAACAACAAAGGGGACCAAACATACCAGATCGCCACC
D17 Mono fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACCAACAACAAAGGTGCCCAAACATACCAGATCGCCACC
D18 Mono fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACCAACAACAAAGGTGGCCAAACATACCAGATCGCCACC
D30 Mono fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACCAACAACAAAGGTGAGGAAACATACCAGATCGCCACC
D39 Mono fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACCAACAACAAAGGTGGGCAAACATACCAGATCGCCACC
R4-H1 Mono fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACCAACAACAATACCGAGCAAACATACCAGATCGCCACC
R4-F5 Mono fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACCAACAACAATCGCGAGCAAACATACCAGATCGCCACC
R4-E8 Mono fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACCAACAACAATCAGCCAGAAACATACCAGATCGCCACC
R3-E4 Mono fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACCAACAACAAAGGGCATCAAACATACCAGATCGCCACC
R3-E8 Mono fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACCAACAACAAATTGGGCCAAACATACCAGATCGCCACC
R3-B7 Mono fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACCAACAACAAATCGGTGACAAACATACCAGATCGCCACC
Umgedreht fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACAGGTGACCAAACATACCAGATCGCCACCCGCGCTTTAATC TGGAGAGGTGAAGAATACGACC
um template rev	GGTCGTATTCTTCACCTCTCCAGATTAAAGCGCGGGTGGCGATCTGGTATGTTTTACCGGTTGTTGTTGTTGTTGTTGTTG TTGAGGTGACCGGTGGTCGTATTCTTCACCTCTCC
um rev	TCCTTTGCTAGCCATTTTTTACCGGTAGGTGGTCGTATTCTTCACCTCTC
CAA 5UTR lang Pac K fwd	CAAGCTCCGCGGCTTACAACAACAACAACAACAACAACAACAAACCCCGGTAAAACATACCAGATCGCCACC
Tobra Va22 HB fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACAACAACAAAGGTGACCAAACATACCAGATCGCCGGTGTTTCGGAAAG
TobraVa22HB rev Temp	GGTGGTCGTATTCTTCACCTCTCCAGATTAAAGCGCGGGTAGTGCTCGGTAGAAGCCAGTAGGTCCAACTTTCCGAAACACCG GCGATCTGGTATGTTT
Tobra hunchback rev	TCCTTTGCTAGCCATTTTTTAGGTGACCGGTGGTCGTATTCTTCACCTC
TobraVa22 TC R4-F5 fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACAACAACAATCGCGAGCAAACATACCAGATCGCCGGTG
Tobra Va22 TC R4-F5 rev	CTCCTTTGCTAGCCATTTTTTTCGCGAGCGGTGGTCGTATTCTTCACCTC
Tobra HB linker Template	CGTATTCTTCACCTCTCCAGATTAAAGCGCGGGTAAAAATTGTTGTTGAGTGCTCGGTAGAAGCCAGTAGGTCCAACTTTCCG AAACACCTTGTTGTTGTTTTTGGCGATCTGGTATG
TobraTC R3G6 linker rev	TCCTTTGCTAGCCATTTTTTAGTAATCTGCGGTGGTCGTATTCTTCACCTCTCCAG
TobraTC R3G6 linker fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACAACAACAAAGTAATCTGCAAACATACCAGATCGCCAAAAAC
TobraTC R3 G6 fwd	CAAGCTCCGCGGCTTAGTTTCGAATAAACAACAACAAAGTAATCTGCAAACATACCAGATCGCCGGT

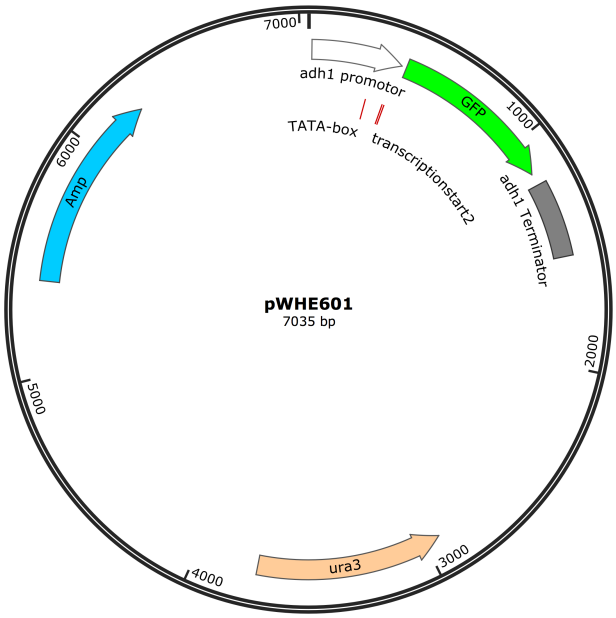
Tabelle 5.2.B Verwendete Plasmide

Bezeichnung	Beschreibung	Herkunft
601	pWHE601	Julia Weigand
IBB	pWEH601* Deletion ATG	Julia Weigand
pGFP3	pWHE601 SacII	Süß 2003
pGFP3_ag	pWHE601 SacII PacI	Süß 2003

5.3 Übersicht über die in dieser Arbeit verwendeten Basisvektoren

5.3.1 pWHE601

Der Vektor pWHE601 (Julia Weigand) diente als Positivkontrolle für das Reportergen-Assay. Der Vektor verfügt über ein ß-Laktamase-Gen, welches Bakterien (hier *E. coli*) eine Resistenz gegen das Antibiotikum Ampicillin verleiht. Die Selektion nach Plasmid-beinhaltenden Kolonien wird so ermöglicht. Das ebenfalls auf dem Plasmid enthaltende URA3-Gen ermöglicht die Selektion auf dem Plasmid in Hefe. URA-3 kodiert für Orotidine 5'-Phosphat (OMP)-Decarboxylase, ein Enzym, das eine Reaktion bei der Synthese von Pyrimidin-Ribonukleotiden katalysiert. Der verwendete Hefestamm verfügt nicht über ein URA3-Gen, was zu einem verlangsamten Zellwachstum führen würde, wenn sich kein Uracil im Medium befindet. Wird das URA3-Gen auf dem Plasmid der Hefe zur Verfügung gestellt, wird die ODCase-Aktivität wiederhergestellt und die Hefen können auf Medien ohne Uracil wachsen, eine Plasmid-abhängige Selektion wird so ermöglicht.
Des Weiteren enthält der Vektor einen 2µ Replikationsursprung (ori) für seine Replikation in Hefezellen. Die Expressionskassette besteht aus einem ADH1-Promotor, mit zwei Transkriptionsstartstellen, sowie -Terminator und einem GFP+-Reportergen.



Sequenz 5´UTR mit Nhe-Schnittstelle (rot)
5´TATTTCAAGCTATACCAAGCATACAATCAACTCCAAGCTAGATCTCTTAAGATGGCTAGCA3´

Abbildung 5.3.1 Vektorkarte des Plasmid pWHE601

5.3.2 pWEH601*

Als Negativkontrolle für das Reportergen-Assay wurde ein Derivat des Vektors pWEH601 verwendet. pWHE601* verfügt im 5´UTR über eine AgeI Schnittstelle, wohingegen pWEH601 über eine AflII-Restriktionsstelle verfügt. Das Startcodon (AUG) wurde durch die Sequenz CTCTTC ersetzt.

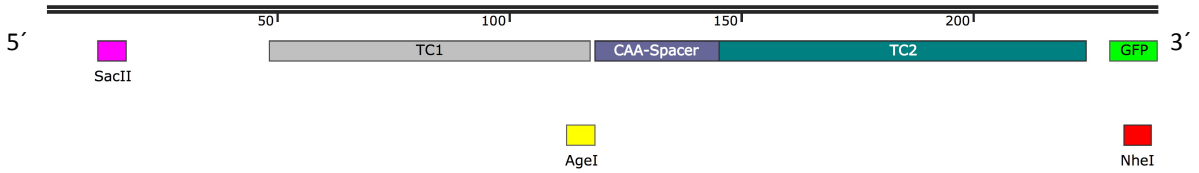
Sequenz 5´UTR pWEH601* mit AgeI (gelb) und NheI-Schnittstelle (rot), ohne ATG:

5´TATTTCAAGCTATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTCTCTTCGCTAGCA3´

5.3.3 pGFP3

Der Vektor pGFP3 diente als Basisvektor für den größten Teil der Klonierungen und wurde von Dr. Julia Weigand und Lara Gorini bereitgestellt. Der Vektor wurde in seinem Grundgerüst bereits 2003 veröffentlicht (Süß 2003). Er verfügt über eine zusätzliche SacII-Schnittstelle im adh1 Promotor, die gegen einen Transkriptionsstartpunkt ausgetauscht worden war. In der 2003 veröffentlichten Studie von Süß *et. al.*, in welcher der Vektor (hier pWEH602 genannt) erstmals verwendet wurde, konnte aber kaum ein Unterschied in der GFP-Expression zum ursprünglich verwendeten Vektor pWH601

gefunden werden. Der Vektor enthält zwei Kopien des TC-Aptamers welche von einem 27 nt langen CAA-Spacer und einem zusätzlichen U am 5´Ende des Spacers separiert sind (Abbildung 5.3.3). Die Kozak-Sequenz (AAAAA-ATG) liegt zwischen dem zweiten TC-Aptamer und der GFP-Gen-Sequenz (Abbildung 5.3.3).



5´TATTTCAAGCTCCGCGGCTTAGTTTCGAATAAACACACATAAACAAACCCGGTAAAACATACCAGAT
CGCCACCCGCGCTTTAATCTGGAGAGGTGAAGAATACGACCACCTACCGGTCAACAACAACAACAACA
ACAACAACAACTCGAGGCCTAAAACATACCAGATCGCCACCCGCGCTTTAATCTGGAGAGGTGAAGAA
TACGACCACCTAGGCCTCGAGAAAAAATG3´

Abbildung 5.3.3 Aufbau des für die Klonierungen relevanten Bereichs des Vektors pGFP3.

5.3.4 pGFP3_ag

Mit dem Vektor pGFP3_ag wurde ein nachfolgender Teil der Konstrukte kloniert. Zusätzlich zu den in pGFP3 vorhandenen Restriktionsschnittstellen enthält dieser Vektor eine PacI Schnittstelle, welche direkt auf das TC1 folgt und 8 Basen des CAA-Spacers ersetzt.

6 Methoden

6.1 Molekularbiologische Methoden

6.1.1 Polymerasekettenreaktion

Die Polymerasekettenreaktion wird im Allgemeinen zur Vervielfältigung von DNA-Molekülen verwendet. Die vervielfältigten DNA-Moleküle wurden im Zuge dieser Arbeit vor Allem für die Amplifikation bestimmter Sequenzen verwendet, welche anschließend in einen verdauten Vektor eingebracht wurden. Des Weiteren wurde die Methode zur Kontrolle von Klonierungsschritten durchgeführt. Üblicherweise dienen als Matrize (Template) der PCR doppelsträngige DNA-Stränge, zum Beispiel Plasmide. Die in dieser Arbeit am häufigsten durchgeführte PCR Methode verwendete aber kein doppelsträngiges DNA-Molekül als Template, sondern ein weiteres Oligonukleotid. Diese Methode wird auch Non-template PCR genannt (ChuWon2010) Die Publikation kann ich nicht finden. Neben dem Vorwärts- (forward; fwd.) und dem Rückwärtsoligonukleotid (reverse; rev.) wurde noch ein weiteres Oligonukleotid eingesetzt welches mit den beiden Oligonukleotiden in Sequenzbereichen überlappt. Dieses Brückenoligonukleotid wurde mit 3 pmol (1/10 der normalen Menge) in die Reaktion gegeben.

In 50 µl Reaktionen wurden 1-100 ng DNA-Matrize mit 30 pmol Vorwärts- und Rückwärts-oligonukleotid, je 20 nmol dNTP und 4U/µl Q5 High-Fidelity DNA-Polymerase mit dem entsprechenden (5x) Q5-Puffer eingesetzt. Das PCR-Programm wurde in einem Thermocycler durchgeführt. Wenn ein doppelsträngiges Templates (Matrizen) vorlag, wurde dieses zunächst initial denaturiert Anschließend erfolgte durch 25 – 35 Zyklen eine Denaturierung, mit anschließender Hybridisierung und Elongation, eine Amplifizierung des Templates. Nach Abschluss der Zyklen wurden eine finale Elongation durchgeführt.

Tabelle 6.1.1 PCR-Programm

Schritt	Dauer (s)	Temperatur (°C)
Initiale Denaturierung	30 - 180	98
Denaturierung	10 - 30	98
Hybridisierung	15 - 30	50-60
Elongation	10 -60	72
Finale Elongation	180	72

6.1.2 Hybridisierung von Oligonukleotiden

Für die Insertion kleinerer Fragmente in einen Vektor, war es oft nicht notwendig eine PCR durchzuführen. Hierfür wurden 2 Oligonukleotide in Vorwärts- und Rückwärtsrichtung bestellt und durch Hybridisierung zu einem Doppelstrang zusammengebracht. Für diesen Reaktionsansatz wurden je 100 pmol Oligonukleotid in MQ gelöst und für 5 min auf 95°C erhitzt. Anschließend wurde die Reaktion langsam auf RT abgekühlt.

6.1.3 PCR Aufreinigung

PCR-Produkte wurden mit Hilfe des entsprechenden Kits der Firma Promega aufgereinigt. Dazu wurde das gesamte Volumen des PCR-Ansatzes nach der PCR in ein dafür vorgesehenes Säulchen überführt und für eine Minute inkubiert. Danach wurde bei 16.000g abzentrifugiert und der Überstand verworfen. Nach zwei Waschschritten mit einer Ehtanol-haltigen Waschlösung und anschließender Zentrifugation wurde die DNA schließlich mit 50 µl MQ eluiert.

6.1.4 Agarose-Gelelektrophorese

Zur Kontrolle und auch zur Aufreinigung von DNA-Fragmenten und Plasmiden wurde eine Agarose-Gelelektrophorese durchgeführt. Das Agarosegel wurde in drei verschiedenen Prozentigkeiten angesetzt: 1%, 2% bzw. 3%. Mit dem 1%igen Agarosegel wurden große DNA-Moleküle, wie Plasmide aufgetrennt, während mit dem 2- bzw. 3%igen Gelen entsprechend kleinere Fragmente aufgetrennt wurden. Die in 1xTAE Puffer erhitzte und gelöste Agarose, wurde in dafür vorgesehene Kammern gegossen, dabei ermöglichte ein Kamm das Einbringen von Ladetaschen. Nach dem Erkalten wurden die Gele in eine dafür vorgesehene, mit TAE-Puffer gefüllte, Laufkammer gelegt und mit den in 6xLadepuffer gemischten DNA-Proben beladen, welche bei 1 – 5 V/cm laufen gelassen wurden. Im Anschluss wurden die Gele in einem Ethidiumbromid-Bad (0,5 µg/ml) für mindestens 10 min gefärbt und dann unter UV-Licht (254 nm) zur Analyse fotografiert. Präparative Agarosegele wurden unter längerwelligem UV-Licht (366 nm) ausgeschnitten, um DNA-Schäden zu vermeiden.

6.1.5 Aufreinigung von Nukleinsäuren aus Gelen

Um DNA-Fragmente (meist Plasmide) aus Agarosegelen aufzureinigen wurden diese nach Ethidiumbromidfärbung unter UV-Licht ausgeschnitten und mittels dem Gel Extractions Kit nach Herstellerangaben gereinigt. Durch doppeltes Waschen der Silica-Membran, für eine bzw. 5 Minuten wurden die Probe entsalzt und anschließend mit 50 µl MQ und Zentrifugation von der Säule entfernt und in einem 1,5 ml Eppi aufgefangen.

6.1.6 Ethanolische Fällung mit Natriumacetat

Da die DNA-Fragmente nach der PCR meistens weiterverarbeitet (mit Restriktionsenzymen verdaut) werden sollten, musste zunächst eine Fällung der Proben durchgeführt werden Enzymen und Pufferbestandteilen zu entfernen.
Die Nukleinsäure-Mischung wurde dazu mit 0,1 Vol. 3 M NaAc pH 6,5 und 2,5 Vol. Ethanol p.a. versetzt und für mindestens 30 min bei -20°C inkubiert. Nach der anschließenden Zentrifugation, welche für 30 min bei 17000x g und 4°C durchgeführt wurde, wurde der Überstand entfernt und das Pellet einmal mit 0,5 Vol. 70% Ethanol gewaschen und für weitere 10 min zentrifugiert. Nach Abziehen des Überstandes wurde das Pellet 5 Minuten bei 37°C getrocknet und dann in einer adäquaten Menge MQ aufgenommen.

6.1.7 Verdau von DNA durch Restriktionsendonukleasen

Mit Hilfe der entsprechenden Enzyme und Puffersysteme wurde doppelsträngige DNA (dsDNA) laut Herstellerangaben endonukleolytisch gespalten. Die einzusetzende Enzymmenge (Units) wurde über Formel 1 berechnet. Falls notwendig wurden die Restriktionsenzyme durch Inkubationen für 20 min bei 65 - 80°C inaktiviert. Um ein inhibieren der Enzyme zu verhindern sollte die Menge der Glycerinkonzentration nicht höher sein als 5% des Reaktionsansatztes.

Formel 1:
$$u = m_p \times \left(\frac{l_r \times n_p}{l_p \times n_r}\right) \div h$$

u = Enzymeinheiten; mp = Masse der zu spaltenden dsDNA [ng]; lr = Länge der Referenz-DNA [bp]; np = Anzahl der Schnittstellen des Enzyms in der zu schneidenden DNA; lp = Länge der zu spaltenden DNA [bp]; nr = Anzahl der Schnittstellen des Enzyms in der Referenz-DNA; h = Dauer des Restriktionsverdau in Stunden.

6.1.8 Ligation von DNA-Molekülen

Für die Ligation von vorverdautem Plasmid und dem gewünschten DNA-Insert wurden 25 ng Plasmid mit 5-fach molaren Überschuss an verdautem DNA-Insert, 1 mM ATP und 400 U T4-DNA-Ligase im T4-Ligations Puffer und einem endgültigen Reaktionsvolumen von 20 µl für 1 h bei RT inkubiert. Anschließend wurde der gesamte Ligationsansatz in CaCl2-kompetente Top10 transformiert.

6.1.9 Konzentrationsbestimmung von DNA

Die Konzentration von Nukleinsäuren wurde spektralphotometrisch mit dem NanoDrop ND-1000 bei 260 nm bestimmt. Da die DNA in MQ gelöst war, wurde dieses als Referenz genutzt.

6.1.10 Klonierungen der Tandem-Konstrukte für das maschinelle Lernen

Da für das machine learning viele Konstrukte getestet wurden, wurde das Klonierungsschema vereinfacht und standardisiert. Es wurde eine Non-template PCR durchgeführt, bei welcher die Vorwärts und Rückwärtsprimer die jeweils veränderten Stamm-Sequenzen des zweiten Tetrazyklin-Aptamers enthielten, während für die PCR immer das Selbe Template-Oligonukleotid verwendet wurde (Kapitel 5.2). In 96-Well PCR-Tubes wurden die Vorwärts- und Rückwärtsoligonukleotide vorgelegt und mit einem Master-Mix auf ein Reaktionsvolumen von 50 µl aufgefüllt. Die Zusammensetzung des PCR-Ansatzes ist in Tabelle 6.1.10.A aufgeführt, das PCR-Programm in Tabelle 6.1.10.B.

Tabelle 6.1.10.A PCR-Ansatz für die Klonierung der Tandem-Konstrukte

Konzentration	Reagenz	Volumen / µl	Endkonzentration
100%	H2O	30	
5x	Q5-Puffer	10	1x
100%	DMSO	1,5	3%
10mM	dNTPs	1	200 µM
10 ng	Template	1	1 ng/µl
10 µM	fwd. Primer	3	600 nM
10 µM	rev. Primer	3	600 nM
2 U/ µl	Q5-Polymerase	0,5	1 U

Tabelle 6.1.10.B PCR-Programm für die Klonierung der Tandem-Konstrukte

Schritt	Dauer (s)	Temperatur (°C)
Initiale Denaturierung	30	98
Denaturierung	10	98
Hybridisierung	15	55
Elongation	10	72
Finale Elongation	180	72

Nach der PCR wurde anschließend mittels ethanolischer Natriumacetat-Fällung aufgereinigt. Dazu wurde der Reaktionsansatz mit 5 µl Natriumacetat und 125 µl 100%iges Ethanol gemischt und für 1 h bei -20°C inkubiert. Die anschließenden Zentrifugationsschritte erfolgten in der Tisch-Zentrifuge bei 17.000 g für 30 min. Nach dem Waschen in 70% Ethanol und dem Trockenen des Pellets, wurden diese in 50 µl MQ aufgenommen. Für den Verdau mit den Restriktionsenzymen NheI-HF und AgeI-HF wurden 30 µl des PCR-Produktes in eine neue 96-Well Platte überführt und mit jeweils einem µl der Enzyme, 5 µl CutSmart®-Puffer und 13 µl MQ für 3 h bei 37°C inkubiert. Der Vektor pGFP3 wurde ebenfalls mit jeweils einem µl der Enzyme, allerdings ÜN, bei 37°C inkubiert und im Anschluss über eine Gelextraktion gereinigt. Die Ligation von Insert und Vektor erfolgte, ebenso wie die Transformation in E. coli Zellen, ebenfalls in 96-Well Platten. Dazu wurde 1 µl verdautes Insert vorgelegt und mit 19 µl, den Vektor enthaltenen, Mastermix vorsichtig gemischt. Die Konzentration des Inserts wurde nicht bestimmt und auf die Bestimmung der Plasmid-Insert Ration wurde verzichtet.

6.2 Methoden mit *E. coli*

6.2.1 Anzucht, Ernte und Lagerung

Die Anzucht von *E. coli* erfolgte üblicherweise in flüssigem oder auf festem LB-Medium, supplementiert mit dem entsprechendem Antibiotikum, meist Ampicillin. Die Inkubationszeit betrug ca. 16 h bei einer Temperatur von 37°C. Flüssigkulturen wurden konstant bei 150 rpm geschüttelt. Das Animpfen geschah immer von frisch ausgestrichenen Platten bzw. von Vorkulturen in einem Verhältnis von 1:500. Kleine Volumina Flüssigkulturen wurden bei 17000x g für eine Minute mittels Zentrifugation geerntet, größere Flüssigkultur-Ansätze bei 5800x g (50-500 ml)
Die Kurzzeitlagerung (< 4 Wochen) erfolgte auf Petrischalen bei 4°C. Für die Langzeitlagerung wurden Glycerin-Stocks angelegt. Hierfür wurde 1 ml Flüssigkultur mit 15% (v/v) Glycerin versetzt und bei -80°C gelagert.

6.2.2 Herstellung chemokompetenter E. coli mittels CaCl2

Zunächst wurde der gewünschte E. coli Stamm aus dem gefrorenen Stock entnommen und auf einer LB-Platte ausgestrichen, die Inkubationszeit betrug 16 h bei 37°C. Anschließend wurde eine Kolonie der Platte gepickt und in eine 4 ml Übernachtkultur (ÜNK) überimpft. Am Folgetag wurde mit dieser ÜNK 300 ml LB-Medium angeimpft. Das Wachstum erfolgte bei 37°C und 120 rpm bis zu einer oD_{600} von 0,4 - 0,5. Der Kolben wurde anschließend für 10 Minuten auf Eis inkubiert. In vorgekühlten 50 ml Reaktionsgefäßen wurden die Zellen geerntet, mit 100 ml eiskaltem 0,1 M $CaCl_2$ gewaschen und final in 20 ml eiskaltem 0,1 M $CaCl_2$ aufgenommen.
Die Bestimmung der Kompetenz der chemokompetenten E. coli-Zellen wurde nach Hanahan et al. (Hanahan 1983) durchgeführt unter Verwendung von des Vektors pUC19.

6.2.3 Transformation von Ligationsansätzen

Für die Transformation wurden die bei -80°C gelagerten E. coli-Zellen langsam auf Eis aufgetaut. Je 20 µl Ligationsansatz wurden mit 100 µl Top10 Zellen gemischt und für 1 h auf Eis inkubiert. Anschließend wurden die Zelle 50 sec bei 42°C im Wasserbad inkubiert (Hitzeschock) und wieder 10 min auf Eis inkubiert. Nach Zugabe von 900 µl LB-Medium wurden die Kultur eine Stunde bei 37°C und 850 rpm geschüttelt, bevor sie auf Selektiv-Agar ausplattiert und für 16 h bei 37°C inkubiert wurde.

6.2.4 Retransformation von bereits präparierten Plasmiden

Für bereits präparierten Plasmide konnte das Transformationsprotokoll, auf Grund ihrer leichteren Transformierbarkeit, an einigen Stellen gekürzt werden. Meist wurde 1 µl Plasmid mit 50 µl kompetenten Zellen vorsichtig gemischt und für 30 Minuten auf Eis inkubiert. Der Hitzeschock erfolgte ebenfalls bei 42°C im Wasserbad für 50 sec mit anschließender 10-minütiger Inkubation auf Eis. Nach Zugabe von 900 µl LB-Medium wurden die Kultur für 30 Minuten bei 37°C und 850 rpm geschüttelt, und anschließend auf Selektiv-Agar ausplattiert und für 16 h bei 37°C inkubiert.

6.2.5 Präparation von Plasmiden

Nach der Ligation mit anschließender Transformation bzw. der Retransformation von Plasmiden wurden einzelne *E. coli*-Kolonien von der LB-Amp-Platte gepickt und in 4 ml flüssigem LB-Amp Medium für 16 h bei 37°C und 140 rpm in Reaktionsgläsern inkubiert. Nach der Inkubationszeit wurde der gesamte Ansatz pelletiert und in 600 ml MQ aufgenommen. Die Präparation erfolgte dann nach dem von Promega angegebenen Mini-Prep Protokoll.

6.3 Methoden mit *S. cerevisiae*

6.3.1 Herstellung kompetenter Hefezellen

Die Hefezellen des Stamms RS453α wurden dem Stock entnommen und auf einer YEPD-Platte ausgestrichen, vereinzelt und für 72 h bei 30°C inkubiert. Anschließend wurde eine Kultur gepickt und in einer 4 ml YPD ÜNK bei 30°C und 140 rpm inkubiert. Mit dieser ÜNK wurde am nächsten Tag 10 ml YPD in einer 1:10 Verdünnung angeimpft und bis zu einer OD zwischen 0,8 und 1,0 bei 500 rpm und 30°C im Inkubator wachsen gelassen. Die Zellen wurden anschließend bei 500 g für 4 min Pelletiert und in 10 ml EZ1 Solution aufgenommen um im darauffolgenden Arbeitsschritt wieder pelletiert zu werden. Dann wurden die Hefezellen in 1 ml EZ2 Solution aufgenommen und langsam auf -80°C heruntergekühlt.

6.3.2 Transformation kompetenter Hefezellen mit dem Frozen Kit von Zymo

Die kompetenten Hefezellen wurden aufgetaut und jeweils 10 µl Zellen wurden mit ca. 100 ng Plasmid und 100 µl EZ3 gemischt. Die Inkubation erfolgte bei 30°C. Während der Inkubationszeit Zeit wurden die Zellen ein - bis zweimal mit dem Vortexer gemischt. Nach einer Stunde wurde der gesamte Reaktionsansatz auf SCD-Ura Selektivmedium ausplattiert und für 72 h bei 30°C in einem befeuchteten Inkubator inkubiert.

6.3.3 Anzucht und Messungen der Zellen für Zytometrie

Die mit dem gewünschten Plasmid transformierten Zellen wurden auf Selektiv-Agar ausplattiert und für 72 h inkubiert. Anschließend wurden die Zellen in flüssig Selektiv-Medium (SCD-Ura) überführt.

Dazu wurde mit einer sterilen Pipettenspitze mehrere Male über die Platte gestrichen um verschiedene Kolonien aufzunehmen, die Kolonien wurden 1,5 ml Selektiv-Medium gelöst. Die Inkubation erfolgten in 24-Well Platten, ÜN bei 450 rpm und 30 °C. Am darauffolgenden Tag wurden die Zellen 1:1000 verdünnt und mit und ohne 250 µM Tetrazyklin für 24 h inkubiert.

Für die Messung wurden die 20 µl Zellen in 180 µl PBS-Puffer, in 96-Well Platten verdünnt. Die zytometrischen Messungen wurden auf einem CytoFlex S-Gerät von Beckman Coulter durchgeführt. Dieser ist mit einem 561-nm-Laser zur Anregung von GFP ausgestattet. Das Emissionslicht wurde bei 510/20 nm bandpassgefiltert. Für die Messungen wurde kein Gating durchgeführt, alle Datenpunkte flossen in die Berechnung des Mittelwerts ein. Pro Versuch und Konstrukt wurden je zwei Wells mit und ohne Tetrazyklin gemessen. Aus diesen Messungen wurde ein Mittelwert gebildet und dieser für die weiteren Berechnungen verwendet. Für alle gemessenen Konstrukte sowie die Positiv-Kontrolle (*S. cerevisiae* transformiert mit Plasmid pWHE601) wurde vor der Berechnung des Schaltfaktors der jeweilige Hintergrund (S. cerevisiae transformiert mit Plasmid IBB ohne AUG), ebenfalls inkubiert mit und ohne 250 µM Tetrazyklin abgezogen. Das Schaltverhalten der Konstrukte wird durch ihren Schaltfaktor (SF/x-fach, Quotient aus An-Zustand (hier: ohne Ligand) und Aus-Zustand (hier: mit Liganden) und die basale Expression (BE, %, Fluoreszenz des An-Zustandes im Verhältnis zur Fluoreszenz eines Konstrukts ohne Aptamer-Insertion) beschrieben.

7 Literaturverzeichnis

Agrawal, N. et al., 2003. SUMMARY. *Microbiology and Molecular Biology Reviews*, 67(4), pp.657–685.

Alipanahi, B. et al., 2015. Predicting the sequence specificities of DnA- and RnA-binding proteins by deep learning. *Nature Biotechnology*, 33(8), pp.831–838.

Altmann, M. & Linder, P., 2010. Power of Yeast for Analysis of Eukaryotic Translation Initiation. *Journal of Biological Chemistry*, 285(42), pp.31907–31912.

An, C.-I., Trinh, V.B. & Yokobayashi, Y., 2006. Artificial control of gene expression in mammalian cells by modulating RNA interference through aptamer-small molecule interaction. *RNA*, 12(5), pp.710–716.

Andronescu, M. et al., 2013. The Determination of RNA Folding Nearest Neighbor Parameters. In *Methods in Molecular Biology*. Methods in Molecular Biology. Totowa, NJ: Humana Press, pp. 45–70.

Bartley, B.A. et al., 2017. Synthetic Biology: Engineering Living Systems from Biophysical Principles. *Biophysj*, 112(6), pp.1050–1058.

Beilstein, K. et al., 2014. Conditional Control of Mammalian Gene Expression by Tetracycline-Dependent Hammerhead Ribozymes. *ACS Synthetic Biology*, 4(5), pp.526–534.

Berens, C., Groher, F. & Suess, B., 2015. RNA aptamers as genetic control devices: the potential of riboswitches as synthetic elements for regulating gene expression. *Biotechnology Journal*, 10(2), pp.246–257.

Berens, C., Thain, A. & Schroeder, R., 2001. A tetracycline-binding RNA aptamer. *Bioorganic & Medicinal Chemistry*, 9(10), pp.2549–2556.

Biau, G., 2012. Analysis of a Random Forests Model. *Journal of Machine Learning Research*, pp 1063-1095.

Bocobza, S. et al., 2007. Riboswitch-dependent gene regulation and its evolution in the plant kingdom. *Genes & development*, 21(22), pp.2874–2879.

Boussebayle, A., Groher, F. & Suess, B., 2019. RNA-based Capture-SELEX for the selection of small molecule-binding aptamers. *METHODS*, 161, pp.10–15.

Boussebayle, A., Torka, D., et al., 2019. Next-level riboswitch development—implementation of Capture-SELEX facilitates identification of a new synthetic riboswitch. *Nucleic Acids Research*, 47(9), pp.4883–4895.

Brantl, S., 2007. Regulatory mechanisms employed by cis-encoded antisense RNAs. *Current opinion in microbiology*, 10(2), pp.102–109.

Breaker, R.R., 2011. Prospects for riboswitch discovery and analysis. *Molecular cell*, 43(6), pp.867–879.

Breaker, R.R., 2012. Riboswitches and the RNA World. *Cold Spring Harbor Perspectives in Biology*, 4(2), pp.a003566–a003566.

Brophy, J.A.N. & Voigt, C.A., 2014. Principles of genetic circuit design. *Nature Methods*, 11(5), pp.508–520.

Callura, J.M., Cantor, C.R. & Collins, J.J., 2012. Genetic switchboard for synthetic biology applications. *Proceedings of the National Academy of Sciences*, 109(15), pp.5850–5855.

Camacho, D.M. et al., 2018. Next-Generation Machine Learning for Biological Networks. *Cell*, 173(7), pp.1581–1592.

Cameron, D.E., Bashor, C.J. & Collins, J.J., 2014. PERSPECTIVES. *Nature Publishing Group*, 12(5), pp.381–390.

Chappell, J. et al., 2015. A renaissance in RNA synthetic biology: new mechanisms, applications and tools for the future. *Current opinion in chemical biology*, 28, pp.47–56.

Cheah, M.T. et al., 2007. Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature*, 447(7143), pp.497–500.

Ching, T. et al., 2018. Opportunities and obstacles for deep learning in biology and medicine. *Journal of the Royal Society, Interface*, 15(141).

Collins, J.A. et al., 2007. Mechanism of mRNA destabilization by the glmS ribozyme. *Genes & development*, 21(24), pp.3356–3368.

Crick, F., 1970. Central dogma of molecular biology. *Nature*, 227(5258), pp.561–563.

Davuluri, R.V. et al., 2000. CART classification of human 5' UTR sequences. *Genome research*, 10(11), pp.1807–1816.

Desai, S.K. & Gallivan, J.P., 2004. Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. *Journal of the American Chemical Society*, 126(41), pp.13247–13254.

Duchardt-Ferner, E. et al., 2010. Highly modular structure and ligand binding by conformational capture in a minimalistic riboswitch. *Angewandte Chemie (International ed. in English)*, 49(35), pp.6216–6219.

Dvir, S. et al., 2013. Deciphering the rules by which 5'-UTR sequences affect protein expression in yeast. *Proceedings of the National Academy of Sciences*, 110(30), pp.E2792–801.

EC, Vancompernolle, K. & Ball, P., 2005. Synthetic Biology Applying Engineering to Biology. *European Commission*, pp.1–44.

Ellington, A.D. & Szostak, J.W., 1990. In vitro selection of RNA molecules that bind specific ligands. *Nature*, 346(6287), pp.818–822.

Faber, F.A. et al., 2016. Machine Learning Energies of 2 Million Elpasolite (ABC_{2}D_{6}) Crystals. *Physical review letters*, 117(13), p.135502.

Falaleeva, M. et al., 2017. C/D-box snoRNAs form methylating and non-methylating ribonucleoprotein complexes: Old dogs show new tricks. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 39(6).

Förster, U. et al., 2011. Conformational dynamics of the tetracycline-binding aptamer. *Nucleic Acids Research*, 40(4), pp.1807–1817.

Furuichi, Y., LaFiandra, A. & Shatkin, A.J., 1977. 5'-Terminal structure and mRNA stability. *Nature*, 266(5599), pp.235–239.

Gawronski, A.R. & Turcotte, M., 2014. RiboFSM: frequent subgraph mining for the discovery of RNA structures and interactions. *BMC bioinformatics*, 15 Suppl 13, p.S2.

Green, A.A. et al., 2014. Toehold Switches: De-Novo-Designed Regulators of Gene Expression. *Cell*, 159(4), pp.925–939.

Groher, A.-C. et al., 2018. Tuning the Performance of Synthetic Riboswitches using Machine Learning. *ACS Synthetic Biology*, 8(1), pp.34–44.

Groher, F. & Suess, B., 2016. In vitro selection of antibiotic-binding aptamers. *METHODS*, 106(C), pp.42–50.

Groher, F. & Suess, B., 2014. Synthetic riboswitches - A tool comes of age. *Biochimica et biophysica acta*, 1839(10), pp.964–973.

Groher, F. et al., 2018. Riboswitching with ciprofloxacin—development and characterization of a novel RNA regulator. *Nucleic Acids Research*, 46(4), pp.2121–2132.

Hamilton, R., Watanabe, C.K. & de Boer, H.A., 1987. Compilation and comparison of the sequence context around the AUG startcodons in Saccharomyces cerevisiae mRNAs. *Nucleic Acids Research*, 15(8), pp.3581–3593.

Hanahan, D., 1983. Studies on transformation of Escherichia coli with plasmids. *Journal of Molecular Biology*, 166(4), pp.557–580.

Hanson, S. et al., 2003. Tetracycline-aptamer-mediated translational regulation in yeast. *Molecular Microbiology*, 49(6), pp.1627–1637.

Hanson, S., Bauer, G. & Fink, B., 2005. Molecular analysis of a synthetic tetracycline-binding riboswitch. *RNA*, 11(4), pp.503–511.

Hinnebusch, A.G., 2011. Molecular Mechanism of Scanning and Start Codon Selection in Eukaryotes. *Microbiology and Molecular Biology Reviews*, 75(3), pp.434–467.

Hinnebusch, A.G., 2014. The Scanning Mechanism of Eukaryotic Translation Initiation. *Annual Review of Biochemistry*, 83(1), pp.779–812.

Hollands, K. et al., 2012. Riboswitch control of Rho-dependent transcription termination. *Proceedings of the National Academy of Sciences*, 109(14), pp.5376–5381.

Hyndman, R.J., 2006. Another look at measures of forecast accuracy. pp.1–4.

Ideker, T. et al., 2001. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, 292(5518), pp.929–934.

Isaacs, F.J. et al., 2004. Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotechnology*, 22(7), pp.841–847.

Jeong, H. et al., 2000. The large-scale organization of metabolic networks. *Nature*, 407(6804), pp.651–654.

Kaufman, R.J., 1994. Control of gene expression at the level of translation initiation. *Current opinion in biotechnology*, 5(5), pp.550–557.

Kazanov, M.D., Vitreschak, A.G. & Gelfand, M.S., 2007. Abundance and functional diversity of riboswitches in microbial communities. *BMC Genomics*, 8, p.347.

Kiledjian, M., 2018. Eukaryotic RNA 5'-End NAD+ Capping and DeNADding. *Trends in cell biology*, 28(6), pp.454–464.

Kim, D.-S. et al., 2005. An artificial riboswitch for controlling pre-mRNA splicing. *RNA*, 11(11), pp.1667–1677.

Klauser, B. et al., 2012. Post-transcriptional Boolean computation by combining aptazymes controlling mRNA translation initiation and tRNA activation. *Molecular BioSystems*, 8(9), pp.2242–2248.

Kniskern, P.J. et al., 1986. Unusually high-level expression of a foreign gene (hepatitis B virus core antigen) in Saccharomyces cerevisiae. *Gene*, 46(1), pp.135–141.

Kozak, M., 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, 44(2), pp.283–292.

Kozak, M., 2002. Pushing the limits of the scanning mechanism for initiation of translation. *Gene*, 299(1-2), pp.1–34.

Li, J. et al., 2017. Nucleotides upstream of the Kozak sequence strongly influence gene expression in the yeast S. cerevisiae. *Journal of biological engineering*, 11, p.25.

Lin, Z. & Li, W.-H., 2012. Evolution of 5' untranslated region length and gene expression reprogramming in yeasts. *Molecular biology and evolution*, 29(1), pp.81–89.

Louppe, G., 2015. Understanding Random Forest. *University of Liège Faculty of Applied Sciences*

Department of Electrical Engineering & Computer Science, pp.1–223.

Mandal, M. et al., 2004. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science*, 306(5694), pp.275–279.

Marek, M.S., Johnson-Buck, A. & Walter, N.G., 2011. The shape-shifting quasispecies of RNA: one sequence, many functional folds. *Physical chemistry chemical physics : PCCP*, 13(24), pp.11524–11537.

Mathews, D.H., 2006. Revolutions in RNA secondary structure prediction. *Journal of Molecular Biology*, 359(3), pp.526–532.

McAdams, H.H. & Arkin, A., 2000. Towards a circuit engineering discipline. *Current biology : CB*, 10(8), pp.R318–20.

McAdams, H.H. & Shapiro, L., 1995. Circuit simulation of genetic networks. *Science*, 269(5224), pp.650–656.

McKeague, M., Wong, R.S. & Smolke, C.D., 2016. Opportunities in the design and application of RNA for gene expression control. *Nucleic Acids Research*, 44(7), pp.2987–2999.

Müller, M. et al., 2006. Thermodynamic characterization of an engineered tetracycline-binding riboswitch. *Nucleic Acids Research*, 34(9), pp.2607–2617.

Pan, J., Thirumalai, D. & Woodson, S.A., 1997. Folding of RNA involves parallel pathways. *Journal of Molecular Biology*, 273(1), pp.7–13.

Pesole, G. et al., 2001. Structural and functional features of eukaryotic mRNA untranslated regions. *Gene*, 276(1-2), pp.73–81.

Ruff, K.M. & Strobel, S.A., 2014. Ligand binding by the tandem glycine riboswitch depends on aptamer dimerization but not double ligand occupancy. *RNA*, 20(11), pp.1775–1788.

Saito, K., Green, R. & Buskirk, A.R., 2020. Translational initiation in E. coli occurs at the correct sites genome-wide in the absence of mRNA-rRNA base-pairing. *eLife*, 9.

Sauer, N. & Stadler, R., 1993. A sink-specific H+/monosaccharide co-transporter from Nicotiana tabacum: cloning and heterologous expression in baker's yeast. *The Plant journal : for cell and molecular biology*, 4(4), pp.601–610.

Schmidt, J. et al., 2017. Predicting the Thermodynamic Stability of Solids Combining Density Functional Theory and Machine Learning. *Chemistry of Materials*, 29(12), pp.5090–5103.

Schneider, C. & Suess, B., 2015. Identification of RNA aptamers with riboswitching properties. *METHODS*, pp.1–7.

Schneider, C. et al., 2017. ROC'n'Ribo: Characterizing a Riboswitching Expression System by Modeling Single-

Cell Data. *ACS Synthetic Biology*, 6(7), pp.1211–1224.

Serganov, A. & Nudler, E., 2013. A Decade of Riboswitches. *Cell*, 152(1-2), pp.17–24.

Sharma, V., Nomura, Y. & Yokobayashi, Y., 2008. Engineering Complex Riboswitch Regulation by Dual Genetic Selection . *Journal of the American Chemical Society*, pp.1–6.

Slomovic, S., Pardee, K. & Collins, J.J., 2015. Synthetic biology devices for in vitro and in vivo diagnostics. *Proceedings of the National Academy of Sciences*, 112(47), pp.14429–14435.

Soukup, G.A. & Breaker, R.R., 1999. Engineering precision RNA molecular switches. *Proceedings of the National Academy of Sciences of the United States of America*, 96(7), pp.3584–3589.

Stoltenburg, R., Reinemann, C. & Strehlitz, B., 2007. SELEX--a (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomolecular engineering*, 24(4), pp.381–403.

Suess, B. et al., 2003. Conditional gene expression by controlling translation with tetracycline-binding aptamers. *Nucleic Acids Research*, 31(7), pp.1853–1858.

Tang, J. & Breaker, R.R., 1997. Rational design of allosteric ribozymes. *Chemistry & Biology*, 4(6), pp.453–459.

Tuerk, C. & Gold, L., 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, 249(4968), pp.505–510.

Tuleuova, N. et al., 2008. Modulating endogenous gene expression of mammalian cells via RNA-small molecule interaction. *Biochemical and biophysical research communications*, 376(1), pp.169–173.

Valenzuela, P. et al., 1979. Nucleotide sequence of the gene coding for the major protein of hepatitis B virus surface antigen. *Nature*, 280(5725), pp.815–819.

Vega Laso, M.R. et al., 1993. Inhibition of translational initiation in the yeast Saccharomyces cerevisiae as a function of the stability and position of hairpin structures in the mRNA leader. *The Journal of biological chemistry*, 268(9), pp.6453–6462.

Venkata Subbaiah, K.C. et al., 2019. Mammalian RNA switches: Molecular rheostats in gene regulation, disease, and medicine. *Computational and structural biotechnology journal*, 17, pp.1326–1338.

Wachsmuth, M. et al., 2013. De novo design of a synthetic riboswitch that regulates transcription termination. *Nucleic Acids Research*, 41(4), pp.2541–2551.

Wachter, A. et al., 2007. Riboswitch Control of Gene Expression in Plants by Splicing and Alternative 3' End Processing of mRNAs. *The Plant Cell*, 19(11), pp.3437–3450.

Weigand, J.E. et al., 2010. Mechanistic insights into an engineered riboswitch: a switching element which confers riboswitch activity. *Nucleic Acids Research*, 39(8), pp.3363–3372.

Weigand, J.E. et al., 2007. Screening for engineered neomycin riboswitches that control translation initiation. *RNA*, 14(1), pp.89–97.

Werstuck, G. & Green, M.R., 1998. Controlling gene expression in living cells through small molecule-RNA interactions. *Science*, 282(5387), pp.296–298.

Westerhoff, H.V. & Palsson, B.O., 2004. The evolution of molecular biology into systems biology. *Nature Biotechnology*, 22(10), pp.1249–1252.

Wieland, M. & Hartig, J.S., 2008. Improved aptazyme design and in vivo screening enable riboswitching in bacteria. *Angewandte Chemie (International ed. in English)*, 47(14), pp.2604–2607.

Winkler, W.C. et al., 2004. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature*, 428(6980), pp.281–286.

Wittmann, A. & Suess, B., 2012. Engineered riboswitches: Expanding researchersâ€™ toolbox with synthetic RNA regulators. *FEBS Letters*, 586(15), pp.2076–2083.

Wittmann, A. & Suess, B., 2011. Selection of tetracycline inducible self-cleaving ribozymes as synthetic devices for gene regulation in yeast. *Molecular BioSystems*, 7(8), p.2419.

Xiao, H., Edwards, T.E. & Ferré-D'Amaré, A.R., 2008. Structural Basis for Specific, High-Affinity Tetracycline Binding by an In Vitro Evolved Aptamer and Artificial Riboswitch. *Chemistry & Biology*, 15(10), pp.1125–1137.

Xue, Y. et al., 2013. Direct conversion of fibroblasts to neurons by reprogramming PTB-regulated microRNA circuits. *Cell*, 152(1-2), pp.82–96.

Yip, K.Y., Cheng, C. & Gerstein, M., 2013. Machine learning and genome annotation: a match meant to be? *Genome biology*, 14(5), p.205.

Zhang, S. et al., 2016. A deep learning framework for modeling structural features of RNA-binding protein targets. *Nucleic Acids Research*, 44(4), pp.e32–e32.

Zuber, J. et al., 2018. Analysis of RNA nearest neighbor parameters reveals interdependencies and quantifies the uncertainty in RNA secondary structure prediction. *RNA*, 24(11), pp.1568–1582.

Zuker, M. & Stiegler, P., 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, 9(1), pp.133–148.

8 Anhang

8.1 Abkürzungen

% (v/v)	% (Volumen/Volumen)
% (w/v)	% (Gewicht/Volumen)
3'SS	3'Spleißstelle
5'SS	5'Spleißstelle
Ac	Acetat
ADH1	Alkoholdehydrogenase 1
Amp	Ampicillin
BE	Basalexpression
DMSO	Dimethylsulfoxid
dNTP	Desoxyribonukleosidtriphosphate
DNA	Desoxyribonukleinsäure
ds	doppelsträngig
E. coli	*Escherichia coli*
et al.	et alii (lat.: “und andere”)
EtOH	Ethanol
fwd	vorwärts (*forwar*d)
GFP	grün fluoreszierendes Protein
KD	Dissoziationskonstante
LB	Nährmedium (lysogeny broth)
MQ	Milli-Q®-Wasser
mRNA	messenger RNA
nt	Nukleotide
ORF	offener Leserahmen (open reading frame)
p.a.	pro analysis
PBS	Phosphatgepufferte Salzlösung
PCR	Polymerase-Kettenreaktion
rev	rückwärts (reverse)
RBS	Ribosomenbindestelle
RNA	Ribonukleinsäure
RNAi	RNA-Interferenz
rRNA	Ribosomale RNA
RT	Raumtemperatur
S. cerevisiae	*Saccharomyces cerevisiae*
SCD	*Synthetic complete dextrose*
SD	Shine-Dalgarno
SDS	Natriumlaurylsulfat
SELEX	Systematische Evolution von Liganden durch exponentielle Anreicherung
SF	Schaltfaktor
snRNP	*small nuclear ribonucleoprotein particle*
ss	einzelsträngig
TC	Tetrazyklin

Ura	Uracil
UTR	untranslatierter Bereich
ÜN	über Nacht
ÜNK	Übernachtkultur
YEPD	Yeast-Extrakt-Pepton-Dextrose Medium

8.2 Nukleobasen

A	Adenin
G	Guanin
T	Thymin
C	Cytosin
U	Uracil

8.3 Dimensionen

k	kilo (10^{3})
m	milli (10^{-3})
µ	mikro (10^{-6})
n	nano (10^{-9})
p	pico (10^{-12})

8.4 Einheiten

°C	Grad Celsius
bp	Basenpaare
Da	Dalton
g	Gramm
h	Stunden
l	Liter
M	Molar
min	Minuten
mol	Einheit der Stoffmenge
rpm	Umdrehungen pro Minute (engl. *revolutions per minute*)
sec	Sekunde
U	Einheit der Enzymaktivität; 1U = µmol Substratumsetzung pro Minute

8.5 Zusätzliche Tabellen und Abbildungen

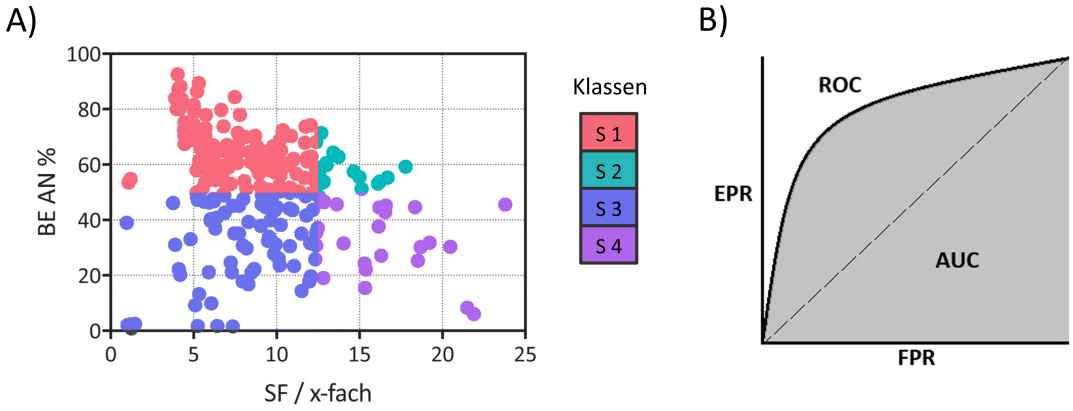


Abbildung 8.5: A) Die in 4 Klassen eingeteilten SF in Bezug auf die Basalexpression; **B)** Beispiel einer ROC-Kurve mit AUC (***A**rea **u**nder the ROC-**c**urve*) EPR= Echte positive Rate; FPR = falsche positive Rate.

Tabelle 8.5: Zusammenstellung der Sequenz der 5´Stämme der in dieser Arbeit klonierten Dimere mit den entsprechenden biophysikalischen Parametern.

Name	Stamm [5´-3]	BE [%]	Aus [%]	SF [X-fach]	Tm [°C]	dG [kcal/mol]	GC [%]	Entropie [bit]	freie Basen	gepaarte Basen	Länge Stamm
Derivat 1	TGGTGACC	2,21	1,30	1,47	44,06	-24,79	63	1,3209	31	42	8
Derivat 2	CGGTGACC	31,46	1,81	18,62	46,84	-25,19	75	1,2555	33	40	8
Derivat 3	GGGTGACC	16,51	1,58	11,15	50,14	-25,83	75	1,2130	31	42	8
Derivat 4	AAGTGACC	38,19	4,99	8,14	33,79	-22,98	50	1,3209	31	42	8
Derivat 5	ATGTGACC	3,02	2,80	1,23	34,68	-23,07	50	1,3863	31	42	8
Derivat 6	ACGTGACC	43,50	4,45	9,87	41,02	-24,22	63	1,3209	31	42	8
Derivat 7	AGATGACC	2,15	2,37	0,91	35,02	-23,50	50	1,3209	31	42	8
Derivat 8	AGTTGACC	44,88	5,78	8,07	33,79	-22,98	50	1,3863	31	42	8
Derivat 9	AGCTGACC	33,82	3,04	12,45	43,47	-25,39	63	1,3209	31	42	8
Derivat 10	AGGAGACC	36,13	4,33	8,64	44,13	-25,43	63	1,0822	31	42	8
Derivat 11	AGGCGACC	32,19	2,03	16,41	49,75	-26,36	75	1,0822	31	42	8
Derivat 12	AGGGGACC	8,63	2,07	4,16	52,88	-26,75	75	1,0397	31	42	8
Derivat 13	AGGTAACC	56,68	5,97	9,59	34,80	-22,71	50	1,3209	31	42	8
Derivat 14	AGGTTACC	49,53	4,17	12,55	34,80	-22,71	50	1,3863	31	42	8
Derivat 15	AGGTCACC	37,15	3,30	11,36	43,92	-25,00	63	1,3209	31	42	8
Derivat 16	AGGTGTCC	30,53	2,89	10,77	43,92	-25,00	63	1,3209	31	42	8
Derivat 17	AGGTGCCC	16,12	2,06	7,87	52,26	-26,71	75	1,2555	31	42	8
Derivat 18	AGGTGGCC	62,87	11,31	5,72	52,26	-26,71	75	1,2130	31	42	8
Derivat 19	AGGTGAAC	51,02	4,61	11,29	33,79	-22,98	50	1,2555	31	42	8
Derivat 20	AGGTGATC	51,75	3,68	12,53	35,02	-23,50	50	1,3209	31	42	8
Derivat 21	AGGTGAGC	41,23	3,47	11,92	43,47	-25,39	63	1,2130	31	42	8
Derivat 22	AGGTGACA	43,59	4,80	9,40	36,95	-23,15	50	1,2555	31	42	8
Derivat 23	AGGTGACT	40,62	3,86	10,82	36,79	-22,27	50	1,3209	31	42	8
Derivat 24	AGGTGACG	53,64	10,58	5,29	40,18	-22,69	63	1,2130	31	42	8
Derivat 25	CCGTGAAG	53,00	6,15	9,01	36,38	-21,60	63	1,3209	31	42	8
Derivat 26	CCAGTGGA	45,87	5,11	9,09	43,28	-23,76	63	1,3209	31	42	8
Derivat 27	AGGCAACC	34,26	2,82	12,14	43,01	-24,69	63	1,0822	31	42	8

Derivat 28	AGGCTGAC	33,02	2,68	12,35	43,47	-25,39	63	1,3209	31	42	8
Derivat 29	AGGTTGCC	41,42	3,06	13,97	43,01	-24,69	63	1,3209	31	42	8
Derivat 30	AGGTGAGG	32,23	3,13	10,99	43,13	-23,47	63	0,9003	31	42	8
Derivat 31	ACGTCACC	49,72	4,32	12,28	41,02	-24,22	63	1,2130	31	42	8
Derivat 32	AGGAGAGC	29,70	3,02	9,99	43,68	-25,82	63	0,9743	31	42	8
Derivat 33	AGGAGTCC	32,45	2,93	11,53	44,13	-25,43	63	1,3209	31	42	8
Derivat 34	TCGTGACC	43,88	4,12	11,64	41,41	-24,44	63	1,3209	31	42	8
Derivat 35	AGGCTACC	27,62	2,16	13,23	44,64	-25,12	63	1,3209	31	42	8
Derivat 36	AGTGACCC	29,05	3,02	9,74	43,92	-25,00	63	1,3209	31	42	8
Derivat 37	AGTGGACC	36,28	2,96	12,35	43,92	-25,00	63	1,3209	31	42	8
Derivat 38	AGGTGGAC	36,28	3,25	11,14	43,92	-25,00	63	1,2130	31	42	8
Derivat 39	AGGTGGGC	20,53	2,27	9,01	52,26	-26,71	75	1,0735	31	42	8
Derivat 40	AGGGGTCC	10,35	2,63	3,97	52,88	-26,75	75	1,2130	31	42	8
Derivat 41	AGGACACC	29,99	3,02	9,92	43,92	-25,00	63	1,0822	31	42	8
Derivat 42	ACCTGACC	34,42	3,17	10,93	43,92	-25,00	63	1,2130	31	42	8
ML-RD_AA1	AGGCTACG	36,27	2,95	12,35	40,87	-22,81	63	1,3209	31	42	8
ML-RD_AA3	AGGCGATC	53,75	4,38	12,33	41,24	-24,86	63	1,3209	31	42	8
ML-RD_AA5	AGGCCATC	10,66	0,71	15,24	43,85	-25,21	63	1,3209	31	42	8
ML-RD_AB1	ACCCTGTAC	33,37	3,45	9,74	47,23	-25,69	56	1,2730	31	44	9
ML-RD_AB2	ACCCTGTTC	38,30	3,47	11,03	45,95	-25,69	56	1,2149	31	44	9
ML-RD_AB3	GCTGAGCC	38,84	2,45	15,87	49,20	-26,61	75	1,2555	31	42	8
ML-RD_AC1	TCGCGTAC	57,32	4,74	12,17	39,11	-23,78	63	1,3209	31	42	8
ML-RD_AC2	ACCGGCTAAC	14,07	0,96	14,67	53,73	-27,74	60	1,2799	31	46	10
ML-RD_AC5	AGCGCATC	15,64	1,50	10,34	40,58	-24,82	63	1,3209	31	42	8
ML-RD_AD2	TCGGCAAC	50,04	3,45	14,54	40,44	-24,13	63	1,3209	31	42	8
ML-RD_SA1	CGCGTGAAT	64,40	7,11	9,08	39,83	-22,58	56	1,3689	33	42	9
ML-RD_SA2	AGAGGTA	83,45	23,02	3,67	22,06	-20,28	43	1,0042	31	40	7
ML-RD_SA3	GAACGCA	61,12	7,49	8,18	25,56	-20,61	57	1,0790	31	40	7
ML-RD_SA4	CCTAAAG	75,44	15,40	4,90	11,31	-17,38	43	1,2770	31	40	7
ML-RD_SA5	GGTCCAA	77,90	8,28	9,42	29,72	-20,68	57	1,3518	31	40	7
ML-RD_SB1	ACCGGAT	59,39	12,67	4,70	29,56	-20,35	57	1,3518	31	40	7
ML-RD_SB2	GTACTGT	72,22	10,92	6,62	18,72	-18,37	43	1,2770	31	40	7
ML-RD_SB3	TCGCATA	45,04	8,72	5,16	18,37	-19,50	43	1,3518	31	40	7
ML-RD_SB4	CTTACCA	71,39	10,56	6,79	16,48	-19,08	43	1,0790	31	40	7
ML-RD_SB5	CGACGTCT	46,47	3,70	12,54	37,42	-22,11	63	1,3209	33	40	8
R1_A1	TCTCCC	58,53	6,34	9,53	25,92	-21,55	67	0,6365	31	38	6
R1_A10	AGGCTTTGA	41,66	5,40	7,97	39,09	-23,87	44	1,3108	31	44	9
R1_A11	ACGGACAGAG	23,34	2,71	8,60	52,55	-26,79	60	1,0549	31	46	10
R1_A12	ATCAGAATAT	30,50	3,08	9,88	27,47	-20,47	20	1,1683	31	46	10
R1_A2	CCGAGA	47,57	7,92	6,42	20,03	-20,51	67	1,0986	31	38	6
R1_A3	CCGAGCT	43,64	4,94	10,02	36,52	-21,85	71	1,2770	31	40	7
R1_A4	ATCCGAA	49,91	4,88	10,23	16,57	-19,32	43	1,2770	31	40	7
R1_A5	GCGATCA	54,80	8,18	6,72	27,39	-21,13	57	1,3518	31	40	7
R1_A6	CGATCCCG	35,11	2,81	12,48	43,91	-23,13	75	1,2130	33	40	8
R1_A7	TTTTGCCC	46,78	4,97	9,38	32,21	-22,19	50	0,9743	31	42	8
R1_A8	CGAAACACG	40,28	4,50	11,34	36,31	-21,55	56	1,0609	33	42	9
R1_A9	TGTCCTAGT	33,08	5,94	3,76	41,38	-23,18	44	1,2730	31	44	9

R1_B1	AGCACA	64,29	11,28	5,72	12,11	-19,44	50	1,0114	31	38	6
R1_B10	GGACTCACT	24,61	2,09	11,79	46,24	-24,49	56	1,3689	31	44	9
R1_B11	GATCACAGTT	38,98	4,28	9,06	40,55	-23,25	40	1,3662	31	46	10
R1_B12	ACATATCTAG	53,50	10,31	5,19	34,28	-21,85	30	1,2799	31	46	10
R1_B2	GTCCCA	46,29	6,99	8,00	25,26	-20,31	67	1,2425	31	38	6
R1_B3	TAAACCC	49,00	4,00	12,19	16,83	-19,52	43	1,0042	31	40	7
R1_B4	TCGTTAC	61,46	11,87	5,18	14,24	-19,44	43	1,2770	31	40	7
R1_B5	TGAGCAC	1,70	1,42	1,20	30,04	-22,47	57	1,3518	31	40	7
R1_B6	AAATAGCC	48,54	6,54	7,45	23,34	-20,90	38	1,2130	31	42	8
R1_B9	TGTTAGTAC	19,01	3,87	4,82	30,05	-21,87	33	1,2730	31	44	9
R1_C1	CCAGCC	56,84	4,47	13,42	33,12	-22,22	83	0,8676	31	38	6
R1_C10	CGGGGGAAT	21,10	1,69	12,56	51,41	-24,72	67	1,1491	33	42	9
R1_C11	CCGGACCTTC	9,22	1,25	7,36	57,14	-28,98	70	1,2206	31	46	10
R1_C12	TATCAAGTAA	22,26	3,54	6,30	27,20	-20,64	20	1,1683	31	46	10
R1_C2	AGTCTG	58,95	10,02	5,94	8,90	-18,05	50	1,3297	31	38	6
R1_C3	AGTCCAC	33,95	3,73	9,11	30,19	-22,29	57	1,2770	31	40	7
R1_C4	ACAGTTG	57,12	10,71	5,34	14,83	-18,31	43	1,3518	31	40	7
R1_C6	ATTGACCA	30,22	4,75	6,43	26,08	-20,95	38	1,3209	31	42	8
R1_C8	GCTGTTCCA	39,26	4,19	9,32	45,63	-25,06	56	1,3108	31	44	9
R1_C9	CACACCCCG	23,59	1,84	12,82	55,70	-25,90	78	0,8487	31	44	9
R1_D1	ATCGTC	59,59	10,10	5,91	7,29	-19,05	50	1,3297	31	38	6
R1_D10	AATTTGCCA	53,11	7,63	7,09	28,87	-21,33	33	1,3108	31	44	9
R1_D11	CTTTCTCCTC	46,58	4,72	9,84	45,06	-26,42	50	0,6931	31	46	10
R1_D12	CTTCTTCTGG	53,00	7,30	7,23	44,30	-24,46	50	1,0297	31	46	10
R1_D2	TAGAGC	55,21	8,51	6,96	12,25	-19,92	50	1,3297	31	38	6
R1_D3	AGGGAGC	37,86	3,43	11,06	40,84	-24,43	71	0,9557	31	40	7
R1_D4	GTGACCC	39,73	3,68	10,81	38,54	-23,12	71	1,2770	31	40	7
R1_D6	GGTGTGCC	20,24	1,08	18,68	49,54	-25,79	75	1,0397	31	42	8
R1_D7	AATAGTCT	59,19	9,51	6,23	15,24	-18,48	25	1,2555	31	42	8
R1_D8	TCGACTATT	36,90	4,72	8,30	29,13	-20,63	33	1,2730	31	44	9
R1_D9	ACGCTGTAC	54,34	7,00	7,96	44,21	-25,30	56	1,3689	31	44	9
R1_E1	AACACC	64,85	10,06	6,60	8,09	-18,88	50	0,6931	31	38	6
R1_E10	CACATAACAG	39,90	6,59	6,05	37,66	-22,10	40	1,1683	31	46	10
R1_E11	ACATAGTTCT	20,96	3,71	5,65	35,38	-22,15	30	1,2799	31	46	10
R1_E12	TAATGTCTCA	1,72	1,61	1,08	35,92	-23,25	30	1,2799	31	46	10
R1_E2	GCCGTC	47,66	3,92	12,14	30,89	-21,77	83	1,0114	31	38	6
R1_E3	AGAATGG	2,04	2,13	0,96	16,25	-18,56	43	0,9003	31	40	7
R1_E3	ATAAACAC	57,50	11,00	5,24	11,33	-18,76	25	1,0042	31	42	8
R1_E4	GTACAGT	53,74	8,94	6,02	18,72	-18,37	43	1,3518	31	40	7
R1_E7	GTAGTGCT	51,57	6,90	8,49	34,88	-21,47	50	1,2555	31	42	8
R1_E8	AGCCAAAGA	32,49	4,15	8,67	39,09	-24,41	44	0,9950	31	44	9
R1_E9	CACAACACG	55,47	8,80	6,30	39,18	-22,13	56	0,9650	31	44	9
R1_F10	CAAACCTACA	49,17	5,17	9,64	40,19	-23,44	40	0,9433	31	46	10
R1_F11	TTAAGAACGT	46,46	7,69	6,04	31,47	-20,80	30	1,2799	31	46	10
R1_F2	TCTGAG	56,18	9,18	6,19	10,17	-18,27	50	1,3297	31	38	6
R1_F3	TGTCTGT	1,56	1,54	1,01	20,60	-19,35	43	0,9557	31	40	7
R1_F4	AACTCGA	66,43	11,69	5,70	15,92	-20,83	43	1,2770	31	40	7

R1_F5	ACCACCCT	31,04	2,18	14,65	46,48	-23,59	63	0,9003	31	42	8
R1_F6	CCACCAAA	19,66	4,11	4,19	32,08	-21,04	50	0,6931	31	42	8
R1_F7	AACTCTAT	52,09	9,45	5,55	15,24	-18,48	25	1,0822	31	42	8
R1_F8	AGCAACTGT	46,09	5,23	9,49	39,00	-22,92	44	1,3689	31	44	9
R1_F9	ATAGACAAG	54,33	8,60	6,32	28,52	-20,64	33	1,1491	31	44	9
R1_G1	AGAGGA	58,72	9,03	6,85	14,15	-19,59	50	0,6931	31	38	6
R1_G10	GGCTCAAGTC	21,94	2,30	9,54	52,36	-28,30	60	1,3662	31	46	10
R1_G11	AAACCGCACC	17,83	1,76	10,20	52,51	-27,31	60	0,9433	31	46	10
R1_G2	GACGGA	40,76	6,12	6,77	21,50	-19,64	67	1,0114	31	38	6
R1_G3	GCTTCAT	50,75	10,02	5,09	17,47	-18,58	43	1,2770	31	40	7
R1_G5	GAAAAAGG	37,09	9,16	3,89	17,14	-17,97	38	0,6616	31	42	8
R1_G6	TACCCTGC	29,76	2,44	12,07	44,78	-24,91	63	1,2130	31	42	8
R1_G9	GCCCGATTC	24,28	1,96	12,26	49,12	-26,38	67	1,2730	31	44	9
R1_H1	ACACAC	60,12	10,65	5,73	9,62	-19,15	50	0,6931	31	38	6
R1_H2	CTGATC	44,50	6,97	6,38	8,55	-19,01	50	1,3297	31	38	6
R1_H3	GTCCCTT	50,38	4,86	10,34	29,53	-20,12	57	1,0042	31	40	7
R1_H4	TCCGATA	57,03	7,61	7,55	19,29	-19,54	43	1,3518	31	40	7
R1_H6	ACGATCCA	44,62	5,48	8,19	34,51	-22,62	50	1,2555	31	42	8
R1_H7	CCGATTCT	50,32	5,64	9,02	30,36	-20,65	50	1,2555	31	42	8
R1_H9	CGTCGCATA	30,82	2,21	15,31	41,51	-23,57	56	1,3689	33	42	9
R2_28	CACCCTCAA	45,71	5,01	9,09	45,39	-24,45	56	0,9369	31	44	9
R2_A1	ACAGTT	72,97	16,23	4,50	-7,41	-16,15	33	1,3297	31	38	6
R2_A10	TGCGAAAGTT	2,46	1,67	1,46	38,98	-22,78	40	1,3138	31	46	10
R2_A11	TCGCGCCGTA	22,14	1,44	15,38	58,91	-28,39	70	1,2799	31	46	10
R2_A12	AACTAC	79,17	15,72	5,04	-7,69	-17,29	33	1,0114	31	38	6
R2_A2	GCACGT	63,99	10,26	6,24	19,51	-18,61	67	1,3297	31	38	6
R2_A3	CAACAGG	64,50	7,70	8,37	24,06	-19,24	57	1,0790	31	40	7
R2_A4	CACTAAC	67,84	12,35	5,49	14,00	-19,18	43	1,0042	31	40	7
R2_A5	CTATTAA	84,70	20,91	4,05	-14,23	-14,97	14	1,0042	31	40	7
R2_A6	ATTCGGAG	64,49	8,36	7,70	30,36	-20,92	50	1,3209	31	42	8
R2_A7	AACTGGGT	62,05	5,62	11,03	36,03	-21,57	50	1,3209	31	42	8
R2_A8	AGCCAGTAT	49,78	4,61	10,79	40,97	-23,17	44	1,3689	31	44	9
R2_A9	CTTTAGTTT	86,40	16,63	5,20	14,41	-17,56	22	1,0027	31	44	9
R2_B1	CTGTTC	68,13	12,28	5,54	5,29	-18,49	50	1,0114	31	38	6
R2_B10	CTCGCCTCCG	15,46	1,01	15,34	60,80	-29,08	80	0,9503	31	46	10
R2_B3	AAAACGG	65,87	12,16	5,41	7,46	-16,99	43	0,9557	31	40	7
R2_B4	AAAGTGC	70,59	12,37	5,70	15,01	-19,96	43	1,2770	31	40	7
R2_B5	TATCGCT	63,20	8,75	7,21	18,14	-18,94	43	1,2770	31	40	7
R2_B6	CTCTTTAC	89,38	16,78	5,32	20,19	-20,30	38	0,9743	31	42	8
R2_B7	ACACAGAA	68,88	12,20	5,65	25,56	-20,81	38	0,9003	31	42	8
R2_B8	GCCTAGAGG	35,07	3,04	11,50	51,73	-25,81	67	1,2730	31	44	9
R2_B9	AGACCTCCG	26,16	2,58	10,13	51,23	-25,83	67	1,2730	31	44	9
R2_C1	ATATAC	83,66	19,65	4,27	-20,94	-15,61	17	1,0114	31	38	6
R2_C10	TTAAGGCTTA	45,09	6,09	7,36	35,17	-22,26	30	1,2799	31	46	10
R2_C11	CAAAAT	92,51	22,96	4,04	-36,83	-12,86	17	0,8676	31	38	6
R2_C12	GTACCT	79,69	12,06	6,66	11,09	-17,41	50	1,3297	31	38	6
R2_C2	TCAGAC	72,16	13,37	5,43	11,83	-19,80	50	1,3297	31	38	6

R2_C3	ATTTGCT	69,55	13,84	5,02	3,32	-17,05	29	1,1537	31	40	7
R2_C5	AAACCAAC	66,79	9,60	6,96	20,88	-20,26	38	0,6616	31	42	8
R2_C6	CCGACGGA	46,56	3,61	12,88	47,09	-24,38	75	1,0822	31	42	8
R2_C7	CGACGAGT	46,36	3,61	12,85	37,42	-22,11	63	1,3209	33	40	8
R2_C8	TCAGTCCTG	48,32	3,83	12,57	45,70	-24,65	56	1,3108	31	44	9
R2_C9	TGGTACCAA	51,72	4,78	10,82	41,04	-23,04	44	1,3689	31	44	9
R2_D1	TCTGAC	65,93	11,30	5,82	11,83	-19,80	50	1,3297	31	38	6
R2_D10	AAACCAGTTC	56,67	6,40	8,94	39,53	-24,36	40	1,2799	31	46	10
R2_D11	ACTTTT	81,35	19,69	4,13	-34,71	-13,86	17	0,8676	31	38	6
R2_D12	ATAAGA	80,01	20,08	3,98	-22,87	-17,53	17	0,8676	37	32	6
R2_D3	CGGGCGA	56,35	5,53	10,21	44,14	-24,75	86	0,9557	33	38	7
R2_D4	ATATAGT	80,06	19,37	4,16	-5,84	-15,59	14	1,0042	31	40	7
R2_D5	AAACCGCT	60,45	4,83	12,98	31,69	-20,91	50	1,2555	31	42	8
R2_D6	ACGCCTCG	56,96	6,47	8,95	46,30	-24,05	75	1,2130	31	42	8
R2_D7	CGCTTTAT	67,46	15,12	4,45	18,26	-18,36	38	1,2130	33	40	8
R2_D8	TTGCAAGAT	54,06	6,74	8,05	29,82	-20,94	33	1,3108	31	44	9
R2_D9	AATCGACAA	59,54	9,41	6,32	27,30	-20,97	33	1,1491	31	44	9
R2_E11	GGCTAG	46,96	7,86	6,18	21,61	-19,00	67	1,2425	31	38	6
R2_E12	CCCGAG	64,61	6,92	9,34	28,73	-19,95	83	1,0114	31	38	6
R2_E3	GATAAGA	70,21	15,80	4,45	6,48	-18,36	29	0,9557	31	40	7
R2_E5	GTAAGAGT	67,08	9,57	6,99	24,36	-19,49	38	1,0822	31	42	8
R2_E7	GGATAACCC	31,75	1,65	19,22	45,06	-24,75	56	1,3108	31	44	9
R2_E8	CTCGTACAC	48,54	4,22	11,75	41,66	-24,52	56	1,2730	31	44	9
R2_E9	CTGCAAAATA	62,43	10,41	6,04	31,05	-21,31	30	1,2206	31	46	10
R2_F1	CCGAGC	72,41	6,99	10,37	29,67	-21,87	83	1,0114	31	38	6
R2_F10	CTTAATAAAA	68,85	14,29	4,87	10,51	-17,04	10	0,8979	31	46	10
R2_F11	AATCGA	88,18	21,19	4,20	-9,43	-17,94	33	1,2425	31	38	6
R2_F12	CGTCAA	79,02	15,66	5,07	2,11	-17,33	50	1,3297	33	36	6
R2_F2	GATAGT	81,42	16,14	5,01	-3,26	-15,91	33	1,0986	31	38	6
R2_F3	ATTGTGT	75,32	16,85	4,46	4,87	-16,93	29	0,9557	31	40	7
R2_F6	GCTTTCGC	67,33	8,81	7,65	37,00	-23,54	63	1,0822	31	42	8
R2_F7	CAATCGGACG	49,61	4,10	12,10	47,69	-24,74	60	1,3138	31	46	10
R2_F8	TCCGGACAC	27,12	1,67	16,31	51,95	-27,15	67	1,2730	31	44	9
R2_F9	GCCTCAACTA	44,09	4,09	10,78	48,71	-25,86	50	1,2799	31	46	10
R2_G1	ACGAGC	60,42	8,05	7,62	20,29	-20,94	67	1,0986	31	38	6
R2_G10	TGGCTGGATC	0,83	0,68	1,24	55,03	-29,10	60	1,2799	31	46	10
R2_G11	TGCTCG	47,20	8,98	5,24	19,04	-19,20	67	1,0986	31	38	6
R2_G12	AAATTC	87,35	21,45	4,08	-31,45	-15,09	17	1,0114	31	38	6
R2_G2	CTGCGC	66,62	6,77	9,80	28,79	-21,83	83	1,0114	31	38	6
R2_G3	TCTACAG	66,93	10,87	6,16	18,29	-18,96	43	1,3518	31	40	7
R2_G4	CCTGTAA	64,48	11,10	5,80	16,48	-18,76	43	1,3518	31	40	7
R2_G5	CGAGCCGC	30,80	3,19	12,39	51,89	-26,94	88	0,9743	33	40	8
R2_G6	CGTATAGA	73,11	12,79	5,72	22,32	-20,43	38	1,3209	33	40	8
R2_G7	GTTCGACCT	31,20	3,20	9,94	43,53	-23,44	56	1,3108	31	44	9
R2_G8	ACTAATAGA	74,33	14,78	5,03	22,81	-20,93	22	1,1491	31	44	9
R2_G9	TCCCCCTTAC	13,21	2,48	5,33	56,49	-28,35	60	0,8979	31	46	10
R2_H10	CGGTCCTTAA	47,92	4,93	9,72	46,09	-24,56	50	1,3662	33	44	10

R2_H11	CTCACT	77,90	13,83	5,73	8,90	-17,78	50	1,0114	31	38	6
R2_H12	GAATAT	83,95	21,92	3,88	-23,03	-13,71	17	1,0114	31	38	6
R2_H3	CACTTAT	72,56	16,46	4,41	2,53	-16,27	29	1,0790	31	40	7
R2_H5	TAACCACG	50,95	5,37	9,48	30,57	-20,19	50	1,2555	31	42	8
R2_H6	CTCGCAGC	44,54	2,74	16,20	45,93	-25,93	75	1,2130	31	42	8
R2_H7	GCACACGAT	48,91	12,78	6,85	43,47	-23,49	56	1,3108	31	44	9
R2_H8	TCCCCCAAT	21,09	2,89	7,30	49,01	-24,32	56	0,9950	31	44	9
R2_H9	TGAGCCAAGA	2,29	1,96	1,17	50,11	-27,61	50	1,2799	31	46	10
R3_A1	ATAGTTCCG	65,29	7,27	8,98	35,15	-21,61	44	1,3689	31	44	9
R3_A10	CCCGTG	53,79	4,19	12,85	28,13	-19,52	83	1,0114	31	38	6
R3_A11	ACCTTCTG	62,89	5,42	12,11	32,83	-21,45	50	1,2555	31	42	8
R3_A12	CCAGAC	70,29	6,47	10,88	21,07	-20,51	67	1,0114	31	38	6
R3_A2	CCCACTA	53,30	4,20	12,70	30,22	-20,78	57	0,9557	31	40	7
R3_A3	CGGACCG	35,39	2,88	12,39	40,66	-21,92	86	1,0042	33	38	7
R3_A4	CTCTACT	62,77	8,64	7,34	18,06	-18,90	43	1,0042	31	40	7
R3_A5	GGAGAGGAC	9,91	1,70	6,06	51,70	-27,65	67	0,9369	31	44	9
R3_A6	CTTACGCTT	64,00	7,14	8,96	33,65	-21,21	44	1,2149	31	44	9
R3_A7	GGTGCT	55,20	4,73	11,68	24,08	-19,39	67	1,0114	31	38	6
R3_A8	CACACCG	58,00	6,10	9,51	32,77	-20,48	71	0,9557	31	40	7
R3_B1	GGTTGTCG	53,91	3,39	16,22	37,27	-21,50	63	0,9743	31	42	8
R3_B10	TCGTAAACCT	67,88	5,83	11,71	40,88	-22,82	40	1,3138	31	46	10
R3_B11	TCACAGAGT	58,92	4,78	12,36	40,31	-23,45	44	1,3689	31	44	9
R3_B12	TCAGAGTCA	61,13	5,23	11,84	40,67	-24,76	44	1,3689	31	44	9
R3_B2	TTAGGGAGG	45,61	3,79	13,62	46,69	-24,11	56	0,9950	31	44	9
R3_B4	CTGCGG	60,54	5,89	10,38	27,82	-19,91	83	1,0114	31	38	6
R3_B6	TCTCGTGC	53,13	3,35	16,12	40,99	-24,83	63	1,0822	31	42	8
R3_B7	ATCGGTGAC	45,53	1,93	23,79	44,06	-25,43	56	1,3689	31	44	9
R3_B8	CGTCATC	14,36	1,99	11,51	23,93	-20,98	57	1,2770	33	38	7
R3_B9	TACGACAAG	52,33	5,29	10,15	34,90	-21,58	44	1,2730	31	44	9
R3_C1	AGGTTATC	71,41	5,70	12,69	24,87	-21,21	38	1,3209	31	42	8
R3_C10	GAGGAGA	22,19	5,41	4,10	31,30	-22,58	57	0,6829	31	40	7
R3_C11	TCTTACAC	77,97	10,09	7,79	24,54	-21,18	38	1,0822	31	42	8
R3_C12	GTAGAGT	71,47	9,32	7,67	19,42	-18,80	43	1,0790	31	40	7
R3_C2	TCCTGCATTA	17,82	1,56	11,97	43,71	-24,64	40	1,2799	31	46	10
R3_C4	CCTCTCTGA	60,89	6,31	9,69	45,85	-25,81	56	1,2149	31	44	9
R3_C5	GTCTAGT	69,99	8,63	8,12	19,42	-18,80	43	1,2770	31	40	7
R3_C6	CGTCAGTATT	42,65	6,21	6,87	38,28	-22,34	40	1,3322	33	44	10
R3_C7	TCTCGCT	57,58	5,65	10,18	29,40	-21,14	57	1,0042	31	40	7
R3_C8	GAAGAAACAC	58,21	5,94	9,99	37,52	-23,87	40	0,9503	31	46	10
R3_C9	CACGGG	65,16	6,30	10,27	28,13	-19,52	83	1,0114	31	38	6
R3_D1	GCTGCTGG	57,58	5,65	10,18	48,39	-24,65	75	1,0397	31	42	8
R3_D10	AGGCATCC	30,27	1,50	20,47	43,85	-25,21	63	1,3209	31	42	8
R3_D11	TTCCAGAAG	61,77	7,11	8,69	36,48	-22,36	44	1,3689	31	44	9
R3_D12	CTGGGC	56,72	5,09	11,18	33,12	-22,22	83	1,0114	31	38	6
R3_D2	TCCCACATC	6,07	0,28	21,90	46,69	-26,00	56	0,9950	31	44	9
R3_D4	CTGACAGAAC	43,57	3,97	10,96	45,06	-26,26	50	1,2799	31	46	10
R3_D5	CCACAC	66,80	7,59	8,80	20,33	-20,08	67	0,6365	31	38	6

R3_D6	CCGCCCA	42,85	2,59	16,55	47,09	-24,05	86	0,7963	31	40	7
R3_D7	GTAGTCGAT	73,68	6,27	11,75	36,64	-21,94	44	1,3108	31	44	9
R3_D8	CACTGGAC	55,33	3,31	16,70	40,55	-24,18	63	1,3209	31	42	8
R3_D9	TAGCTTAG	57,92	11,19	5,17	23,94	-19,77	38	1,3209	31	42	8
R3_E10	GCGACG	60,31	8,24	7,32	25,55	-19,46	83	1,0114	31	38	6
R3_E11	GCATACCCA	62,52	5,33	11,85	47,23	-24,88	56	1,2149	31	44	9
R3_E12	CAACCCCC	27,83	2,84	9,85	49,23	-25,23	75	0,5623	31	42	8
R3_E2	CCCCGGAA	25,76	2,09	12,33	49,58	-24,46	75	1,0397	31	42	8
R3_E3	TACGACGAA	35,15	4,56	7,74	35,65	-22,34	44	1,2730	31	44	9
R3_E4	AGGGCATC	8,29	0,39	21,49	43,85	-25,21	63	1,3209	31	42	8
R3_E5	GGATACT	60,84	7,23	8,45	20,01	-18,62	43	1,3518	31	40	7
R3_E6	CTCGTCATTT	16,76	2,03	8,31	37,05	-22,30	40	1,1683	31	46	10
R3_E7	CGCAGC	71,56	6,95	10,35	28,79	-21,87	83	1,0114	33	36	6
R3_E8	ATTGGGCC	44,59	2,48	18,35	43,40	-24,51	63	1,3209	31	42	8
R3_E9	TCGGATAG	62,08	6,49	9,67	32,40	-21,14	50	1,3209	31	42	8
R3_F1	CACTGGA	55,02	8,29	6,73	29,25	-21,05	57	1,3518	31	40	7
R3_F10	CGTTCCATA	31,65	2,64	12,03	35,30	-21,94	44	1,3108	33	42	9
R3_F11	CAGCTCAG	68,14	5,61	12,39	39,35	-23,04	63	1,3209	31	42	8
R3_F12	GCGGGAGT	55,47	3,70	15,00	49,75	-24,46	75	1,0735	31	42	8
R3_F2	CACTTTGTC	61,67	6,61	9,36	33,70	-22,85	44	1,2149	31	44	9
R3_F3	CCTGGGTG	59,19	3,35	17,78	48,65	-23,97	75	1,0397	31	42	8
R3_F4	GGCTCCT	37,63	2,33	16,16	40,84	-22,53	71	1,0790	31	40	7
R3_F5	CGGCGA	73,77	10,92	6,80	28,46	-22,04	83	1,0114	33	36	6
R3_F6	AGTGTCG	61,04	8,36	7,33	25,21	-19,98	57	1,2770	31	40	7
R3_F7	CCTTAGC	70,38	8,12	8,70	26,41	-21,32	57	1,2770	31	40	7
R3_F8	GCTTGGAC	62,75	4,59	13,74	40,65	-24,20	63	1,3209	31	42	8
R3_G1	AAGCTCTTAG	60,55	5,63	10,83	39,88	-23,65	40	1,3662	31	46	10
R3_G10	ACCACAGT	55,87	5,47	10,22	36,46	-21,84	50	1,2555	31	42	8
R3_G11	CACGCCTG	51,21	3,39	15,12	45,47	-23,58	75	1,2130	31	42	8
R3_G12	CATCCAGG	59,77	4,61	13,02	40,13	-22,47	63	1,3209	31	42	8
R3_G12	ACATCCAG	59,77	4,61	13,02	33,76	-21,54	50	1,2555	31	42	8
R3_G2	AGTGGTGTA	55,00	4,42	12,48	41,16	-23,52	44	1,0609	31	44	9
R3_G3	GGCCCC	31,55	2,28	14,04	45,90	-23,87	100	0,6365	31	38	6
R3_G4	TAGCTC	66,04	9,02	7,36	12,25	-19,92	50	1,3297	31	38	6
R3_G6	AGTAATCTGC	74,20	6,28	12,06	41,14	-25,27	40	1,3662	31	46	10
R3_G7	CACAGC	84,41	11,27	7,49	20,13	-20,47	67	1,0114	31	38	6
R3_H1	TCTCGG	65,86	9,60	6,90	20,03	-19,24	67	1,0986	31	38	6
R3_H10	GCAGACCA	43,56	4,43	9,90	43,61	-24,37	63	1,0822	31	42	8
R3_H11	GATACCTTGG	38,20	3,73	10,25	46,38	-24,18	50	1,3662	31	46	10
R3_H12	AGGCTCA	45,15	5,85	7,74	32,68	-22,58	57	1,3518	31	40	7
R3_H2	GGCGC	59,02	5,66	10,44	24,89	-20,77	100	0,6730	31	36	5
R3_H3	CTTACAGCT	68,94	6,74	10,22	37,02	-22,26	44	1,3108	31	44	9
R3_H7	GGATTGGA	51,26	6,52	8,18	34,75	-21,89	50	1,0397	31	42	8
R3_H8	TCCGTATTGA	59,27	6,87	8,76	41,78	-24,13	40	1,3322	31	46	10
R3_H9	CTGGCTAAC	45,21	2,72	16,58	43,44	-24,99	56	1,3689	31	44	9
R4_A1	TCAGGGTC	42,68	3,06	14,01	44,27	-25,22	63	1,3209	31	42	8
R4_A10	AGTAGGCC	48,74	3,51	14,05	44,64	-25,12	63	1,3209	31	42	8

R4_A11	TAGGGCTC	43,31	3,02	14,56	44,98	-25,34	63	1,3209	31	42	8
R4_A12	AGGTCAGC	42,78	3,19	13,25	43,47	-25,39	63	1,3209	31	42	8
R4_A2	TCAGGTCC	48,23	3,63	13,43	44,27	-25,22	63	1,3209	31	42	8
R4_A3	CGGCGCTA	35,03	2,16	16,26	47,19	-24,50	75	1,2555	33	40	8
R4_A4	TAGCACGC	38,65	3,38	11,54	41,43	-24,52	63	1,3209	31	42	8
R4_A5	CCTGCGGA	39,31	2,95	13,50	49,19	-25,12	75	1,2555	31	42	8
R4_A6	TTAGGGCC	22,76	2,28	10,04	44,57	-24,64	63	1,3209	31	42	8
R4_A7	GACCGTGC	35,02	2,83	11,91	47,09	-25,44	75	1,2555	31	42	8
R4_A8	CGGATCCG	38,06	2,46	15,88	43,91	-23,13	75	1,2555	33	40	8
R4_A9	GCACGGCT	25,85	2,38	11,09	49,12	-24,42	75	1,2555	31	42	8
R4_B1	AGGACGCT	42,05	4,33	9,73	43,25	-23,63	63	1,3209	31	42	8
R4_B10	GAGGCCTC	30,30	2,77	9,78	49,79	-26,65	75	1,2555	31	42	8
R4_B11	AGGCGTTC	52,63	3,10	17,34	40,29	-24,34	63	1,3209	31	42	8
R4_B12	CGCTGGCA	42,94	2,63	16,39	48,55	-25,44	75	1,2555	33	40	8
R4_B2	ATCTGGGC	59,68	5,18	11,89	43,85	-25,21	63	1,3209	31	42	8
R4_B3	CGAGCGCT	41,30	2,18	18,90	45,83	-24,21	75	1,2555	33	40	8
R4_B4	ACTCTGGC	53,21	4,24	12,64	43,47	-25,39	63	1,3209	31	42	8
R4_B5	AGCGTCCG	48,66	3,75	13,23	46,30	-24,05	75	1,2555	31	42	8
R4_B6	ACGGTCAC	49,40	6,21	9,00	41,02	-24,22	63	1,3209	31	42	8
R4_B7	TAGAGCGC	46,71	4,14	11,31	41,67	-24,95	63	1,3209	31	42	8
R4_B8	AGGCTCCT	31,71	2,07	15,58	46,17	-24,41	63	1,3209	31	42	8
R4_B9	CGGCTCAG	39,77	2,93	13,93	45,65	-24,05	75	1,2555	33	40	8
R4_C1	AGCGCCTG	42,89	2,73	15,74	48,41	-24,79	75	1,2555	31	42	8
R4_C10	CGGCCGAT	30,47	1,72	17,80	46,67	-23,64	75	1,2555	33	40	8
R4_C11	AGCGAGCT	49,76	3,10	16,19	42,80	-24,02	63	1,3209	31	42	8
R4_C12	CGCAGCTG	47,95	3,61	13,41	45,02	-24,01	75	1,2555	33	40	8
R4_C2	CGCCAGGT	44,62	3,98	11,25	48,91	-24,17	75	1,2555	33	40	8
R4_C3	TCCAGCTG	44,62	3,98	11,25	42,83	-23,65	63	1,3209	31	42	8
R4_C4	ATACGGCC	30,83	2,78	11,10	42,11	-24,16	63	1,3209	31	42	8
R4_C5	ACTGGGAC	39,69	2,37	16,73	43,92	-25,00	63	1,3209	31	42	8
R4_C6	CGGAGCCT	51,77	5,17	10,12	49,05	-24,60	75	1,2555	33	40	8
R4_C7	ATTCCGGC	52,43	6,48	8,94	40,69	-24,16	63	1,3209	31	42	8
R4_C8	ACTCCGGA	37,27	2,66	13,72	44,09	-24,23	63	1,3209	31	42	8
R4_C9	GCCGAGTC	24,21	2,24	8,80	47,25	-25,87	75	1,2555	31	42	8
R4_D1	AGGCTCGT	37,88	3,31	11,65	43,25	-23,63	63	1,3209	31	42	8
R4_D10	TCGCCTGA	48,63	4,43	11,15	43,77	-24,64	63	1,3209	31	42	8
R4_D11	GAGCTGCC	48,42	4,40	11,10	49,20	-26,61	75	1,2555	31	42	8
R4_D12	TCGGAACC	47,39	3,51	13,60	41,12	-24,17	63	1,3209	31	42	8
R4_D2	AGTGCCCG	30,37	2,80	10,94	48,91	-24,40	75	1,2555	31	42	8
R4_D3	AGCTTGCC	45,02	3,91	11,94	42,56	-25,08	63	1,3209	31	42	8
R4_D4	TAGCGCGA	44,33	5,41	8,31	41,50	-25,08	63	1,3209	31	42	8
R4_D5	TTGGCCAC	31,81	2,89	11,10	43,16	-24,48	63	1,3209	31	42	8
R4_D6	ACGGGATC	48,59	4,23	11,58	41,66	-24,47	63	1,3209	31	42	8
R4_D7	AGCCGTCG	39,58	2,47	15,89	46,30	-24,05	75	1,2555	31	42	8
R4_D8	AGACCTGC	38,77	3,18	12,46	43,47	-25,39	63	1,3209	31	42	8
R4_D9	CCCAGGTG	51,52	4,70	10,97	48,65	-23,97	75	1,2555	31	42	8
R4_E1	CCGCTGGA	54,85	4,20	13,46	49,19	-25,12	75	1,2555	31	42	8

R4_E10	TCAGAGGC	52,43	3,52	14,52	43,82	-25,61	63	1,3209	31	42	8
R4_E11	CGTGACCG	30,59	2,36	12,64	43,32	-22,88	75	1,2555	33	40	8
R4_E12	ACGGTGAC	54,09	3,86	14,02	41,02	-24,22	63	1,3209	31	42	8
R4_E2	AGCTCGGA	46,20	4,94	9,42	43,62	-24,62	63	1,3209	31	42	8
R4_E3	TCCAGGCT	35,48	2,01	17,78	46,31	-24,20	63	1,3209	31	42	8
R4_E4	AGTCGGCA	43,08	3,42	12,60	43,40	-24,51	63	1,3209	31	42	8
R4_E5	CGGTCGAC	35,00	2,07	16,93	44,30	-24,84	75	1,2555	33	40	8
R4_E6	CCAGGCTA	47,71	3,11	15,34	43,99	-23,88	63	1,3209	31	42	8
R4_E7	GCCTGAGC	47,01	2,49	18,59	49,20	-26,61	75	1,2555	31	42	8
R4_E8	TCAGCCAG	58,14	3,29	17,45	42,83	-23,65	63	1,3209	31	42	8
R4_E9	CCTAGGTC	53,93	3,65	14,54	41,88	-24,34	63	1,3209	31	42	8
R4_F1	TCGAGCGC	44,48	3,50	12,84	46,88	-26,19	75	1,2555	31	42	8
R4_F10	AGACCTGG	48,08	3,52	13,66	43,13	-23,47	63	1,3209	31	42	8
R4_F11	CTGGACGC	56,02	3,61	15,51	46,38	-25,54	75	1,2555	31	42	8
R4_F12	CTGGCGAC	49,44	2,81	17,79	46,38	-25,54	75	1,2555	31	42	8
R4_F2	AGGTTCCC	37,55	3,62	10,38	43,69	-24,73	63	1,3209	31	42	8
R4_F3	TCCAGGAG	40,09	3,48	11,53	43,49	-23,69	63	1,3209	31	42	8
R4_F4	ACTGCCGG	49,95	3,35	14,96	48,91	-24,40	75	1,2555	31	42	8
R4_F5	TCGCGAGC	48,38	2,48	19,68	46,88	-26,19	75	1,2555	31	42	8
R4_F6	TCGCAGGA	40,09	2,63	15,76	43,77	-24,41	63	1,3209	31	42	8
R4_F8	ACGCCTGA	54,86	3,92	13,94	43,40	-24,42	63	1,3209	31	42	8
R4_F9	TCTGGGCA	45,97	4,62	9,91	46,45	-25,08	63	1,3209	31	42	8
R4_G1	AAGCGTCC	50,74	4,92	11,53	40,29	-24,34	63	1,3209	31	42	8
R4_G10	ACGGAGCT	49,94	3,35	14,86	43,25	-23,63	63	1,3209	31	42	8
R4_G11	TCACGGCA	41,93	3,48	12,16	43,55	-24,30	63	1,3209	31	42	8
R4_G12	GACGTCGC	16,79	1,37	13,37	44,61	-25,09	75	1,2555	31	42	8
R4_G2	TCGTGGAC	47,18	3,00	16,12	41,41	-24,44	63	1,3209	31	42	8
R4_G3	CGCGGCAT	40,72	3,02	11,57	46,02	-23,60	75	1,2555	33	40	8
R4_G4	TCGGCTCA	43,60	4,04	11,10	43,77	-24,73	63	1,3209	31	42	8
R4_G5	CGTCGGCA	42,41	2,30	18,41	46,44	-24,70	75	1,2555	33	40	8
R4_G6	CGTGGACC	54,53	5,07	10,97	46,84	-25,19	75	1,2555	33	40	8
R4_G7	GCCTGCTA	50,25	11,65	4,30	44,32	-24,17	63	1,3209	31	42	8
R4_G8	AGGTGACC	25,21	0,66	39,50	43,92	-25,00	63	1,3209	31	42	8
R4_G9	TACGGACC	43,05	3,14	14,04	42,55	-24,17	63	1,3209	31	42	8
R4_H1	TACCGAGC	41,60	2,32	17,97	42,11	-24,56	63	1,3209	31	42	8
R4_H10	AGTCCCGA	46,07	4,72	9,87	44,09	-25,56	63	1,3209	31	42	8
R4_H11	TCGCCAGA	45,30	4,08	11,23	43,77	-25,18	63	1,3209	31	42	8
R4_H12	AACTCGGC	44,32	4,20	10,63	40,29	-24,34	63	1,3209	31	42	8
R4_H2	TTGCGCCA	42,86	4,13	10,47	42,60	-23,99	63	1,3209	31	42	8
R4_H3	ACGTGGCA	57,78	5,65	10,27	43,17	-24,08	63	1,3209	31	42	8
R4_H4	AGGCGCTA	49,62	4,90	10,97	44,14	-24,31	63	1,3209	31	42	8
R4_H5	ACCTCAGG	49,46	4,75	10,54	43,13	-23,47	63	1,3209	31	42	8
R4_H6	GCCCTGAG	44,36	3,91	11,49	49,00	-24,69	75	1,2555	31	42	8
R4_H7	TCCAGAGG	46,19	4,77	9,97	43,49	-23,69	63	1,3209	31	42	8
R4_H9	CGACTCGG	26,73	2,85	9,41	43,54	-23,31	75	1,3209	33	40	8
R4_H9	AGCCCAGT	37,07	2,88	13,05	45,98	-23,98	63	1,2555	31	42	8

9 Curriculum Vitae

Persönliche Daten

Name	Ann-Christin Groher, geb. Reuter
	Geboren am 29.11.1982 in Offenbach am Main

Bildungsweg

1999-2003	Georg-Kerschensteiner-Schule Obertshausen
	Abschluss: Abitur
2008-2013	Studium Biologie (Diplom); Johannes-Gutenberg-Universität Mainz
	Abschluss: Diplom Biologe (sehr gut)
2017-2020	Promotionsstudium an der TU-Darmstadt
	Tätigkeit im Labor von Prof. Dr. Beatrix Süß
	Angestrebter Abschluss: Doktor der Naturwissenschaften (Doctor rerum naturalium)

Beruflicher Werdegang

2013-2015	Mitarbeiterin in Forschung und Entwicklung (R&D) bei der inno-train Diagnostik GmbH
Seit 2020	Projektmanager im Bereich Entwicklung (R&D) bei der Virotech Diagnostics GmbH

Besuchte Veranstaltungen und Präsentationen

2017	**LOEWE Schwerpunkt CompuGene, 2. Retreat, Annweiler-Bindersbach** Vortrag: A Massive Approach towards the *in silico* Prediction of Riboswitch Performance
2017	**35. Rabensteiner Kolleg**, Pottenstein, Deutschland
2017	**LOEWE Schwerpunkt CompuGene Symposium: Computer-aided Engineering of Synthetic Genetic Circuits, Darmstadt** Poster: Tetrazykline Dimers - A Massive Approach towards the *in silico* Prediction of Riboswitch Performance
2017	**LOEWE Schwerpunkt CompuGene, 3. Retreat, Darmstadt** Poster: Machine learning with Tetrazykline Dimers - A large scale apporach towards the *in silico* Prediction of Riboswitch Performance
2018	**LOEWE Schwerpunkt CompuGene, 4. Retreat Obergurgl, Österreich**

Ehrenwörtliche Erklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit entsprechend den Regeln guter wissenschaftlicher Praxis selbstständig und ohne unzulässige Hilfe Dritter angefertigt habe.

Sämtliche aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie sämtliche von Anderen direkt oder indirekt übernommenen Daten, Techniken und Materialien sind als solche kenntlich gemacht. Die Arbeit wurde bisher bei keiner anderen Hochschule zu Prüfungszwecken eingereicht.

Die eingereichte elektronische Version stimmt mit der schriftlichen Version überein.

Darmstadt, den ___.___._____

.....

(Ann-Christin Groher)